Bekhzod Anvarov

Machine Learning Engineer
Self-Supervised Learning (JEPA), Deep Learning, Anomaly Detection

Turin, Italy
virtsudo@proton.me | virtsudo.vercel.app | linkedin.com/in/virtsudo | github.com/virtsudo

SUMMARY

Computer Engineering graduate (MSc, Cybersecurity track) from Politecnico di Torino, applying deep learning to security data. For my thesis I built a self-supervised Joint-Embedding Predictive Architecture (JEPA) in PyTorch and tested whether its representation keeps a classifier more accurate than raw features under concept drift. I have also trained FFNN, RNN, GNN and BERT models for security tasks and have used Python for 5+ years. Open to junior roles and internships in machine learning, on-site or remote.

EDUCATION

Master of Science in Computer Engineering (Cybersecurity track)
Politecnico di Torino, 2026

Bachelor's Degree in Computer Engineering
Politecnico di Torino, 2022

RESEARCH EXPERIENCE

JEPA Representation Learning for Intrusion Detection on Real-World Drifting Network Traffic
Master's thesis, Politecnico di Torino, 2026
Supervisors: Prof. Guido Marchetto, Prof. Alessio Sacco, Cristiano Serra, MSc
Does a representation learned from unlabeled network traffic keep an intrusion classifier more accurate than raw flow features on traffic recorded years later?

- **Built** a self-supervised JEPA model in PyTorch with a context encoder, an EMA target encoder, a predictor and grouped masking of correlated features. Pre-trained one encoder per year on unlabeled flows of real Internet backbone traffic from 2011, 2016 and 2021 (MAWIFlow benchmark).
- **Evaluated** it against raw flow features by training four classifiers (logistic regression, decision tree, random forest, XGBoost) on each year and testing them on every year (ROC-AUC, PR-AUC). Replaced the benchmark's scaling, which lost most of the resolution in the features.
- **Found** that the representation makes the classes more linearly separable. In domain, logistic regression on it closes most of its ROC-AUC gap to XGBoost on raw features, but the representation does not reduce the drop across years, and the 2011 labels mean something different from the later ones. Also implemented VICReg and SIGReg regularizers, which did not improve the classifiers.

Stack: Python, PyTorch (CUDA), scikit-learn, XGBoost, Polars, NumPy, Matplotlib

PROJECTS

Deep Learning for Cybersecurity Labs
Stack: Python, PyTorch, Jupyter, Hugging Face
Built an FFNN for network intrusion detection and FFNN, RNN and GNN models for malware classification.
Fine-tuned BERT to analyze SSH logs and tag them with MITRE ATT&CK tactics.
GitHub: github.com/virtsudo/dl-cybersecurity-labs

SKILLS

Machine learning: PyTorch (CUDA), scikit-learn, XGBoost, Hugging Face, Polars, NumPy, Matplotlib, Jupyter
Programming: Python, C/C++, SQL (MySQL, SQLite)
Tools & environments: Linux shell, Git/GitLab, LaTeX
Focus areas: Self-supervised learning (JEPA), Representation learning, Anomaly detection, Concept drift

RELEVANT COURSEWORK

- **AI and Cybersecurity:** deep learning (CNN, RNN, LSTM), NLP, anomaly detection, traffic classification.
- **Data Science and Database Technology:** classification, clustering, association rules, data warehouses, OLAP.

LANGUAGES

Uzbek (Native), Russian (C1), English (B2), Italian (B1)