Muhammad Khalifa

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

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 Muhammad**, Shaalan Khaled. Character convolutions for Arabic Named Entity Recognition with Long Short-Term Memory Networks. Computer Speech & Language. 2019 Nov 1; 58: 335-46.
- **Khalifa Muhammad**, Hussein Noura. 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.

PREPRINTS

- **Khalifa, Muhammad**, and Aminul Islam. "Will Your Forthcoming Book be Successful? Predicting Book Success with CNN and Readability Scores." arXiv preprint arXiv:2007.11073 (2020).
- **Khalifa, Muhammad**. "Semantic Source Code Search: A Study of the Past and a Glimpse at the Future." arXiv preprint arXiv:1908.06738 (2019).

AWARDS

- Intern Day *best presentation award* during Naver Labs Europe internship among 8 presentations.
- Won the first place in the IDAT@Fire 2019 Shared Task (as YOLO team) for Irony Detection in Arabic Tweets among 18 competing teams.
- *Best paper award* for a short survey submitted during a master's course on research methods given by Prof. Amr Kamel.
- 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

Applied Scientist Intern (Full-Time) **Amazon Inc.** **10/2020 – Present**

- Applied scientist intern in the Amazon Comprehend Team
- Working with Miguel Ballesteros on Unsupervised Summarization with Pre-trained LMs.

Research Intern (Full-Time) **NAVER Labs Europe** **05/2020 – 09/2020**

- Completed a 5-months research internship in NAVER Labs Europe.

- Working with Marc Dymetman and Hady Elsahar on Controlled Text Generation with Distributional Constraints and Energy-based models. Goal was to impose *moment constraints* on the text generated from pre-trained language models while minimizing the deviation from the original model.
- Designed and implemented the experiments, use cases, and comparisons with baselines. A paper titled "A Distributional Approach to Controlled Text Generation" was submitted to ICLR 2021.
- Presented the work during the "intern-day" winning *the best presentation award* among 8 presentations.

ML Research Engineer (Part-Time) **Sypron Solutions** **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** **01/2017 – 5/2017**

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

SELECTED 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*.

- 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.

COURSEWORK HIGHLIGHTS

- Advanced Deep Learning (Cairo Univrisrty, 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)