Muhammad Khalifa

PERSONAL DETAILS

Address	33, Fostat St. Cairo, Egypt, 17611
Mobile	(+20) 114 522 3109
E-Mail	muhammad.e.khalifa@gmail.com
Other	Github, LinkedIn, Blog

ACADEMIC QUALIFICATIONS

M.Sc Computer Science 2018 - 2021
Cairo University, Egypt
Supervisors: Dr. Aly Fahmy and Hesham Hassan.
Finished one year of masters-level studies with a GPA of 90.1%.
Thesis Title: *Transfer Learning for Natural Language Processing in Low-Resource Scenarios.*

B.Sc Computer Engineering, Distinct with Honors 2011 - 2016
Mansoura University, Mansoura, Egypt
Cumulative GPA: 87.06% (Excellent), Class Rank: 8/119
Graduation Project: *Educational Interactive 2D Game for Primary School Students.*

PUBLICATIONS, PREPRINTS, AND SUBMISSIONS

Muhammad Khalifa, Khaled Shaalan. Character convolutions for Arabic Named Entity Recognition with Long Short-Term Memory Networks. Computer Speech & Language. 2019 Nov 1; 58: 335-46.
Muhammad Khalifa, Noura Hussein. 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.
Muhammad Khalifa, Aminul Islam. Will Your Forthcoming Book be Successful? Predicting Book Success with CNN and Readability Scores. arXiv preprint arXiv:2007.11073.
Muhammad Khalifa, Hady Elsahar, Marc Deymetman. A Distributional Approach to Controlled Text Generation *(Submitted to ICLR 2021)*. OpenReview.net Link.
Muhammad Khalifa, Muhammad Abdulmageed, Khaled Shaalan. Self-Training Pre-Trained Language Models for Zero- and Few-Shot Multi-Dialectal Arabic Sequence Labeling *(Submitted to EACL 2021)*.
Muhammad Khalifa, Hesham Hassan, Aly Fahmy, Muhammad Abdulmageed, Khaled Shaalan. Self-Training Pre-Trained Language Models for Zero-Shot Multi-Dialectal Arabic Natural Language Understanding *(Submitted to ACM Transactions on Asian and Low-resource Language Processing)*.

PROFESSIONAL EXPERIENCE

Applied Scientist Intern Oct 2020 - Present
Amazon Inc.
Supervisors: Kathleen Mckeown and Miguel Ballesterous.
Topic: *Summarization of Movie Transcripts for Synopsis Generation.*

Research Intern May 2020 - Sep 2020
Naver Labs Europe
Supervisors: Marc Dymetman and Hady Elsahar.
Topic: *A Distributional Approach to Controlled Text Generation.*

Details: The goal of the work was to *impose moment constraints* on the text generated from pre-trained language models while minimizing the deviation from the original model. I designed and implemented the experiments, use cases, and comparisons with baselines. I participated in writing the manuscript, which was submitted to *ICLR 2021*. I also presented the work during the "intern-day" winning the Best Presentation Award.

Machine Learning Research Engineer (Part-Time) April 2018 - Feb 2020
Sypron Solutions

Manager: Dr. Alaa Khamis

Worked on the research and development of *Anomaly Detection in time-series models for a Predictive Maintenance system*. I investigated different architectures for anomaly detection in time-series including *Supervised classifiers and Unsupervused Autoencoders and Variational Autoencoders*. I also deployed real-time Anomaly Detection models to production with `TensorFlow-Serving` API and `Flask`. I developed a `Kafka` and `Spark Streaming` pipeline for stream processing of the IoT sensor data.

Teaching Assistant Dec 2016 - May 2017
Mansoura University

Taught Course: *Introduction to Computer Programming* for freshman students.
Handled guiding and evaluating students' assignments. Assisted with final exams grading.

AWARDS AND GRANTS

Best Presentation Award
Won the Best Presentation Award Intern Day during the Naver Labs Europe internship against with 8 other interns.

IDAT@Fire 2019 Shared Task
Won the first place in the IDAT@Fire 2019 Shared Task (as YOLO team) for Irony Detection in Arabic Tweets against 18 competing systems.

Best Assignment Paper Award
Won the best paper award for a short survey on Semantic Source Code Search submitted for the master's course on research methods given by Professor Amr Kamel.

1st Place at IEEE Code Door 2014
Won the first place in IEEE Code Door 2014 algorithmic problem-solving competition held at Mansoura University.

1st Place at NASA's Space Apps Challenge
Won 1st place NASA's Space Apps Challenge 2016 edition in the Journey to Mars Challenge.

SELECTED PROJECTS

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

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.

ULMFit Implementation
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).

BLMPM Implementation
Implemented the Bilateral Multi Perspective Matching for Natural Language Sentences paper with `PyTorch`*. Obtained a validation accuracy of 85% on Quora Question Pairs Dataset.*

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

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

COURSEWORK HIGHLIGHTS

- Advanced Deep Learning (Cairo University, By Prof. Aly Fahmy)
- Natural Language Processing using Deep Learning (CS224n).
- Bayesian Methods for Machine Learning (Coursera).
- DS-GA 1012: Natural Language Understanding and Computational Semantics (NYU)