

Wissam Antoun

MACHINE LEARNING SCIENTIST · GRADUATE RESEARCHER

Beirut, Lebanon

☎ (+961) 71-778924 | ✉ wissam.antoun@gmail.com | 🏠 wissamantoun.github.io | 📺 [wissamantoun](https://www.linkedin.com/company/wissamantoun) | 🎓 [Wissam Antoun](#)

Summary

Artificial Intelligence and Machine Learning Engineer specialised in state-of-the-art Natural Language Processing techniques. During my masters, I had developed my NLP and published one of the most famous and powerful AI models for Arabic text representation. I also ranked in top places in almost all AI competitions that I participated in and was nominated for the Abdul Hadi Debs Endowment Award for Academic Excellence.

In addition, I have experience in teaching as a Graduate Teaching Assistant at the Eng. faculty at AUB and as an AI instructor Zaka. I joined Siren Analytics as Data Scientist where I worked on various NLP, IR and Anomaly Detection problems that directly impacts the Lebanese public sector. I'm now a Research Engineer in the ALMAnaCH team at Inria-Paris, continuing my research on enhancing text representation and model robustness to noisy text. I'm self-motivated, self-taught and a performer when working on tight deadlines. I'm also a tech geek who loves computer hardware and technology, and enjoys DIY projects and gaming.

Education

Masters Of Engineering In Electrical And Computer Engineering

Beirut, Lebanon

AMERICAN UNIVERSITY OF BEIRUT

Feb. 2018 - Sep. 2020

- Major Area: Artificial Intelligence and Machine Learning systems
- Minor Area: Software, Networking and Security
- Awarded Graduate Fellowship with full tuition coverage
- Nominated for the Abdul Hadi Debs Endowment Award for Academic Excellence

Bachelor Of Engineering In Computer And Communication Engineering

Zouk Mosbeh, Lebanon

NOTRE DAME UNIVERSITY - LOUAIZE

Sept. 2013 - Jul. 2017

- GPA: 3.58 Dean's List for 6 semesters

Work Experience

ALMAnaCH, Inria

Paris, France

RESEARCH ENGINEER

Mar. 2022 - Current

- Researching language models for languages displaying high variability, in particular Arabic dialects used on social media.
- Implementing new pre-training techniques for French language representation.

Siren Analytics

Beirut, Lebanon

MACHINE LEARNING SCIENTIST

Jan. 2021 - Feb. 2022

- Researched and developed semantic search engines using state-of-the-art NLP and IR approaches.
- Built an OCR API for ID card scanning.
- Designed and developed document AI solutions to extract information from scanned documents.
- Developed a real-time anomaly detection system for lockdown mobility permit.
- Developed encryption solutions for model IP protection.
- My work also involved researching and applying latest AI solutions, evaluating software solution vendors, and internship mentoring.

MIND Lab - American University of Beirut

Beirut, Lebanon

GRADUATE RESEARCH ASSISTANT

Sept. 2018 - Jan. 2021

- Co-founded AUB's Machine Intelligence Development (MIND) Lab <https://sites.aub.edu.lb/mindlab>
- My research focused on state-of-the-art NLP technologies for natural language processing, generation, and chatbots.
- Worked on pre-training various Arabic language models that ranges from RNN-based to the Transformer-based models, using Cloud TPU and GPU instances.
- Researched improving temporal and spatial resolution of remote sensing ML systems for soil moisture prediction.

ECE Dept. - American University of Beirut

Beirut, Lebanon

GRADUATE TEACHING ASSISTANT

Feb. 2018 - Jan. 2020

- EECE 435L: Software Tools Lab, Co-Instructor (3 semesters). Delivered a range of teaching and assessment activities including tutorials directed towards the delivery of modern Software Tools at the undergraduate level.
- EECE 696: Applied Parallel Programming, Lab instructor (1 semester). Delivered hands-on experience to graduate students on parallel programming and GPU computing using the CUDA C/C++ framework on local computers and on AWS.
- EECE 330: Data Structures and Algorithms, Lab instructor (1 semester). Delivered hands-on experience on fundamental algorithms and data structures used in software applications at the undergraduate level using C++.
- Introduced a plagiarism detection tool for computer code to the Electrical Engineering Department.

Academic Reviewer

REVIEWER FOR THE FOLLOWING JOURNALS:

2020 - Current

- Journal of King Saud University - Computer and Information Sciences (IF: 13.47)
- Expert Systems with Applications (IF: 6.95)
- Workshop on Arabic Natural Language Processing (h-index: 24, i10-index: 51)
- Workshop on Open-Source Arabic Corpora and Processing Tools
- ACM Transactions on Asian and Low-Resource Language Information Processing - TALLIP (IF: 1.42)
- Language Resources and Evaluation Journal (IF: 1.36)

Zaka

Online

AI INSTRUCTOR

May, 2020

- Gave a 4-part online workshop series titled *A Comprehensive NLP series* that was focused on delivering the major NLP concepts through hands-on coding examples along with the NLP fundamentals theory.

Stars of Science

Doha, Qatar

MACHINE LEARNING EXPERT

Feb. 2020 - Apr. 2020

- Provided Machine Learning and Artificial Intelligence solutions as part of the support team that helps contestants throughout their Stars of Science journey

Huawei Technologies (Lebanon) S.A.R.L.

Beirut, Lebanon

INTERN

June 2016 - Aug. 2016

- Learned about various LTE principles, base station hardware, and Microwave transmissions systems.

Honors & Awards

INTERNATIONAL

- | | | |
|------|---|-------------------|
| 2021 | First Place , Arabic Sentiment Analysis 2021 @ KAUST
(wti.kaust.edu.sa/solve/Arabic-Sentiment-Analysis-Challenge) | Kaggle, Online |
| 2020 | Second Place , OSACT4-Shared task on Offensive Language Detection | Marseille, France |
| 2019 | Best Domain Detection system , Qatar International Fake News Detection and Annotation Contest | Doha, Qatar |

DOMESTIC

- | | | |
|------|--|--------------|
| 2021 | Nominee , Abdul Hadi Debs Endowment Award for Academic Excellence | AUB, Lebanon |
| 2019 | Second Place , Mindsets-OLX Datathon, Best car price prediction score.(www.kaggle.com/c/mindsets) | AUB, Lebanon |
| 2019 | First place , Big Data, AI, and Media Hackathon Powered by Touch and Anghami | AUB, Lebanon |

Publications

My research metrics:

- **Citation score: 450**
- **h-index: 8**
- **i10-index: 8**

AraGPT2: Pre-Trained Transformer for Arabic Language Generation

Antoun, Wissam, Fady Baly, and Hazem Hajj

Proceedings of the Sixth Arabic Natural Language Processing Workshop co-located with EACL, Kyiv, Ukraine

2021

Cited by: 17

AraELECTRA: Pre-Training Text Discriminators for Arabic Language Understanding

Antoun, Wissam, Fady Baly, and Hazem Hajj

Proceedings of the Sixth Arabic Natural Language Processing Workshop co-located with EACL, Kyiv, Ukraine

2021

Cited by: 19

Empathetic BERT2BERT Conversational Model: Learning Arabic Language Generation with little data

Naous, Tarek, **Wissam Antoun**, Reem Mahmoud, and Hazem Hajj

Proceedings of the Sixth Arabic Natural Language Processing Workshop co-located with EACL, Kyiv, Ukraine

2021

Cited by: 3

AraBERT: Transformer-based Model for Arabic Language Understanding

Antoun, Wissam, Fady Baly, and Hazem Hajj

Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools co-located with LREC 2020, Marseille, France

2020

Cited by: 296

Multi-Task Learning using AraBert for Offensive Language Detection

Djandji, Marc, Fady Baly, **Wissam Antoun**, and Hazem Hajj

Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools co-located with LREC, Marseille, France

2020

Cited by: 20

State of the Art Models for Fake News Detection Tasks

Antoun, Wissam, Fady Baly, Rim Achour, Amir Hussein, and Hazem Hajj

2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT'2020), Doha, Qatar

2020

Cited by: 22

hULMonA: The Universal Language Model in Arabic

ElJundi, Obeida, **Wissam Antoun**, Nour El Droubi, Hazem Hajj, Wassim El-Hajj, and Khaled Shaban
Proceedings of the Fourth Arabic Natural Language Processing Workshop co-located with ACL, Florence, Italy

2019
Cited by: 36

Real-Time Fire Detection Using CUDA (Poster)

Antoun, Wissam and Manal Jalloul
ISC High Performance, Frankfurt, Germany

2019
Cited by: 1

A Comprehensive Approach for A Smart Medication Dispenser

Kassem, Abdallah, **Wissam Antoun**, Mustapha Hamad, and Chady El-Mou Cary
International Journal of Computing and Digital Systems, University of Bahrain

2019
Cited by: 15

Smart Medicine Dispenser (SMD)

Antoun, Wissam, Ali Abdo, Suleiman Al-Yaman, Abdallah Kassem, Mustapha Hamad, and Chady El-Mou Cary
2018 IEEE 4th Middle East Conference on Biomedical Engineering (MECBME), IEEE

2018
Cited by: 21

Presentations

AraBERT: pretraining, use cases and future improvements

Online

PYDATA KHOBAR AND KSU'S IWAN RESEARCH GROUP MEETUPS

May. 2020

- Introduced the AraBERT model for the Arab community and explained the pretraining and fine-tuning procedure

High Performance Computing for Lebanon (HPC4L) Brown Bag

AUB, Lebanon

PRESENTER FOR CURRENT AND FUTURE/POTENTIAL USE CASES FOR HPC4L

Feb. 2020

- Introduced the use case of HPC for pretraining transformer-based model for Arabic language Understanding
- Explained parallelization and scaling techniques for transformer models

Skills

Python TensorFlow PyTorch Transformers Haystack Detectron2 Elastic Stack Keras Docker FastAPI Rasa Bash
Git CUDA Parallel Programming C++ Matlab Java SQL LaTeX VSCode