Yahia Salaheldin Shaaban

yahia.shaaban@mbzuai.ac.ae | LinkedIn | GitHub | +971502631423

Education

- **Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)** — Abu Dhabi, UAE
  M.Sc. in Machine Learning: 4.0 — *Aug. 2024 – May 2026*
  I am working under the supervision of Dr. Salem Lahlou and Dr. Martin Takac on the intersection of Foundation Models and Reinforcement Learning, with a focus on reasoning over time series data and LLMs for combinatorial optimization (LLM4CO).

- **Faculty of Engineering, Alexandria University** — Alexandria, Egypt
  BSc in Computer and Communication Engineering; Graduated in the top 5% — *Sept. 2018 – Jul. 2023*
  - **Thesis:** Designed a personalization scheme for federated learning for neural collaborative filtering. Developed secure multi-party computation (SMPC) aggregation and integrated it into the Flower library. Supervised by Dr. Ahmed Kosba
  - **Courses Completed:** Algorithms and Data Structures, Operating Systems, Databases, Networks, Distributed Systems.

Experience

- **AIQ Intelligence** — Abu Dhabi, UAE
  Research Intern — *May 2025 – Present*
  - **Thesis Collaboration:** Conducting my master's thesis in collaboration with AIQ, focusing on time-series large language models for oil-well drilling sensor data.
  - **Multimodal Alignment for Time Series and LLMs:** Enhanced a CLIP-style multimodal encoder for aligning time-series signals and language descriptions. Developed a novel technique leveraging hard negatives extracted from both sensor data and associated NLP text to improve signal quality and retrieval performance.

  e

- **AIC-MCIT** — Alexandria, Egypt
  Full Time R&D Engineer — *Sept. 2023 - Sept. 2024*
  - **Remote Sensing:** Developed a multi-spectral model for crop field segmentation. Built an internal multi-source satellite dataset benchmark, including Planet Labs and Landsat data. Supervised the annotation pipeline and implemented human-in-the-loop techniques to enhance model performance.
  - **Biomedical Imaging:** Developed a data pipeline for DICOM images and a production-level segmentation model for breast cancer screening for Baheya hospital.
  - **Model Optimization and Deployment:** Optimized deep learning models using pruning and quantization. Deployed models on-premise and utilized Ray on top of Slurm, Docker, and TorchServe.

- **Flower Labs** — Remote - Cambridge University
  Flower Summer of Reproducibility — *Jul. 2023 - Oct. 2023*
  - **Open Source Contribution:** Replicated the FedPara paper (ICLR 2022) and integrated the implementation into the Flower framework.

- **Fatima Fellowship** — Remote
  Predoctoral Research Fellow — *Jul. 2023 - Sept. 2024*
  - **HyperKKL:** Worked with Dr. Umar Niazi (MIT postdoc) to extend the learning-based Luenberger observer to non-autonomous nonlinear dynamical systems (distribution shift problem in time series data). Employed low-rank meta-learning to adapt to new exogenous inputs. Built a Python framework to accelerate the PINN experiments.

- **DELL Technologies** — Cairo, Egypt
  Software Engineering Intern — *Sept. 2022 - Aug. 2023*
  - **AI4Research for Scientific Breakthroughs:** Replicated the Delphi paper and utilized the S2AG dataset to build a 10M-node citation graph. Developed temporal PageRank centrality to identify influential papers. Built a pipeline to parse paper concepts, problems, and solutions, formalizing the scientific discoveries problem as a link prediction problem.
  - **5G Digital Twin:** Utilized Graph Neural Networks to leverage network topology to reduce latency. Built and benchmarked datasets through NS3 simulation for the Lena dual stripe model.

- **Incorta** — Alexandria, Egypt
  Machine Learning Intern — *Jul. 2021 - Oct. 2021*
  - **Aspect-Based Sentiment Analysis:** Developed an aspect-based sentiment analysis model on clients' data.

Projects

- **Semester Master Course:** Developed a multi-granular text retrieval model for the Arabic language, supervised by Dr. Nagwa ELmakky and Dr. Marwan Torki.
- **Network and Database Course:** Developed a car rental website using Flask, PostgreSQL, and Bootstrap with live search and a multi-threaded C++ server supporting HTTP/2 pipelining and caching.
- **Blockchain and Cybersecurity Course:** Implemented smart contracts and decentralized apps, studied encryption and cryptographic systems for secure communication.
- **Computer Vision Course:** Implemented CV algorithms including segmentation, detection, and recognition to optimize performance.

Publications

- **SVRPBench: A Realistic Benchmark for Stochastic Vehicle Routing Problem:** Preprint - Submitted to NeurIPS conference 2025
- **First-Place Solution to NeurIPS 2024 Invisible Watermark Removal Challenge:** Published at ICLR 2025 workshop.

Awards and Achievements

- **Mediterranean Machine Learning Summer School (M2L) 2025:** Awarded full grant by Google DeepMind.
- **Eastern European Machine Learning Summer School (EEML) 2025:** Awarded full grant by Google DeepMind.
- **Rotman International Trading Competition:** Participating in the 21st edition of RITC in 2025
- **InSilico Medicine:** Ranked first in InSilico Drug Discovery hackathon.
- **NeurIPS Workshop:** Ranked first in both tracks, black box and beige box, for watermarking attacks in the "Erasing the Invisible" workshop. Developed a novel approach for watermarking classification, and diffusion models for adaptive attacks.
- **Google Hash Code:** Ranked in the top 5% in Hash Code 2022.
- **Mate ROV Competition:** 3rd place worldwide, 1st place in the Arab Regional, and awarded best ML team in 2021.
- **Microsoft Azure:** 1st team in Microsoft Azure ML ROV challenge 2021.

Certificates

- **McKinsey Forward Program:** Joined the McKinsey Forward Program, focused on leadership, problem-solving, and business skills development.
- **Entrepreneurship Workshop:** Participated in a 2-month entrepreneurship workshop organized by the Incubation and Entrepreneurship Center at MBZUAI.
- **Virginia Tech Hardware Design for Machine Learning Summer Training 2022:** Developed an FPGA accelerator (adopting a systolic array architecture) for accelerating convolutional neural network inference.
- **Lean Six Sigma - Yellow Belt:** Project on accelerating the research cycle.

Skills and Languages

- **Programming Languages:** C++/C, Python, JavaScript, Java, Matlab
- **Libraries:** OpenCV, PyTorch Geometric, TensorFlow, SpaCy, NS3, ROS, Catapult, Flask, PyQt, Graph-tools
- **Frameworks:** Ray, Dusk, Slurm, Flower, Flask
- **Database Systems:** MySQL, PostgreSQL
- **Miscellaneous:** gRPC, Protobuff, Socket Programming, Docker, Git, Web Scraping, Automation Scripts
- **Languages:** Arabic (Native), English (C1), French (C1), German (A1)

Extracurricular Activities

- **M.I.A Robotics** — Alexandria, Egypt
  Machine Learning Lead — *Aug. 2020 - Jul. 2022*
  - **Team Lead:** Trained and guided the 2022 ML team, developed models such as underwater object detection and segmentation, and autonomous UAV systems.
  - **ROV Control System:** Designed an adaptive 6 DOF controller for ROV using gain scheduling.
- **IEEE SSCS ALEXSC** — Alexandria, Egypt
  Head of Activities Team — *Oct. 2019 - Aug. 2021*
  - **Event Planning:** Planned events for undergraduates, including visits, tech summits, and competitions (Chipions) to bridge the gap between academia and industry.
- **Pro Athlete** — Egypt
  Short Distance Swimmer — *2020*
  - **National Competitor:** Competed professionally at national level until 2020.