

AHMED OUMAR

Masdar ◇ Abu Dhabi, UAE

ahmed.oumar@mbzuai.ac.ae ◇ linkedin.com/in/ahmed-oumar ◇ Google Scholar

EDUCATION

PhD Student, NLP, MBZUAI Start 2025

Relevant Coursework: Advanced NLP, Multimodal Modeling, Interpretability of LLMs, Agentic AI.

Bachelor of Math and Computer Science, Cadi Ayyad University 2021

EXPERIENCE

Founder & CEO Dec 2025 - Present
DataBayt.AI *Abu Dhabi, UAE*

- Leading the development of open-source software and datasets for the AI community.
- Website: databayt.ai — GitHub: github.com/DataBayt-AI

Guest Researcher June 2025 - Aug 2025
UKP Lab - Darmstadt University *Darmstadt, Germany*

- Conducted research on the interpretability of Vision Language Models (VLMs).

NLP Research Fellow Aug 2024 - Present
Riken Advanced Intelligence Project *Japan*

- Working on interpretability of Large Language Models.

NLP Research Associate March 2023 - Aug 2025
MBZUAI *Abu Dhabi, UAE*

- Published several papers at top-tier conferences (ACL, EMNLP, NAACL).
- Focused on Generative AI for low-resource languages, Machine Translation, and dataset design.
- Investigated reasoning mechanisms and interpretability in Large Language Models (LLMs).

NLP Research Assistant Intern October 2022 - Feb 2023
Qatar Computing Research Institute (QCRI) *Remote, Morocco*

- Developed state-of-the-art propaganda detection systems for Arabic and multilingual settings.
- Achieved 2nd place in SemEval 2023 TASK 3.

NLP Engineer Intern February 2022 - June 2022
Walmart Global Tech *Remote, Morocco*

- Optimized production-scale LLMs for product type prediction.
- Achieved a 40-60% reduction in model size while maintaining performance.

SELECTED PUBLICATIONS

From Detection to Correction: Activation-Based Error Probes for VLM Counting (Under review)

- Ahmed Oumar El-Shangiti et al., 2026

The Geometry of Numerical Reasoning: Language Models Compare Numeric Properties in Linear Subspaces

- Ahmed Oumar El-Shangiti et al., NAACL 2025 (oral)

Arabic Automatic Story Generation with Large Language Models

- Ahmed Oumar El-Shangiti et al., ArabicNLP Conference 2024 (oral)

Palm: A Culturally Inclusive and Linguistically Diverse Dataset for Arabic LLMs

- Alwajih et al., ACL 2025
- *Best Resource Paper Award*

Number Representations in LLMs: A Computational Parallel to Human Perception

Casablanca: Data and Models for Multidialectal Arabic Speech Recognition

- Talafha et al., EMNLP 2024

Dolphin: A Challenging and Diverse Benchmark for Arabic NLG

- Nagoudi et al., Findings of EMNLP 2023

SKILLS, WORKSHOPS, AND AWARDS

Skills	Python, NumPy, Pandas, NLTK, scikit-learn, PyTorch, Hugging Face Transformers, Git/GitHub
Workshops	Generative AI in the Era of LLMs (I2COMSAPP April 2024), UBC NLP Workshop
Awards	Best Resource Paper Award , ACL 2025; Facebook Secure and Private AI Scholarship 2019; 1st in Arabic Misogyny Identification Shared Task; 2nd in SemEval 2023 Task 3
Languages	Arabic, English, French, German, Japanese, Urdu