Muhammad Khalifa

33, Fostat St, Cairo, Egypt • (+20) 114 522 3109
Muhammad.e.khalifa@gmail.com

EDUCATION

Mansoura, Egypt **Mansoura University** **Fall 2011 – June 2016**

B.Sc. in Computer Systems Engineering.
Grade: 87.0%. (Distinct with Honors).
Rank: 8 / 119.

Cairo, Egypt **Cairo University** **Spring 2018 – Now**

Full-time MSc. in Computer Science.
Finished one year of masters-level studies. Excellent (90.1%) average grade in all courses.
Thesis Topic: *Transfer Learning for Natural Language Processing in Low-Resource Scenarios.*

PUBLICATIONS

- **Khalifa M**, Shaalan K. Character convolutions for Arabic Named Entity Recognition with Long Short-Term Memory Networks. Computer Speech & Language. 2019 Nov 1; 58: 335-46.
- **Khalifa M**, Hussein N. Ensemble Learning for Irony Detection in Arabic Tweets. Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. In: CEUR-WS.org, Kolkata, India, December 12-15.
- **Khalifa M**. Semantic Source Code Search: A Study of the Past and a Glimpse at the Future, *Arxiv Preprint.*

AWARDS

- Won the first place in the IDAT@Fire 2019 Shared Task (as YOLO team) for Irony Detection in Arabic Tweets among 18 competing teams.
- Won the first place in IEEE Code Door 2014 algorithmic problem-solving competition.
- Won 1st place Nasa's Space Apps Challenge 2016 edition in the Journey to Mars Challenge.

WORK EXPERIENCE

Research Intern (Full-Time) **NAVER Labs Europe** **05/2020 – Present**

- Doing a 6-months research internship in NAVER Labs Europe.
- Working with Marc Deymetman and Hady Elsahar on Controlled Text Generation with Deep Reinforcement Learning and Energy-based models.

ML Research Engineer (Part-Time) **Sypron Solution** **04/2018 – 2/2020**

- Research and Development of Anomaly Detection models for Predictive Maintenance.
- Built and tested multiple architectures for anomaly detection in time-series including LSTM classifiers and Autoencoder and Variational AutoEncoder (VAE) using *TensorFlow*.
- Deployed real-time anomaly detection models with *TensorFlow-Serving API* and *Flask*. Built and deployed a *Kafka* and *Spark Streaming* pipeline for stream processing of IoT sensor data.

Software Engineer (Full Stack) **CITC Mansoura University** **07/2017 - 10/2017**

- Worked as part of a team in CITC on Equipment's Maintenance System for Oncology Center of Mansoura University.
- Implemented main functions of the systems such as equipment's archiving and placing (HTML, CSS, Node JS and MySQL).

Teaching Assistant **Mansoura University** **1/2017 – 5/2017**

- Courses: Introduction to Computer Programming (Visual Basic) for freshman students.
- Handled guiding and evaluating students' assignments. Assisted with final exams grading.

PROJECTS

- **Fairseq-tagger:** Fairseq library adapted for sequence-labeling tasks (NER, POS Tagging, etc). Supports BPE, Finetuning pretrained models (Roberta, XLM-R) and logging seqeval metrics. *Link*

- **Transpoemer**: This is an ongoing side project that explores the generative ability of Transformer-Based Architectures with focus on poetry generation. Currently Implemented models: BERT. *Link*.

- **Fine-tunable:** a Keras-based library for Transfer Learning in NLP. Aims to enable pre-training a language model and fine-tuning it on a target task in just a few lines of code. *Link*

- Implementation of ULMFit training and fine-tuning scheme with PyTorch including Discriminative Fine-tuning and Slanted Learning rates. Trained on Arabic Wikipedia, fine-tuned on Arabic Dialect Identification AOC dataset. Got an accuracy of 80% on word-level LM and 82% on character-level LM (4-5% improvement over not using Transfer Learning). *Link*

- *Bilateral Multi Perspective Matching for Natural Language Sentences.* Implemented the paper model with PyTorch. Got a validation accuracy of 85% on Quora Question Pairs Dataset. *Link*

- Arabic to English machine translation with the Transformer model (6 self-attention blocks, 8 attention heads). Dataset used: OpenSubtitles v2018 (~262M tokens). *Link*

- Siamese CNNs for Duplicate Question Detection on Quora Question Pairs Dataset. Got an accuracy of 82% on a validation set.

- Arabic News Summarization with Seq2Seq model using PyTorch. Dataset used was SaudiNewsNet. Got good summaries but overall bad BLEU score due to the small size of the dataset. *Link*

- Movie Genre Prediction: Predicting movie genre from title using Naive Bayes and Support Vector Machines with bag of words model using word2vec. Got an f1-score of 63%. *Blog*

- Arabic Poetry Generation: Poetry generation model based on a word-level language model with TensorFlow using a 2-layer LSTM network. Corpus used: 10000 Arabic dialect poems scraped from *www.adab.com*.

COURSEWORK HIGHLIGHTS

- Natural Language Processing using Deep Learning (CS224n).
- Bayesian Methods for Machine Learning (Coursera).
- DS-GA 1012: Natural Language Understanding and Computational Semantics (New York University).