

# ABDELRAHIM A. ELMADANY, PH.D.

Deep Learning and Natural Language Processing Group

University of British Columbia (UBC)

Vancouver, BC, Canada

☎ (+1) 604-442-4346

✉ [a.elmadany@ubc.ca](mailto:a.elmadany@ubc.ca)

🌐 [LinkedIn](#)

🎓 [Google Scholar](#)

## HIGHLIGHTS

---

- Versatile researcher-engineer with a track record of delivering high-impact, deadline-driven projects across academia *and* industry.
- Expert at framing open research questions, translating findings into production systems, and communicating results to technical, executive, and public audiences.
- Skilled mentor and collaborator: co-supervised graduate students, led multi-institutional grant teams, and fostered inclusive lab cultures that attract top talent.
- **5 + years** at the forefront of LLM innovation—pre-training, fine-tuning, and safety alignment of models (up to 13B parameters) used in both research and commercial applications.
- Secured competitive funding from the Gates Foundation, Google, and AMD; managed \$ > 1 M in compute and grant resources.
- Authored 30 + peer-reviewed publications in premier venues (ACL, EMNLP, EACL, INTERSPEECH, ACM CSCW) and released widely-adopted open-source datasets & toolkits.

## EDUCATION

---

### Ph.D. in Computer Science

**2013 – 2017**

*Cairo University, Egypt*

*4 years*

**Dissertation Title:** “Automatic Act Classification for Arabic Dialogue Context”

**Supervisors:** Prof. Mervat Gheith and Prof. Sherif Abdou

### Master’s Degree in Computer Science

**2007 – 2010**

*Cairo University, Egypt*

*3 years*

**Thesis Title:** “Building Relevance Judgments without Pooling”

**Supervisors:** Prof. Mahmoud Riad and Prof. Kareem Darwish

### Diploma in Computer Science

**2002 – 2003**

*Cairo University, Egypt*

*2 years*

## CURRENT RESEARCH INTERESTS

---

My research is primarily focused on advancing natural language processing (NLP) for low-resource languages, with a particular emphasis on African languages and Arabic. I am passionate about leveraging machine learning and deep learning techniques to improve access to language technologies in underrepresented linguistic communities.

- NLP for African & Arabic Languages
- Machine Translation (MT)
- Low-Resource Languages
- Dialogue Language Understanding
- Large Language Models (LLMs)
- Speech Processing & Speech Understanding

- Big Data & Scalable Learning
- Human-Computer Interaction
- Code-Switched
- Arabic Dialects

## EXPERIENCE

**Research Associate (NLP/ML/DL/LLM Research Scientist)** **Nov. 2024 – Present**  
*School of Information, UBC* *Vancouver, Canada*

- Co-leading the UBC research team in developing cutting-edge NLP and LLM solutions for low-resource African languages.
- Designing and building benchmarking platforms for African text, speech, and large language models (LLMs).
- Driving large-scale African NLP initiatives across diverse linguistic regions and tasks.
- Collaborating with interdisciplinary research teams across multiple countries and institutions.
- Funded by the Bill & Melinda Gates Foundation and CLEAR Global.

**Consultant — Researcher** **Jun. 2022 – Mar. 2024**  
*Almotaheda* *Cairo, Egypt (Remote)*

- Led the development of industry-grade natural language processing systems, focusing on Arabic dialects.
- Designed and implemented Optical Character Recognition (OCR) and Automatic Speech Recognition (ASR) systems tailored for multiple Arabic dialects.
- Trained and fine-tuned language and acoustic models for real-world deployment, with an emphasis on efficiency and accuracy.
- Managed an NLP engineering team, overseeing the full pipeline from data processing to model optimization for production-level integration.
- Collaborated with product and engineering teams to align model performance with industry requirements and deployment constraints.

**Postdoctoral Researcher** **Oct. 2019 – Oct. 2024**  
*School of Information, UBC* *Vancouver, Canada*

- Trained large-scale models such as BERT, RoBERTa, T5, GPT, Wav2Vec, and Data2Vec, achieving state-of-the-art performance on multiple language understanding and generation tasks through novel approaches.
- Analyzed and segmented billions of data points using Python and PySpark; automated data cleaning, validation, and de-duplication with custom Python scripts.
- Developed benchmarks for evaluating language models on natural language understanding (NLU) and generation (NLG) tasks.
- Built demos and tools to showcase the utility of trained models and to automate benchmark evaluation using Python, FastAPI, and PHP.
- Mentored and supervised undergraduate, MSc., and PhD students and engineers on research and development projects.
- Published over 30 research papers in top-tier NLP venues including ACL, EMNLP, EACL, and NAACL.
- Collaborated with industry partners and academic institutions on joint research initiatives.

**Senior NLP Consultant** **Sep. 2011 – Sep. 2013**  
*Research and Development International (RDI)* *Cairo, Egypt*

- Provided NLP expertise on core components such as search engines, data mining algorithms, and part-of-speech (POS) tagging systems.
- Contributed to the development and optimization of enterprise-scale NLP solutions across RDI's product offerings.
- Worked closely with cross-functional teams to support applied NLP use cases in commercial and research contexts.

## Lecturer

Sep. 2011 – Aug. 2019

*Applied Collage, Jazan University*

*Jazan, Saudi Arabia*

- Taught undergraduate courses including Programming Languages, Algorithms, Databases, Advanced Databases, Data Structures in C++, Object-Oriented Programming in C++, Web Programming, and Python.
- Mentored undergraduate students on capstone projects, programming assignments, and research skills development.
- Led departmental tasks including organizing teaching schedules, coordinating academic activities, and participating in curriculum development.
- Prepared course materials, exams, and grading rubrics to ensure instructional consistency and quality.
- Supported departmental accreditation efforts and engaged in faculty-level meetings and academic planning.

## Senior NLP Researcher

May 2008 – Aug. 2011

*Research and Development International (RDI)*

*Cairo, Egypt*

- Developed cutting-edge Arabic NLP applications including a search engine, diacritization system, morphological analyzer, and part-of-speech (POS) tagger.
- Led the design and implementation of scalable NLP pipelines for Arabic language processing in production environments.
- Collaborated with linguists and software engineers to integrate language-specific rules and resources into core NLP modules.
- Optimized algorithms for morphological analysis and POS tagging to improve performance on large-scale Arabic corpora.
- Conducted evaluation and benchmarking of NLP components using custom datasets and metrics tailored to Arabic morphology and syntax.
- Contributed to R&D publications and patent filings on Arabic language technology innovations.

## Technical Manager, *Akhbar El-Yom* Portal Project

Jan. 2008 – Apr. 2009

*Centre for Studies & Development of Computer & Information Systems*

*Cairo, Egypt*

- Directed the end-to-end design and deployment of the national *Akhbar El-Yom* news portal, coordinating a team of software engineers and designers.
- Integrated Arabic CMS, search, and analytics components; delivered a scalable platform with 24/7 uptime and real-time content ingestion.

## Research Engineer — e-Evaluation & Assessment Project

Jan. 2008 – April 2008

*Cairo Microsoft Innovation Center (CMIC)*

*Smart Village, Egypt*

- Developed an evaluation platform for graduation projects, streamlining submission and assessment workflows.

## NLP Research Engineer

Jun. 2006 – Jun. 2008

*Centre of Excellence for Data Mining & Computer Modeling (DMCM)*

*Cairo, Egypt*

- Built Arabic news topic-tracking and event-detection pipelines, combining IR techniques with statistical machine-learning models.

## PUBLICATIONS

---

2025

**\*Equal Contribution.**

- **Voice of a Continent: Mapping Africa's Speech Technology Frontier** . ArXiv (Submitted to EMNLP 2025)  
[A. Elmadany](#), S. Kwon, H. Toyin, A. Inciarte, H. Aldarmaki, M. Abdul-Mageed
- **Where Are We? Evaluating LLM Performance on African Languages**. ACL 2025  
I. Adebara, H. Toyin, N. Ghebremichael [A. Elmadany](#), M. Abdul-Mageed
- **Palm: A Culturally Inclusive and Linguistically Diverse Dataset for Arabic LLMs** . ACL 2025  
F. Alwajih, A. Mekki, S. Magdy, [A. Elmadany](#), Omer Nacar , ....., M. Abdul-Mageed

2024

- **Cheetah: Natural Language Generation for 517 African Languages**. ACL 2024  
I. Adebara\*, [A. Elmadany](#)\*, M. Abdul-Mageed
- **Benchmarking LLaMA-3 on Arabic Language Generation Tasks**. ArabicNLP 2024  
M. T. I. Khondaker, N. Naeem, F. L. Khan, [A. Elmadany](#), M. Abdul-Mageed
- **NADI 2024: The Fifth Nuanced Arabic Dialect Identification Shared Task**. ArabicNLP 2024  
M. Abdul-Mageed, A. Keleg, [A. Elmadany](#), C. Zhang, I. Hamed, W. Magdy, H. Bouamor, N. Habash
- **WojoodNER 2024: The Second Arabic Named Entity Recognition Shared Task**. ArabicNLP 2024  
M. Jarrar, N. Hamad, M. Khalilia, B. Talafha, [A. Elmadany](#), M. Abdul-Mageed
- **Toucan: Many-to-Many Translation for 150 African Language Pairs**. Findings of ACL 2024  
[A. Elmadany](#)\*, I. Adebara\*, M. Abdul-Mageed

2023

- **Octopus: A Multitask Model and Toolkit for Arabic Natural Language Generation**. ArabicNLP 2023  
[A. Elmadany](#)\*, E. M. B. Nagoudi\*, M. Abdul-Mageed\*
- **VoxArabica: A Robust Dialect-Aware Arabic Speech Recognition System**. ArabicNLP 2023  
A. Waheed, B. Talafha, P. Sullivan, [A. Elmadany](#), M. Abdul-Mageed
- **NADI 2023: The Fourth Nuanced Arabic Dialect Identification Shared Task**. ArabicNLP 2023  
M. Abdul-Mageed, [A. Elmadany](#), C. Zhang, E. M. B. Nagoudi, H. Bouamor, N. Habash
- **WojoodNER 2023: The First Arabic Named Entity Recognition Shared Task**. ArabicNLP 2023  
M. Jarrar *et al.*
- **JASMINE: Arabic GPT Models for Few-Shot Learning**. EMNLP 2023  
E. M. B. Nagoudi, M. Abdul-Mageed, [A. Elmadany](#), A. A. Inciarte, M. T. I. Khondaker
- **SERENGETI: Massively Multilingual Language Models for Africa**. Findings of ACL 2023  
I. Adebara\*, [A. Elmadany](#)\*, M. Abdul-Mageed, A. A. Inciarte
- **ORCA: A Challenging Benchmark for Arabic Language Understanding**. Findings of ACL 2023  
[A. Elmadany](#)\*, E. M. B. Nagoudi\*, M. Abdul-Mageed\*
- **Dolphin: A Challenging and Diverse Benchmark for Arabic NLG**. Findings of EMNLP 2023  
E. M. B. Nagoudi\*, [A. Elmadany](#)\*, A. O. El-Shangiti, M. Abdul-Mageed\*
- **UBC-DLNL at SemEval-2023 Task 12: Impact of Transfer Learning on African Sentiment Analysis**. SemEval 2023  
G. Bhatia, I. Adebara, [A. Elmadany](#), M. Abdul-Mageed

## 2022

- **AraT5: Text-to-Text Transformers for Arabic Language Generation.** ACL 2022  
E. M. B. Nagoudi\*, [A. Elmadany](#)\*, M. Abdul-Mageed\*
- **AfroLID: A Neural Language Identification Tool for African Languages.** EMNLP 2022  
I. Adebara\*, [A. Elmadany](#)\*, M. Abdul-Mageed, A. A. Inciarte
- **TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation.** OSACT-5 (LREC 2022)  
E. M. B. Nagoudi\*, [A. Elmadany](#)\*, M. Abdul-Mageed\*
- **A Benchmark Study of Contrastive Learning for Arabic Social Meaning.** WANLP 2022  
M. T. I. Khondaker, E. M. B. Nagoudi, [A. Elmadany](#), M. Abdul-Mageed, L. V. S. Lakshmanan
- **NADI 2022: The Third Nuanced Arabic Dialect Identification Shared Task.** WANLP 2022  
M. Abdul-Mageed, C. Zhang, [A. Elmadany](#), H. Bouamor, N. Habash

## 2021

- **ARBERT & MARBERT: Deep Bidirectional Transformers for Arabic.** ACL-IJCNLP 2021  
M. Abdul-Mageed\*, [A. Elmadany](#)\*, E. M. B. Nagoudi\*
- **Investigating Code-Mixed MSA–Egyptian → English Machine Translation.** CALCS @ NAACL 2021  
E. M. B. Nagoudi\*, [A. Elmadany](#)\*, M. Abdul-Mageed
- **Mega-COV: A Billion-Scale Dataset of 100 + Languages for COVID-19.** EACL 2021  
M. Abdul-Mageed, [A. Elmadany](#), E. M. B. Nagoudi, D. Pabbi, K. Verma, R. Lin
- **DiaLex: A Benchmark for Evaluating Multidialectal Arabic Word Embeddings.** WANLP 2021  
M. Abdul-Mageed, S. Elbassuoni, J. Doughman, [A. Elmadany](#) *et al.*
- **NADI 2021: The Second Nuanced Arabic Dialect Identification Shared Task.** WANLP 2021  
M. Abdul-Mageed, [A. Elmadany](#), H. Bouamor, N. Habash

## 2020

- **Toward Micro-Dialect Identification in Diaglossic and Code-Switched Environments.** EMNLP 2020  
[A. Elmadany](#), E. M. B. Nagoudi, M. Abdul-Mageed
- **Leveraging Affective Bidirectional Transformers for Offensive Language Detection.** OSACT-4 (LREC 2020)  
E. M. B. Nagoudi, [A. Elmadany](#), M. Abdul-Mageed
- **Machine Generation and Detection of Arabic Manipulated and Fake News.** WANLP 2020  
[A. Elmadany](#), E. M. B. Nagoudi, M. Abdul-Mageed
- **Holy tweets: Exploring the sharing of the Quran on Twitter.** CSCW 2020  
N. Abokhodair, [A. Elmadany](#), W. Magdy

## 2015 - 2019

- **Emoji use in Twitter white nationalism communication.** CSCW 2019  
L. Hagen, M. Falling, O. Lisnichenko, [A. Elmadany](#), P. Mehta, M. Abdul-Mageed, J. Costakis, T. E. Keller
- **Sentence-level bert and multi-task learning of age and gender in social media.** ArXiv 2019  
M. Abdul-Mageed, . Zhang, [A. Elmadany](#), A. Rajendran, M. Przystupa, L. Ungar Michael

- **Dianet: Bert and hierarchical attention multi-task learning of fine-grained dialect.** ArXiv 2019  
M. Abdul-Mageed, . Zhang, [A. Elmadany](#), A. Rajendran, L. Ungar
- **Arabic Tweet-Act: Speech Act Recognition for Arabic Asynchronous Conversations.** WANLP 2019  
B. Algotiml, [A. Elmadany](#), W. Magdy
- **Improving dialogue act classification for spontaneous Arabic speech and instant messages at utterance level.** LREC 2018  
[A. Elmadany](#), S. M. Abdou, M. Gheith
- **An Arabic speech-act and sentiment Corpus of Tweets.** OSACT4 2018  
[A. Elmadany](#), H. Mubarak, W. Magdy
- **Improving Dialogue Act Classification for Spontaneous Arabic Speech and Instant Messages.** LREC 2018  
[A. Elmadany](#), S. M. Abdou, M. Gheith
- **Building Relevance Judgments without Pooling: Relevance Feedback Clustering.** Book, LAP Lambert Academic Publishing 2015  
[A. Elmadany](#)
- **Turn Segmentation into Utterances for Arabic Spontaneous Dialogues and Instant Messages.** IJNLC 2015  
[A. Elmadany](#), S. M. Abdou, M. Gheith
- **JANA: A Human-Human Dialogues Corpus for Egyptian Dialect LDC2016T24.** LDC 2015  
[A. Elmadany](#), S. M. Abdou, M. Gheith
- **RECENT APPROACHES TO ARABIC DIALOGUE ACTS CLASSIFICATIONS.** NLP 2015  
[A. Elmadany](#), S. M. Abdou, M. Gheith
- **Towards Understanding Egyptian Arabic Dialogues.** arXiv 2015  
[A. Elmadany](#), S. M. Abdou, M. Gheith
- **Building a heterogeneous information retrieval collection of printed Arabic documents.** LREC 2006  
[A. Elmadany](#), N. Adly, K. Darwish, O. Emam, W. Magdy, M. Nagi

## RESEARCH FUNDING

---

**Bill & Melinda Gates Foundation & CLEAR Global** Sep. 2024 – Sep. 2025

- Developing a comprehensive benchmark for African NLP that unifies speech and text resources.

**Google Cloud TPU Research Credits** Jan. 2021 – Aug. 2023

- Trained and released a suite of large language models for Arabic and African languages—including ARBERT, MARBERT, AraT5, CHEETAH, and SERENGETI.

**AMD GPU Research Credits** Mar. 2021 – Jun. 2022

- Built high-performance machine-translation and code-switching models for Arabic using AMD GPU clusters.

## TEACHING & SUPERVISION

---

**Teaching Courses** 2011 – 2018

- Programming Languages & Algorithms
- Database Systems (Intro & Advanced)

- Data Structures in C++
- Object-Oriented Programming in C++
- Web Programming
- Python Programming

**Ph.D. Supervision, The University of British Columbia** **2022 – Present**  
*Vancouver, Canada*

- **Sang Yun Kwon** — *African NLP* (2025 – Present)
- **Ife Adebara** — *African NLP* (2022 – 2024)

**M.Sc. Supervision, School of Informatics, University of Edinburgh** **2017 – 2018**  
*Edinburgh, UK*

- **Bushra Algotiml** — *Speech-Act Classification in Arabic Tweets* (2017 – 2018)

## SERVICES

**Academic Services** **2011 – 2018**

- Led departmental tasks including organizing teaching schedules, coordinating academic activities, and participating in curriculum development.
- Served on curriculum committees, organized departmental seminars, and acted as faculty advisor for the annual student programming contest.
- Responsibilities included syllabus design, lecture delivery, lab coordination, assessment creation, and mentoring cohorts of 60–120 students per term.

**Conference/Journal Program Committee Member** **2011 – 2025**

- ARR (2022- Now) ACL (2017–2023), OSACT (2018–2022), LREC (2018–2022), EMNLP (2020–2023), ACL Rolling Review (2022–2024), WANLP (2020–2022), ArabicNLP 2023, and ACM/IEEE *Transactions on Audio, Speech and Language Processing*.

**Shared Task Organiser** **2021 – 2025**

- Dialect Identification (*NADI*, 2021–2025).
- Arabic NER (*Wojood*, 2023–2024).

## HONORS & AWARDS

**Outstanding Project Award** **May 2023**

- **Afrocentric-NLP**: Ranked among the global *Top 10* in the IRCAI *Top 100 Outstanding Projects*, issued under the auspices of UNESCO.

**Best Paper Award** **Jun. 2022**

- “*TURJUMAN: A Public Toolkit for Neural Arabic Machine Translation*,” honoured as Best Paper at the 5th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT5).

**Best Toolkit Award** **2018**

*Kuwait*

- *Holy Qur'an Automatic Integrity Checker*, the first of its kind to electronically check the typing of verses of Qur'an to the level of the symbols on letters. recognised as the Best Toolkit in a national competition, Kuwait.

## INVITED TALKS

---

- Presenter, *OSACT5* Workshop, 20 Jun 2022, Marseille, France (Remotely).
- Presenter, 11th edition of the *Language Resources and Evaluation Conference* (LREC 2018), 9 May 2018, Miyazaki, Japan (in-person).
- Presenter, *OSACT3* Workshop, 8 May 2018, Miyazaki, Japan (in-person).
- Presenter, “*Evaluation Criteria of Digital Qur’an Products*” Workshop, 29 Aug. 2015, Cairo, Egypt (in-person).
- Speaker, “*How to Use RDI Arabic NLP on Islamic Books (Qur’an, Hadith...)*”, Meeting at the Ministry of Awqaf & Islamic Affairs, 9 Aug. 2012, Kuwait (in-person).

## TECHNICAL SKILLS

---

- **Languages:** Python, C, C++, C#, SQL, HTML/CSS, JavaScript
- **Libraries & APIs:** PyTorch, TensorFlow, Hugging Face, scikit-learn, LLMs APIs
- **Big-Data Frameworks:** PySpark
- **Analytics & BI:** PANDAS, NumPy, Tableau
- **Model Engineering:** LLM pre-training / fine-tuning, Prompt Engineering, RLHF, Model Compression
- **Speech Tech:** Automatic Speech Recognition, Text-to-Speech Synthesis, Audio Classification
- **Cloud & DevOps:** Google Cloud Platform, Amazon SageMaker, Docker, Linux, Git

## REFERENCES

---

*References available upon request.*