Salaheldin Mohamed

Doctoral Researcher @ Institut Polytechnique de Paris · Computer Vision & Generative AI
salaheldin.y.mohamed@gmail.com | Personal Website | GitHub | LinkedIn

Research Focus

I am currently a Doctoral Researcher at Télécom Paris, Institut Polytechnique de Paris. Previously, I was a master's thesis student at the Max Planck Institute for Informatics, supervised by Prof. Dr. Christian Theobalt. My work focuses on **generative vision, controlled video, and image synthesis**. I am particularly interested in developing **physically plausible and realistic video and image models.**

Education

Institut Polytechnique de Paris — *Paris, France*
Ph.D. in Computer Science — *2026 – 2029*

- Focus: Physically Plausible and Realistic Video and Image Models .

Saarland University — *Saarbrücken, Germany*
M.Sc. in Computer Science — *2023 – 2026*

- Focus: Computer Vision, Machine Learning, and Advanced Deep Learning Architectures.

Alexandria University — *Alexandria, Egypt*
B.Sc. in Computer and Systems Engineering — *2018 – 2023*

- **Bachelor Thesis:** Masked Controlled Image Generation (Supervisor: Prof. Dr. Ayman Khalafallah)

Publications & Preprints

- **Mohamed, S.**, Hamza, M. M., Dabral, R., & Theobalt, C. *Towards Realistic Co-Speech Video Generation via Reward-Free Reinforcement Learning and Multilevel Speech Guidance.* **NeurIPS 2026**. [Project Page]
  - Formulated a novel reward-free RL (no human preference or expert and reward models) framework combined with multi-level speech representations to generate temporally consistent, high-fidelity human speech videos without requiring scalar reward engineering.
- **Mohamed, S.**, M. Hamza, M., Dabral, R., Theobalt, C. *"Gaussian Windows for Multi-Focal Attention."* (Under Review)
  - Introduced a Gaussian window mechanism that allows the attention to capture multifocal representation of data structure space instead of just locality imposed by positional encoding.
- Han, D.*, **Mohamed, S.***, Li, Y., Denzler, J. *"Diffusion-based Identity-Preserving Facial Privacy Protection."* **IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2025)**. [Paper]
  - Proposed a latent diffusion obfuscation pipeline preserving natural facial expressions and fine spatial details while effectively preventing automated identity matching systems.
- **Mohamed, S.**, Han, D., Li, Y. *"Fusion is all you need: Face Fusion for Customized Identity-Preserving Image Synthesis."* *arXiv preprint arXiv:2409.19111*, 2024. [arXiv]
  - Designed an identity-preserving conditioning pipeline utilizing cross-attention feature injection for zero-shot personalized subject generation in diffusion pipelines.
- Han, D.*, **Mohamed, S.***, Li, Y. *"ShieldDiff: Suppressing Sexual Content Generation from Diffusion Models through Reinforcement Learning."* *Preprints*, 2024. [Preprint]
  - Developed an RL-guided fine-tuning framework that unlearns harmful concept representations in latent space without degrading benign generative capability.
- **Mohamed, S.** *"TCIG: Two-Stage Controlled Image Generation with Quality Enhancement through Diffusion."* *arXiv preprint arXiv:2403.01212*, 2024. [arXiv]
  - Engineered a two-tier generative model combining structural pose layout conditioning with progressive latent diffusion refinement.

Research Experience

Max-Planck-Institut für Informatik *Saarbrücken, Germany*
Master Thesis Researcher (Advisors: Prof. Christian Theobalt, Dr. Rishabh Dabral) *Jan 2025 – Feb 2026*

- Researched generative diffusion models tailored for human video dynamics.
- Engineered a reward-free reinforcement learning framework to align video diffusion trajectories directly with multi-modal acoustic features.
- Scale training across GPU clusters using PyTorch distributed data parallel (DDP) and DeepSpeed backends.

Huawei - Munich Research Center *Düsseldorf, Germany*
Research Intern *May 2024 – Oct 2024*

- Researched safety alignment mechanisms for text-to-image models, culminating in *ShieldDiff* (RL-based latent concept suppression).
- Developed an identity-preserving anonymization pipeline using conditional latent diffusion, resulting in a accepted publication at **ICASSP 2025**.
- Formulated zero-shot face fusion architectures for personalized image synthesis, surpassing existing baselines in identity fidelity.

Industry & Software Engineering Experience

Kapitalwise *New York, USA (Remote)*
Part-time Machine Learning Engineer *Jul 2021 – Oct 2023*

- Fine-tuned Open-Source LLMs (LLaMA-2) using Low-Rank Adaptation (LoRA) and custom financial instructional datasets.
- Architected NLP sequence classification pipelines to analyze credit card transaction metadata and detect user life-stage events.
- Designed high-concurrency caching infrastructures with Redis and RabbitMQ to serve sub-100ms inference responses.

Siemens *Cairo, Egypt*
Software Engineer Intern *Aug 2022 – Sep 2022*

- Implemented modern C++ reflection architectures, dynamic serialization engines (JSON/Binary), and dynamic runtime GUI generation.

Evolv Fit & Master Micro *Remote / Cairo, Egypt*
Machine Learning & Software Engineering Intern *Jun 2021 – Sep 2021*

- Created real-time pose estimation and hand gesture tracking algorithms using OpenCV and MediaPipe for automated workout logging.
- Optimized low-level data plotting tools in C++, reducing latency on large numerical arrays (1M+ points) from 60 seconds to under 1 second.

Technical & Infrastructure Skills

Generative Modeling & Vision: Latent Diffusion Models (LDMs), ControlNet, GANs, Transformers
Machine Learning & Alignment: Prefrence optimization(DPO, SDPO), Reinforcement Learning (GRPO), Reward Modeling, Concept Unlearning, LoRA/PEFT, Fine-Tuning LLMs
Frameworks & Libraries: PyTorch, HuggingFace (Diffusers/Transformers), DeepSpeed, Accelerate, OpenCV, TensorFlow, Scikit-Learn, JAX (Basic)
Distributed Computing & Compute: CUDA, Distributed Data Parallel (DDP), Slurm Cluster Management, Docker, AWS (EC2, S3, ECS), Git, Linux, LaTeX

Selected Open-Source Research Implementations

PyTorch D3PM (Discrete Diffusion Probabilistic Models) *[Colab Implementation]* *2024*

- Clean-room re-implementation of Google's Discrete Diffusion Probabilistic Models (D3PM) in PyTorch, implementing discrete Markov transition matrices for categorical latent spaces.

Attention-Based Skin Cancer Classification *2023*

- Constructed multi-scale spatial attention vision architectures on the HAM10000 medical image dataset, achieving 94% classification accuracy.

Health Monitor Distributed Big Data Log Engine *2022*

- Architected real-time log ingestion and batch analysis pipelines using Hadoop HDFS, MapReduce, and Apache Spark (PySpark).