KONSTANTINOS GEORGIOU

Machine Learning Engineer

☎ +1(865) 978-9244
</> https://gkos.dev

@ gkos.mldev@gmail.com
in linkedin.com/in/gkos

United States
github.com/drkostas

SUMMARY

- Ph.D. Machine Learning Engineer with 8+ years building production ML systems for Large Language Models, NLP, Computer Vision, and real-time data pipelines at scale.
- Proven track record designing and deploying ML solutions that drive measurable business impact through rigorous experimentation, A/B testing, and metrics-driven evaluation. Led cross-functional initiatives integrating GenAI, RAG pipelines, and traditional ML into scalable production platforms, combining deep technical expertise with clear stakeholder communication.
- Presented work at top-tier AI conferences and contributed to open-source initiatives, demonstrating commitment to shipping high-quality ML products and collaborative knowledge sharing.

EXPERIENCE

Summer 2025

Applied Scientist – L5

Amazon- US

Skills: LLM Evaluation · Prompt Engineering · Retrieval Augmentation · Agentic Workflows · NLP · Python · PyTorch · Metrics Design

- Designed and deployed a multi-agent LLM framework to convert natural language into executable code, improving functional success rates by 12% through iterative experimentation and rigorous evaluation, and production monitoring at scale.
- Built a real-time detection system that correctly identified 93% of unfeasible requests, achieving 4.5x performance improvement over baseline and preventing generation of non-functional code through automated quality evaluation.
- Conducted systematic analysis uncovering critical misalignment between heuristic LLM evaluations (98% score) and functional correctness (79% pass rate), delivering actionable insights into limitations of current assessment methods.

2024 – Now

Co-Founder & ML Engineer

XPensAI Ltd – United Kingdom

Skills: Python · Computer Vision · Generative AI · Real-time ML · Production Systems · Cloud Infrastructure · AWS · Azure · SQL

- Launched AI-powered SaaS platform used by 30+ businesses, reducing manual expense entry by 65% through automated ML pipelines and real-time processing.
- Led development and deployment of core ML algorithms for automated expense tracking and receipt processing, improving processing speed by 120% while ensuring system reliability and production monitoring.
- Built and deployed deep learning and computer vision solutions, achieving 95% accuracy on receipt scanning, implementing scalable architecture and continuous monitoring for production quality.

2021 – Now

Machine Learning Researcher

University of Tennessee - US

Skills: Python · PyTorch · Deep Learning · Computer Vision · NLP · LLMs · Transformer Models · Model Evaluation · Experimentation

- Developed self-supervised framework using masking strategies and teacher-guided distillation to learn robust visual representations, validated through rigorous evaluation across multiple downstream tasks.
- Created feature masking strategy for Transformer models that raised Macro F1 by 6% and reduced feature reliance by 15%, improving detection robustness and interpretability through systematic experimentation and evaluation.
- Advanced masked image modeling research by tailoring scale factors for multi-modal data, achieving 5% accuracy improvement over state-of-the-art across 4 datasets through rigorous benchmarking and evaluation.
- Developed fine-tuning strategies for multi-modal self-supervised model, reducing training time by 32% and improving Macro F1 by 5.4% while delivering production-ready implementations for client pipeline.
- Authored foundational study on LLM security ("Occasionally Secure," Arxiv 2024), establishing principles for secure and reliable code generation systems.

2019 – 2021

Data Engineer

Performance Technologies S.A - Greece

Skills: Python · SQL · Data Pipelines · Machine Learning · Batch & Real-time Processing · GCP · Apache Spark · Docker · Monitoring

- Led rapid completion of terabyte-scale data replication project for telecommunications provider, reducing processing time from days to minutes and ensuring real-time data access for ETL, analytics, and production pipelines.
- Developed machine learning model to predict order fulfillment times, collaborating with business stakeholders to analyze operations and deliver 34% improvement over baseline through iterative experimentation.
- Designed and deployed SIP call quality benchmarking service across public institutions, implementing automated monitoring and evaluation systems that enabled performance tracking and service provider assessment.

2018 – 2019

Machine Learning Researcher

University of Patras - Greece

Skills: Python · Algorithm Design · Machine Learning · Performance Optimization · Apache Spark · Docker · Graphs · SQL

- Conducted ML research specializing in graph neural networks and scalable algorithm design for large-scale network analysis.
- Optimized community detection algorithm execution time by 84%, creating first scalable solution while maintaining high accuracy through systematic performance optimization and rigorous evaluation.

2017 - 2018

Software Engineer

Global Voices Ltd - UK

Skills: Python · Software Engineering · SQL · Production Deployment · System Reliability · CI/CD · Code Quality · Monitoring

- Led development on proprietary CMS, implementing key features and reducing critical bugs through rigorous testing and code reviews to improve system reliability.
- Optimized continuous integration and deployment pipelines, enhancing efficiency and reliability of releases, resulting in 50% reduction in rollbacks through automated testing and monitoring.

EDUCATION

2025

- PhD in Data Science & Engineering

  University of Tennessee

  - Received Fellowship Award from the University of Tennessee Graduate School and Tickle College of Engineering.
  - Developed production-ready ML implementations from scratch, including CNNs and RL agents, with expertise in experimentation, A/B testing, evaluation metrics, and advanced statistical modeling for rigorous hypothesis validation.
  - Doctoral research focused on Self-Supervised Learning, NLP, LLM Security, and Trustworthy AI, leading to multiple publications in top-tier conferences.

- Integrated Master's in Computer Science & Engineering

2019

  University of Patras

  - Developed an innovative distributed algorithm for community prediction in social graphs, achieving significant improvements in scalability and accuracy.

PUBLICATIONS

- *Trustworthy AI for Early Dementia Detection: Robust Feature Masking and Clinical Interpretability.* - CHASE 2025
- *Improving Masked Image Modeling with Adaptive Masking and CLIP Distillation.* - ICCV 2025
- *Advancing Multi-scale Remote Sensing Analysis through Self-Supervised Learning Fine-tuning Strategies.* - IEEE IGARSS 2024
- *Koopman-based Transition Detection in Satellite Imagery.* - IEEE IGARSS 2024
- *Occasionally Secure: A Comparative Analysis of Code Generation Assistants.* - Arxiv 2024
- *Cross-Scale MAE: A Tale of Multi-scale Exploitation in Remote Sensing.* – NeurIPS 2023
- *Semantic Segmentation in Aerial Imagery using Multi-level Contrastive Learning with Local Consistency.* – WACV 2023
- *A Distributed Hybrid Community Detection Methodology for Social Networks.* – Algorithms 2019