

Mohamed Abdelfattah

AI PHD CANDIDATE, EPFL SWITZERLAND

(+41) 78 254 95 38 | mohamed.abdelfattah@epfl.ch | Homepage | Google Scholar | GitHub | LinkedIn

Summary

- Research:** PhD at EPFL VITA Lab building self-supervised **multimodal foundation models** (VLMs, LLMs) for video understanding/emodied AI.
- Publications:** 4 first-author publications at **CVPR x2**, **ECCV**, and **NeurIPS**.
- Industry:** AI Research Scientist Intern at **Meta Reality Labs Zurich**; developed multimodal egocentric models for Meta's AR/VR AI systems.
- Scale:** Trained **10B+ parameter**, **10M+ sample** systems on **500+ H200 GPUs** through Switzerland's **Swiss AI Initiative** on Alps.

Research & Industry Experience

EPFL École Polytechnique Fédérale de Lausanne

Lausanne, Switzerland

Ph.D. Researcher (Prof. Alexandre Alahi, VITA Lab)

Sep 2022 – Present

- Develop OSKAR [NeurIPS 2025], a **unified multimodal world model** for video, skeleton, and text, and own the full pipeline from data curation to evaluation; within Switzerland's **Swiss AI Initiative**, train **10B+ parameter** systems on **512 H200 GPUs** on Alps, reaching 86.1% on Kinetics-400, 91.1% on NTU120-XSub, 37.9 mAP on AVA, and 50.4 R@1 on MSRVT on 5 downstream tasks.
- Develop MaskCLR [CVPR 2024] and S-JEPA [ECCV 2024] for **strong action representations**, improving cross-source generalization by 26.6 points, noise robustness by up to 14.9 points vs. MotionBERT baselines, and transfer to PKU-II to 71.4% vs. 58.4% for SkeletonMAE.

Meta Reality Labs

Zurich, Switzerland

AI Research Scientist Intern

May – Nov 2025

- Designed and trained OSIRIS, a LLaMA-based **egocentric multimodal LLM** for human-state tracking from video, audio, eye-tracking, and dialogue; led model, dataset, and benchmark design for Meta's embodied **AR/VR** roadmap.
- Developed OSMO, the largest egocentric emotion dataset (110h, 288 subjects, 1000+ open-vocabulary categories), establishing the first benchmark for egocentric smart-glasses human-state tracking. Published as first-author at **CVPR 2026**.

SpreeAI

Nevada, USA

Computer Vision Applied Scientist

Jul 2022 – Feb 2023

- Architected a **conditional diffusion model** for high-fidelity face reconstruction achieving 20% quality improvement over production baseline; redesigned U-Net head-swapping pipeline preserving pose, skin tone, and illumination for AI-powered virtual try-on at scale.

King Abdullah University of Science and Technology (KAUST)

Thuwal, Saudi Arabia

Deep Learning Intern (Prof. Mohamed Elhoseiny, Vision-CAIR Group)

Mar 2022 – Jan 2023

- Led collection of ArtELingo: a multilingual **vision-language** dataset across 84K artworks and 4 languages enabling culturally diverse multimodal learning; co-authored **EMNLP 2022**.

Boston University

Remote, USA

Remote Visiting Research Student (Prof. Sarah Adel Bargal, IVC Group)

Jun – Dec 2021

- Designed augmentation pipelines and evaluation baselines for deformable industrial waste-object **semantic segmentation**; co-authored the ZeroWaste benchmark at **CVPR 2022**.

First-Author Publications

CVPR 2026 | work done at Meta Reality Labs & EPFL internship

Meta Reality Labs & EPFL

OSMO: Open-vocabulary Self-eMOTION Tracking

- Introduced OSIRIS, a novel LLaMA-based **multimodal LLM** processing **egocentric** video, audio, eye-tracking, and dialogue for continuous emotion tracking. Paired with OSMO, the first and largest egocentric emotion dataset (110h, 288 subjects, 1000+ open-vocab categories). Achieves SOTA on a 5-task benchmark, outperforming EmotionLLaMA, AffectGPT, and DeSTA.

NeurIPS 2025

EPFL

OSKAR: Omnimodal Self-supervised Knowledge Abstraction and Representation

- Developed an omnimodal **self-supervised** abstraction objective training a single **unified encoder**; pretrained and finetuned OSKAR on **512 H200 GPUs** on Alps as part of a Swiss-wide open-science effort, surpassing specialized models across 5 diverse tasks: 86.1% Kinetics400, 91.1% NTU120, 50.4 R@1 MSRVT retrieval, and 37.9 mAP AVA localization, without manual annotations.

CVPR 2024

EPFL

MaskCLR: Attention-Guided Contrastive Learning for Robust Action Representation Learning

- Designed attention-guided masking with **multi-level contrastive learning**. Achieves state-of-the-art across NTU60, NTU120, and Kinetics400; improves cross-pose-estimator generalization by 26.6 percentage points and noise robustness by 5.9 percentage points vs. MotionBERT.

ECCV 2024

EPFL

S-JEPA: A Joint Embedding Predictive Architecture for Self-Supervised Skeletal Action Recognition

- Introduced **latent-space joint-embedding prediction** for skeletal sequences, achieving 93.1% on NTU60 and 71.4% transfer on PKU-II (vs. 58.4% for SkeletonMAE), outperforming reconstruction-based methods across 4 evaluation protocols.

Automated Detection of Isolated REM Sleep Behavior Disorder using Computer Vision 

- Built, with Stanford collaborators, a computer-vision pipeline for automated detection of isolated REM sleep behavior disorder from standard 2D sleep-lab video. Using optical-flow movement features from 172 video-polysomnograms, achieved up to **91.9%** accuracy.

Education

École Polytechnique Fédérale de Lausanne (EPFL)

Lausanne, Switzerland

Ph.D. in Computer Vision

Dec 2026 (Expected)

- Advisor: [Alexandre Alahi](#) , VITA Lab
- Research focus: **multimodal foundation models/LLMs**, self-supervised learning, representation learning, and action understanding

The American University in Cairo (AUC)

Cairo, Egypt

B.S. in Computer Engineering

Jun 2022

- Highest Honors, Rank 5/80, GPA 3.91/4.0; **Dean's List of Honors**
- Double Minor in Mathematics and Business Administration; thesis on fine-grained text-to-image generation

Selected Co-Authored Work

EMNLP 2022

KAUST

ArtELingo: A Million Emotion Annotations of WikiArt with Diversity over Language and Culture 

- Built a multilingual vision-language dataset over 84K artworks for cross-cultural captioning and emotion understanding.

CVPR 2022

Boston University

ZeroWaste Dataset: Towards Deformable Object Segmentation in Cluttered Scenes 

- Built augmentation pipelines and evaluation baselines for deformable waste-object segmentation in weakly supervised real-world perception.

Building and Environment 2023

EPFL

A framework for Machine Learning-based human thermo-physiology modeling with computer vision 

- Co-authored, with EPFL ICE Lab collaborators, a non-intrusive thermal monitoring framework combining RGB/IR sensing, computer vision, and LSTM-based thermo-physiology modeling to estimate local skin temperatures with **<0.5°C RMSE** in most body regions.

Skills

Research	Multimodal Foundation Models, VLMs/MLLMs, Self-Supervised Learning, Representation Learning, Video & Embodied AI
AI	Generative AI, LLM Fine-tuning & Evaluation, Multimodal Learning, Retrieval & Alignment
Code	Python, PyTorch, JAX, CUDA, Hugging Face Transformers, FlashAttention, vLLM, DeepSpeed, NumPy, SciPy
Systems	Distributed Training (FSDP, DDP), Large-Scale GPU Training, High-Throughput Data Pipelines, Slurm, W&B
MLOps	Model Serving, Inference Optimization, Experiment Tracking, Reproducibility, Docker, API Integration, Evaluation Pipelines
Data	Large-Scale Data Curation, Pretraining Pipelines, Dataset Design, Benchmarking, Data Processing & Augmentation
Engineering	Modular Code Design, Debugging Large-Scale Systems, Performance Profiling, Version Control (Git)

Awards

2022	High Academic Achievement Award , top-5 graduate in the class of 2022	AUC
2022	PA Cup , top academic and extracurricular achievement in graduating class	AUC
2021	Research Grant , \$4K grant for computer vision research	AUC
2019	ROV Excellence Award , top-10 Middle East ranking in MATE ROV Competition	MATE ROV
2019	Best Design Award , most efficient mine-detection rover at national robotics competition	Minesweepers
2018	First Place , winner of CSCE Programming Contest	AUC
2018	Highest Achiever and Reader of the Year , strongest essay writing performance in 2018	AUC
2017	AGFE Full Scholarship , \$160K merit scholarship for outstanding STEM potential	AUC

Academic Service

École Polytechnique Fédérale de Lausanne (EPFL)

Lausanne, Switzerland

Lead Teaching Assistant, Deep Learning for Autonomous Vehicles

Feb 2023 – Present

- Lead the PhD TA team, mentor **150+ master's students**, and advise 10 semester-project teams on research scoping, experimentation, and technical communication through an end-to-end autonomous driving stack.

CVPR, ECCV, NeurIPS

Reviewer

2024 – Present

- Contribute collaborative peer review on multimodal learning, egocentric/video understanding, representation learning, and computer vision for flagship venues including CVPR, ECCV, and NeurIPS.