Panagiota Kiourti

pkiourti@gmail.com | Google Scholar | LinkedIn | GitHub

About

AI / ML Engineer with expertise in deep learning security, computer vision, and 3D geometry. Skilled in building ML models, publishing pioneering research on backdoor defenses, and delivering production-ready code that powers high-impact products.

Education

Boston University Boston, MA
Ph.D. in Computer Engineering, GPA: 3.91/4.00 Sept 2018–Jun 2024

- Thesis: Enhancing Deep Learning Security Through Explainability and Robustness
- Pioneered first reinforcement learning backdoor attack methodology, advancing neural network security research; developed defense strategies.
- Implemented defensive strategies against deep learning backdoors, using explainability to improve robustness.
- Theorized a new definition for the robustness of XAI feature attribution methods and their evaluation.
- Taught Cybersecurity (EC521) and Operating Systems (EC440) modules to 100+ students, mentored students in other projects.

National Technical University of Athens Athens, Greece
MEng in Electrical and Computer Engineering, GPA: 7.945/10.0 Sept 2015

- Major in Computer Science; Minor in Bioengineering
- Thesis: Sentiment Analysis using hybrid n-grams

Publications

- Kiourti, Panagiota. Enhancing deep learning security through explainability and robustness. Diss. Boston University, 2024.
- Fu, Feisi, Panagiota Kiourti, and Wenchao Li. "Dormant Neural Trojans." 2023 International Conference on Machine Learning and Applications (ICMLA). IEEE, 2023.
Acceptance Rate: $< 32\%$
- Kiourti, Panagiota, et al. "Misa: Online defense of trojaned models using misattributions." Proceedings of the 37th Annual Computer Security Applications Conference 2021.
Citations: 8, **Acceptance Rate**: 24.5%, **DOI:** 10.1145/3485832.3485908
- Kiourti, Panagiota, et al. "Trojdrl: evaluation of backdoor attacks on deep reinforcement learning." 2020 57th ACM/IEEE Design Automation Conference (DAC). IEEE, 2020.
Citations: 158, **Acceptance Rate**: 22.7%, **DOI:** 10.1109/DAC18072.2020.9218663

Media Coverage

WIRED: "Tainted Data Can Teach Algorithms the Wrong Lessons.", Knight, Will. Nov 25, 2019
URL: https://www.wired.com/story/tainted-data-teach-algorithms-wrong-lessons/

Schneier on Security: "Manipulating Machine Learning Systems by Manipulating Training Data" Nov 29, 2019
URL: https://www.schneier.com/blog/archives/2019/11/manipulating_ma.html

Boing Boing: "Tiny alterations in training data can introduce 'backdoors' into machine learning models," Cory Doctorow Nov 25, 2019
URL: https://boingboing.net/2019/11/25/backdooring-ai.html

Towards Data Science (Medium): "Neural Trojan attacks and how you can help."
URL: https://towardsdatascience.com/neural-trojan-attacks-and-how-you-can-help-df56c8a3fcdc

Academic Service

Reviewer (2019–2025)

- **Conferences:** DAC 2019–2020, HSCC 2020, DATE 2021–2022, DSN 2021–2022, ICCAD 2021, TACAS 2020 & 2023
- **Journals:** IEEE TIFS 2022 and Springer JAR 2024 (invited)

Open Source Impact

GitHub repo: **rl_backdoor**

GitHub repo: **visualizing_nn_adversarial_attacks**

Fellowships

- Boston University Student Fellowship for Grace Hopper Celebration 2020
- DAC Young Fellowship 2020

Teaching

- **Teaching Assistant** at Boston University Spring 2020
  Cybersecurity (EC521)
- **Teaching Assistant** at Boston University Fall 2019
  Operating Systems (EC440)
- **Research Project Mentor** at Boston University Summer 2021
  Trojan Attacks in Face Recognition

Experience

Carbon3D Redwood City, CA
Senior Software Engineer—Geometry July 2024–Nov 2025

- **Invented a novel and fast method** to automate **3D denture orientation for printing** (Computer Vision, Statistical Analysis), achieving **99.7% success rate** and enabling > $**1M/year** in new revenue.
- Leveraged camera calibration techniques to recover 3D tilt of hardware components across cross-functional teams with production-grade accuracy.

Carbon3D Redwood City, CA
Research Scientist Intern—Geometry May 2022–Aug 2022

- Researched alternatives for representing large 3D meshes and developed a method to implicitly represent a 3D surface using a neural network following NVIDIA's NGLOD method using PyTorch, achieving **98% compression rate** of large objects.

SRI International Menlo Park, CA
Computer Scientist Intern—CS Lab Jun 2020–Aug 2020

- **Created a novel backdoor detection method** for image classification Neural Networks (under the funding of the IARPA TROJAI program), **achieving 96% AUC**.

SRI International Menlo Park, CA
Reinforcement Learning Intern May 2019–Aug 2019

- Conducted research to find signatures of Trojan/Backdoor attacks on Neural Network Policies (Reinforcement Learning Agents)
- Extended backdoor attacks in LSTM-based RL agents with **100% attack success rate** on high-dimensional inputs in Atari environments. Trained a recurrent neural network that tried to model the sequence of statistical metrics on explanation maps corresponding to the images fed to the RL policy.
- Total work of 6 months across two teams, while being mentored by the technical director of the SRI CS Lab.

Flashchat Athens, Greece
Full-Stack Developer Aug 2017–Apr 2018

- Built AI Facebook Messenger chatbots to streamline e-commerce for both small and big retail companies through Messenger chatbots.
- Developed the frontend and backend of a UI that **boosted** efficiency of **customer data workflows 4×.**
- Implemented web services for automatically extracting reports about numbers tracked by Google Analytics.

Intrasoft International Athens, Greece
Software & Test Engineer Apr 2016–July 2017

- Built UIs for Lottery Management Software projects.

- Tested RESTful Web Services automatically using SoapUI Groovy Scripts and prepared and executed Test Cases for the product and project components.
- Delivered QA reports for project release, prepared release notes.
- Interacted with the customer during User Acceptance Test.

National Technical University of Athens
Distributed Knowledge & Media Systems Group Athens, Greece
Research Assistant Oct 2015–Apr 2016

- Researched and developed the sentiment classification techniques again n-gram based text summarization applied to Sentiment Classification (Polarity Classification) on larger texts, e.g., IMDB reviews.

Skills

- **ML/AI:** Deep Learning, Computer Vision, Explainable AI, Generative AI, Reinforcement Learning
- **3D/Geometry:** Trimesh, Open3D, PyMeshLab, numpy-stl, PyVista, scipy.spatial.transform, fast-simplification, scikit-geometry, CGAL, libigl, Eigen
- **Programming:** Python, C, C++, Java, JavaScript, ReactJS, NodeJS, SQL, NoSQL
- **Image processing & CV:** OpenCV, scikit-image, scipy
- **ML/modeling:** PyTorch, TensorFlow, Keras, scikit-learn, ONNX, captum.ai
- **Graphs & spatial:** NetworkX, scipy.spatial.KDTree
- **Visualization & Apps:** Plotly (Mesh3d, Scatter3d, custom hovers/camera), Matplotlib and mplot3d (2D/3D), PyVista, Dash, Flask
- **Performance/Systems/Build**: CUDA toolkit, NVIDIA container toolkit, Bazel
- **Databases:** MongoDB, PostgreSQL, MySQL
- **DevOps / CI/CD:** GitHub Actions (workflows, CI/CD automation)
- **Core Strengths:** Analytical Thinking, Communication, Adaptability, Initiative, Dedication, Persistence
- **Full-Stack Development**, **Cybersecurity**
- **Languages**: Greek (Native), English (Fluent - C2)

Professional Development

- **3D Machine Learning with PyTorch3D**
  Certificate of Completion | Educative 2025
- **Neural Radiance Fields (NeRF)**
  Certificate of Completion | Udemy 2025
- **Mathematics Behind LLMs & Transformers**
  Certificate of Completion | Udemy 2025
- **Practical Multi-Armed Bandit Algorithms**
  Certificate of Completion | Udemy 2025

- **YOLOv8 Object Detection for Number Plate Recognition**
  Certificate of Completion | Udemy 2025
- **9th Summer School on Formal Techniques** Summer 2019
  Menlo College, Atherton, CA
- **The Analytics Edge** Honor Code Certificate
  edX MITx | Online 2015
- **Bioinformatics Algorithms**
  Statement of Accomplishment with Distinction, Grade Achieved: 94.53%
  University of California San Diego | Online 2014