# Mostafa M. ElAraby

Applied Scientist | Machine Learning, NLP & Generative AI

Montreal, Canada linkedin.com/in/mostafaelaraby

github.com/mostafaelaraby mostafaelaraby.com Google Scholar

## SUMMARY

Applied ML scientist (**PhD candidate, Mila / Université de Montréal**) specializing in **continual learning, OOD detection**, and model robustness, with publications at top-tier venues (**TMLR, AAAI, CRV**). Brings **5+ years** shipping production **neural machine translation** and conversational AI at scale (Microsoft Skype Translator & Bot Framework, Tensorgraph), plus deep expertise in **LLMs, transformers, NMT, and generative VLMs**.

## EXPERIENCE

**Research Intern** | *BuildCheck.ai* Oct 2025 – Jan 2026

- Designed and deployed a **production CV system** parsing architectural floor plans at **73.08% mIoU**, **outperforming SOTA Transformer baselines** on a long-tail benchmark.
- Engineered novel **loss functions** for extreme class imbalance, recovering rare structural elements baselines missed entirely.

**Research Intern** | *Precision AI (Mitacs)* May 2023 – Dec 2023

- Built a pixel- and instance-level **Out-of-Distribution (OOD) detection** framework on semantic-segmentation models, plus a **generalization-estimation testbed** (AUROC, FPR, Davies–Bouldin) for silent-failure, **model-quality monitoring** of deployed vision systems.
- Audited the production training repo, fixing critical upsampling bugs and shipping scalable improvements for edge deployments.

**Research Intern** | *Microsoft ATL, Cairo* Jun 2019 – Sep 2019

- Led an Arabic-to-English **dialectal NMT** system tuning **beam-search decoding**, and researched **transformer-based conversational context adaptation** for multi-turn dialogue.

**Senior R&D Engineer** | *Tensorgraph* Sep 2018 – Mar 2019

- Architected **retrieval-based conversational AI** for B2B automotive dealers, combining **Seq2Seq generation with retrieval ranking** for real-time response recommendation.
- Owned end-to-end **high-concurrency, low-latency** deployment with A/B comparisons and monitoring.

**R&D Engineer** | *Microsoft ATL, Cairo* Nov 2016 – Jun 2018

- Engineered a **data-augmentation pipeline** for Levantine NMT, raising BLEU from **25.03** → **27.91 (+2.88)** and **shipping to Skype Translator** (live June 2018).
- Shipped **multilingual support for Microsoft Bot Framework (V4)** via real-time translation middleware for global bot localization; published **Gender-Aware Spoken-Language Translation** (**+2.36 BLEU**, ICNLSP 2018).

## EDUCATION

**Ph.D., Computer Science** | *Mila / Université de Montréal* Fall 2020 – Jun 2026

Focus: **OOD detection, generalization, and uncertainty** in deep networks. **GPA: 4.3 / 4.3.** Advisor: Prof. Liam Paull (Mila).

**M.Sc., Computer Science (AI)** | *Mila / Université de Montréal* Winter 2019 – Aug 2020

Thesis: *Neural-network compression via Mixed-Integer Programming* (link). **GPA: 3.85 / 4.3.** Advisors: Prof. Margarida Carvalho (Mila) & Prof. Guy Wolf (Mila).

**B.S.E., Computer Science (Minor: Communications)** | *Alexandria University* Jun 2014

GPA: 3.3 / 4.0.

## SELECTED PUBLICATIONS

*Full list on Google Scholar.*

**GROOD: Gradient-Aware Out-of-Distribution Detection** TMLR, 2025

Post-hoc OOD detection via gradient-space class prototypes; **91.2% AUROC** (CIFAR-10 Near-OOD), **94.8%** (ImageNet-1k Far-OOD). [paper]

**Layerwise Early Stopping for Test-Time Adaptation** AAAI, 2025

LEAST prevents TTA overfitting by dynamically masking which layers to update via a cosine-distance criterion — no validation set needed. [paper]

**Background-Aware Continual Semantic Segmentation** CRV, 2024

Background-shift detector, custom loss, and continual transformer decoder for continual segmentation — **+3.6% mIoU** over prior SOTA. [paper]

## TECHNICAL SKILLS

- **LLMs & Generative AI:** Foundation models, post-training (SFT, RLHF/DPO), **PEFT / LoRA**, prompt engineering, RAG, transformers, generative Vision-Language Models (Qwen-VL).
- **NLP, Recommender & Retrieval:** NMT, Seq2Seq, tokenization, NER, text classification, multi-turn dialogue, dense / sparse retrieval, ranking, real-time response recommendation.
- **Core ML & Production:** Deep / Reinforcement Learning, Computer Vision, OOD detection, model compression; large-scale ML pipelines, MLOps, distributed training, Docker, AWS / GCP, **A/B testing**, low-latency inference.
- **Tooling & Languages:** PyTorch, TensorFlow, Hugging Face, scikit-learn, OpenCV, Git, Linux; **Python (expert)**, C++, Java, MATLAB, SQL, Bash.

## SELECTED PROJECTS, SERVICE & LANGUAGES

- **CVPR21 Continual Learning Challenge** (Jun 2021) — Team *Real-DEEL*, **6th place** (**0.87 CL score**) on the leaderboard; integrated into the Sequoia framework.
- **AI Driving Olympics — Lane Following** (Dec 2019) — DAgger imitation-learning baseline (sim+real video); selected as official baseline in the Duckietown gym.
- **Reviewer:** TMLR, ICLR, NeurIPS. **Languages:** Arabic (native) · English (IELTS 7.0) · French (DELF B2).