Oussama Bouanani

Berlin | E-Mail | Website | Scholar | LinkedIn | GitHub

Experience

Research Assistant — Explainable AI & Mechanistic Interpretability
Fraunhofer HHI, Berlin Mar 2023 – Present

- Adapted explainability methods to AlphaZero-style reinforcement learning models to better understand model decisions and internal behavior.
- Developed and refined neuron-labeling methods for computer vision backbones, producing more faithful and semantically precise descriptions of internal representations.
- Extended neuron-level analysis methods to large language models to better understand and steer model behavior.

Independent Software Developer
Berlin Feb 2017 – Mar 2023

- Built and maintained commercial Unity/C# products for PC and mobile, including Asset Store tools used by more than 1,500 developers.
- Delivered freelance software projects for private clients and small studios, covering interactive applications, educational tools, and digital art projects.

Research Assistant — Privacy & Robustness in ML
Fraunhofer AISEC, Berlin May 2021 – Jul 2021

- Benchmarked membership inference and data reconstruction attacks across model architectures and training choices.
- Evaluated mitigation strategies including augmentation, upsampling, and noise injection.
- Identified hyperparameters that reduced privacy leakage by up to 40% and proposed a theoretical upper bound on membership inference attack success.

Technical Skills

Machine Learning: PyTorch · Hugging Face · scikit-learn · Computer Vision · Interpretability · Reinforcement Learning · LLM Fine-tuning · RAG · LangChain · LangGraph
MLOps & Deployment: Docker · FastAPI · MLflow · Weights & Biases · Slurm · GCP
Data & Programming: Python · C# · SQL · PostgreSQL · pandas · NumPy · Matplotlib
Tools: Linux · Git · Unity

Projects

NeuroLens

- Built a reusable library for describing and labeling individual neurons in vision backbones.
- Designed modular components for dataset setup, label generation, neuron selection, and evaluation to make neuron-labeling workflows more reproducible and easier to inspect.

Real-Time Strategy Engine for Unity

- Built and released a 5-star commercial Unity/C# framework with modular systems for pathfinding, memory pooling, spatial queries, and multiplayer networking.
- Maintained the product since 2017 and supported a community of more than 1,500 developers.

Publications

Contrastive Semantic Projection: Faithful Neuron Labeling with Contrastive Examples
XAI World Conference 2026 2026

- Proposed a contrastive approach to neuron labeling that produces more specific and faithful labels, with improvements shown across experiments and a melanoma detection case study.

Influence of Training Parameters on Neural Networks' Vulnerability to Membership Inference Attacks
Trustworthy AI Workshop at INFORMATIK 2022 2022

- Studied how training choices such as batch size, activation function, and normalization affect privacy risk under membership inference attacks.

Education

Freie Universität Berlin — M.Sc. Computer Science
Grade: 1.4
Thesis: Contrastive Explanations for Neuron Semantic Labeling

Freie Universität Berlin — B.Sc. Computer Science
Grade: 1.3
Thesis: The influence of training parameters and architectural choices on the vulnerability of neural networks to membership inference attacks