

Mohamed Eldesouki

Curriculum Vitae (last update: Dec. 2024)

School of Engineering and Computer Science, Concordia University
Cairo, Egypt
0020 (100) 304-6958 - mohamed@eldesouki.com
<https://mohamed.eldesouki.com/>

SUMMARY

- **Highly skilled Software Engineer** with 13+ years in NLP, 9+ years in ML, 5+ years in IR & Recommendation Systems [Master's degree] , and 2+ years in Speech Recognition,
- **Scalable Systems:** Designed and deployed large-scale software, including microservices architectures, serving millions of users across industries,
- **Key Projects:** Built search engines, Live Arabic Dialects Identification of Arabic Speech , a deep learning library , and Arabic text processing tools (morphological analysis, NER, POS tagging, dialect identification, etc.), Question Answering,
- **Impact:** Enhanced recommendation accuracy and search efficiency for large-scale platforms,
- **Academic Achievements:** Published 17+ research articles in top-tier conferences and Journals in the domain of AI and NLP & IR. See [\[full list\]](#),
- **DevOps:** Skilled in CI/CD pipelines, microservices deployment, and tools like Kubernetes, Ansible, and Terraform,
- **Collaboration:** Led cross-functional teams to deliver end-to-end NLP solutions,
- **Programming:** Proficient in Python, TensorFlow, PyTorch, C/C++, Java, and modern software development practices.

EXPERIENCE

- Co-founder and Natural Language Processing Team Lead* Jan. 2023 - Present
Artitech Startup, (Istanbul, Türkiye - Cairo, Egypt)
projects: Majorayat platform (see details in the projects section)
- Teaching Assistant/Full-time PhD student* Sep. 2019 - Jan. 2023
School of Engineering and Computer Science, Concordia University, Montreal, Canada
projects: WISE
- Software Engineer/Research Associate* Dec. 2015 - Sep. 2019
Qatar Foundation (QCRI), Doha, Qatar
projects: Farasa, DialectID, Natasy, FarSpeech, ClassStrength
- Senior R&D Engineer* Jan. 2015 - Dec. 2015
OMS Company, Cairo, Egypt
projects: Buzzdiggr
- Senior Research Engineer* Aug. 2013 - Dec. 2014
Taya IT Company, Cairo, Egypt
projects: TAPS, Taya RecSys
- Research Engineer* Jul. 2012 - Aug. 2013
MGD Company, Cairo, Egypt
projects: MGD ITS
- Research Assistant* Jul. 2009 - Jun. 2012
Institute of Statistical Studies and Research (ISSR), Cairo University, Cairo, Egypt.

EDUCATION

Master (M.Sc) in Computer Science , Cairo University (Master by Research) Rank: Top of my class (with grade 82.2%) Specialized in Web Information Retrieval and Recommendation systems.	Jan. 2012
Postgraduate Diploma in Computer Science , Cairo University	May 2006
Bachelor of Computers and Information , Cairo University Majoring in: Information Systems	May 2003

ACHIEVEMENTS & RECOGNITION

- I have received a full scholarship from Gina Cody School of Engineering and Computer Science, Concordia University to pursue PhD in Computer Science.
 - I have been awarded a Concordia University International Tuition Award of Excellence, valued at approximately \$37,915.00.
 - Graduated the first of my class in the Masters program with a final grade of 82%, with a recommendation to complete the PhD.
 - Published 20+ research articles in top-tier conferences and Journals in the domain of Artificial Intelligence and Natural Language Processing (NLP). See [\[full list\]](#).
 - Developed an *Arabic dialect identification system* using features extracted from only the transcripts of a speech recognition system. Using this system, I achieved the **FIRST** place in accuracy and **THIRD** place in F1 among 18 participants in the DSL Shared Task 2016 of the VarDial 2016 workshop for Arabic dialect identification [\[LINK\]](#).
 - I obtained a scholarship to join *Wikimania Conference in Poland in 2010* [\[LINK\]](#) due to being an active Wikipedia member who have created and contributed in writing more than 1800 articles in both Arabic and English Wikipedia projects.
-

PUBLICATIONS

Full list of publications can be found at [Google Scholar](#).

PROJECTS

All the details of the following projects are available in [my website](#) with some demonstrations:

Majarayat Platform (try it <https://majaray.at>) Jan. 2023 - Present
I am the co-founder and lead engineer of Majarayat. Majarayat is a modern Arabic news platform that delivers news in a social media-friendly format. Each story is a concise summary generated by aggregating articles from over 500 news sources. Using cutting-edge Artificial Intelligence (AI) technologies, Majarayat fills in missing details to provide a comprehensive view of every story. Majarayat personalized news feed leverages a recommendation system that adapts to each user's preferences, interests, and location, ensuring they see only the most relevant stories.

I have have used cutting-edge Natural Language Processing (NLP) and Machine Learning (ML) algorithms to build the following: (1) Topic Detection and Tracking (to group same-event news articles together into a single story group): I have designed a simple time-text encoder based on entity-aware BERT to combine textual and temporal information to encode news articles before clustering them into groups. (2) Recommendation System (to rank news stories): tracking user interests and preferences by building a dynamic semantic network constructed from concepts and entities extracted from Arabic wikipedia articles. The semantic network changes to reflect the continuous user interests drifting over time. (3) Text Summarization: (to summarize the news articles of a single story group in to a concise summary) I have used off-the-shelf Large Language Models, i.e. Fanar and ChatGPT, and prompt engineering to generate the news summaries (4) Offensive Language Detection (for user comments and posts): using the Bidirectional Gated Recurrent Unit augmented with attention layer (Bi-GRU_ATT).

Technology used: Python, Django, FastAPI, Microservices, Redis, PyTorch, Scikit-learn.

DialectID Project ([\[source code\]](#) [\[demo paper\]](#)) Oct. 2017 - Sep. 2018

At Qatar Computing Research Institute (QCRI), I have worked in collaboration with [MIT CSAIL](#) on the DialectID project that aims to automatic dialect identification in Arabic Broadcast Speech into five Arabic dialects namely; Modern Standard Arabic (MSA), Egyptian dialect, Levantine dialect, Moroccan dialect, and Gulf dialect.

I built DialectID as an online live identification system for the Arabic speech. The key features of the system I built are: full duplex communication based on websockets, very scalable (+100K users), can do speech segmentation, supports Kaldi's GMM and "online DNN" models, and Python, Java, Javascript clients are available.

Technology used: Python, web sockets, Tornado framework, Kaldi toolkit, Tensorflow.

Farasa Project (try it <http://farasa.qcri.org/>) Jan. 2016 - Aug. 2019

Farasa is a fast and accurate text processing toolkit for Arabic text. Farasa can handle both Modern Standard Arabic (MSA) and the different Arabic dialects. Farasa can perform word segmentation, lemmatization, part-Of-speech tagging, text diacritization, dependency and constituency parsing, and spell checking and correction.

My task in Farasa was to improve Arabic Dialect Segmentation, and Part-Of-Speech tagging. Furthermore, I was responsible for providing annotated data. Using a deep neural architecture of Bidirectional LSTM-CRF, I achieved an average of 93.4% and 92.8% for both segmentation and POS tagging, respectively for 4 Arabic dialects + MSA. [\[more details\]](#)

Technology used: Python, Java, Django, Tensorflow, and Tomcat.

Natasy Deep Learning Lib ([\[source code\]](#) or `pip install natasy`) Jun 2017 - Sep 2017

A deep learning library designed and developed to be both easy to use and source code readable. It is a straightforward implementation of different algorithms and techniques of deep learning in Python. You can use it for small projects and/or educational purposes.

I am the maintainer of Natasy open-source project. I implemented the majority of the algorithms from scratch. Natasy supports fully-connected NN, RNN, CNN architectures. I provided implementation for most of the well know functions and algorithms (+25) for activation, initialization, optimization, dropout, and more.

Technology used: Python, and Numpy

Buzzdiggr project (<http://www.buzzdiggr.com/>) Jan. 2015 - Dec. 2015

A real-time monitoring platform that listens to social media platforms and the web for mentions of commercial brands then provides an array of powerful features for analysis.

I was responsible to transfer and maintain the state-of-the-art science and technologies in the field of topic detection, NER, sentiment analysis, and Arabic dialects identifications where, 1) I designed and implemented many parts of the aforementioned technologies 2) I have provided the dataset for building sentiment analyzer, NER, and dialect identification, and 3) I have conducted research building both sentiment analyzer and NER.

Technology used: Python, Java, Scikit-learn, Django, Spark Streaming.

Taya Arabic Processing Suite (TAPS) project Aug. 2013 - Dec. 2014

A multi-purpose Arabic processing port which facilitates Arabic language analysis through advanced tools and techniques, TAPS is designed to adapt and adjust to all search engines. Some of the key features of TAPS are Morphological Analyzes, Named-Entity Recognition (NER), Keyboard Layout Detection/Correction, Spell Correction, Language Identification, Dynamic Document Clustering, and Text Summarization.

I worked on TAPS project for around one year and half when I was a senior research engineer at Taya IT company with [Dr. Ossam Emam](#). I built the NER and the spellcheck error detection & correction, and auto-completion.

Technology used: Python, Java, CRF++, AMIRA.

MGD Intelligent Tutoring System Jul. 2012 - Aug. 2013

An Intelligent tutoring systems (ITS) that can assist human teachers identify the weakness of their students in different math skills (addition, subtraction, multiplications, etc) and then suggests a tailored strategic path of exercises to strengthen these skills for the students.

Technology used: C# .NET

TECHNOLOGY Proficient and familiar with a vast array of programming languages, concepts and technologies, including:

Programming Languages:

Proficient in Python, C++, Java, Prolog and Lisp and familiar with C# .NET and PHP.

Machine learning & Scientific packages; Tensorflow, PyTorch, Keras, Scikit-learn, Octave, NumPy, SciPy, matplotlib, Jupyter notebook, CRF++, YASMET, YamCha.

NLP & IR packages; NLTK, Indri (Lemur project), Solr (Lucene), FARASA toolkit, MADAMIRA, SRILM (SRI Language Model), Kaldi.

Databases: *Relational DBs;* MySQL, MS SQL server 2012 PostgreSQL. *NoSQL DBs;* MongoDB, Neo4j, Redis.

Technologies and Tools: *Web Platforms:* Django, Flask, and Java EE.

Web Frontend Technologies: HTML, XHTML, CSS, JavaScript, HTML DOM, Ajax, XML, XML DOM, JSON and Bootstrap framework, ReactJS.

Other tech.: Git, conda, virtual environment, virtual machines, Docker, Python Packaging.
OS Platforms: Linux, Windows, Mac.

**OTHER
ACTIVITIES**

- Reviewing for: ACL 2018, BJIT, RANLP 2017 <http://lml.bas.bg/ranlp2017/pc.php>,
- Teaching experience at both Concordia University, and Cairo University. At Cairo University, I had the responsibility to prepare and teach two courses, namely introduction to programming using Python and fundamentals of NLP.

REFERENCES Upon request