Apostolos Modas

SENIOR AI ENGINEER | APPLIED SCIENTIST | TECHNICAL LEADERSHIP

Lausanne, Switzerland | apostolos.modas@gmail.com
LinkedIn | GitHub | Google Scholar | Website

PROFESSIONAL SUMMARY

AI engineer and applied scientist with 9+ years of experience in computer vision, machine learning, and generative AI. At Sony AI, my work includes multimodal model evaluation, Intellectual Property (IP) detection, and datasets and tools used by other teams. I have led technical projects from research through delivery and contributed to FHIBE research featured on Nature's December 2025 cover. I completed my PhD at EPFL, with first-author papers at NeurIPS, CVPR, and ECCV.

PROFESSIONAL EXPERIENCE

SONY AI | Senior AI Engineer

Zurich, Switzerland | Oct 2025 – Present

- Built an Intellectual Property (IP) risk benchmark for text-to-image models, evaluating 20 models (13 open-weight, 7 proprietary), 640 IP entities, 6,142 direct and indirect prompts, and approximately 120,000 generated images. Built LLM-based prompt generation and image-generation workflows and analyzed results across entity categories and sample sizes. The manuscript is under review.
- Created a human-annotated image set to compare VLM-based IP detectors using precision, recall, and F1, while accounting for annotator agreement. Selected a detector for the larger benchmark and helped design a Prolific annotation workflow to collect data for fine-tuning.
- Developing IP detection runtime guardrails for prompts, images, and video: evaluating prompt-level approaches, refining an implemented image detector, and testing video-level methods. Also designing a text-to-video benchmark covering implicit IP references, temporal scenarios, and audio-related IP.
- Led the technical handover of FHIBE to Mozilla Data Collective, preparing dataset releases, derivatives, and metadata. Built tools to trace image relationships and handle updates after consent revocation; delivered stakeholder training and remained the main technical contact.
- Led packaging and distribution of a FHIBE-related dataset to Sony business units, collaborating on delivery and the internal commercialization process.
- Exploring synthetic image augmentation and iterative dataset expansion to study how training-data distributions affect out-of-distribution generalization.

SONY AI | Senior Research Scientist

Zurich, Switzerland | Oct 2022 – Sep 2025

- Contributed robustness experiments and cross-dataset annotation analyses to FHIBE research published in Nature (December 2025 cover).
- Technically led a team of three AI researchers and engineers to make a computer vision evaluation framework dataset-agnostic. Owned technical direction and delivered a common data interface, configurable experiments, and automated checks across 9 datasets, 9 CV tasks, and 22 model configurations. Much of the work was released in the team's public repository.
- Built a diffusion-based image anonymization pipeline combining automated processing with human review. Trained three QA reviewers on annotation and quality checks; the pipeline prepared more than 25% of FHIBE images and has since been reused by other Sony teams.
- Contributed to research on responsible data curation presented as a NeurIPS 2023 Oral.

EPFL | Research Assistant

Lausanne, Switzerland | Apr 2018 – May 2022

- Researched adversarial robustness, invariance, and generalization of deep neural networks under distribution shifts. Published first-author papers at NeurIPS, CVPR, and ECCV.
- Highlights: *SparseFool* for sparse adversarial attacks and *PRIME augmentations* for robustness to common corruptions.

IBM RESEARCH | Research Intern

Zurich, Switzerland | Feb 2021 – Jun 2021

- Developed defenses against black-box adversarial attacks and contributed to IBM's open-source Adversarial Robustness Toolbox (ART).

EPFL | Research Intern

Lausanne, Switzerland | Sep 2017 – Feb 2018

- Developed shared-memory synchronization components for dynamic data-flow programs used in HEVC decoding.

TECHNICAL SKILLS

Research: Computer vision, multimodal models, LLMs and VLMs, text-to-image/video generation, model evaluation, data-centric AI, adversarial robustness, fairness, privacy, AI safety.

Engineering: Python, PyTorch, Hugging Face, Pandas, OpenCV, scikit-learn, AWS (EC2, S3), Docker, Linux, Git, CI/CD, C/C++; configurable evaluation pipelines, experiment tracking, data validation.

EDUCATION

Doctor of Philosophy (PhD) | Electrical Engineering | 2018–2022

Swiss Federal Technology Institute (EPFL), Switzerland

Field of study: Deep Learning

Thesis: Robustness and Invariance Properties of Image Classifiers | Best Thesis Award nominee

Master of Science (MSc) | Computer Science | 2015–2017

National and Kapodistrian University of Athens, Greece

Field of study: Machine Learning

Valedictorian | GPA: 9.77/10

Bachelor of Science (BSc) | Informatics and Telecommunications | 2009–2015

National and Kapodistrian University of Athens, Greece

Field of study: Signal Processing

SELECTED PUBLICATIONS

- Fair human-centric image dataset for ethical AI benchmarking. Nature, 2025 (December cover).
- Ethical considerations for responsible data curation. NeurIPS, 2023 (Oral).
- PRIME: A few primitives can boost robustness to Common Corruptions. ECCV, 2022 (first author).
- Neural Anisotropy Directions. NeurIPS, 2020 (first author).
- Hold me tight! Influence of discriminative features on deep network boundaries. NeurIPS, 2020 (first author).
- SparseFool: A few pixels make a big difference. CVPR, 2019 (first author).

PUBLIC RESEARCH AND DATA

Fairness benchmark: github.com/SonyResearch/fairness-benchmark-public

FHIBE dataset releases: Mozilla Data Collective

LANGUAGES

Greek (native) | English (proficiency) | German (basic)