Abdelrahman "Boda" Sadallah

NLP, MSc

📞 +971 543138923

✉ abdelrahman.sadallah@mbzuai.ac.ae

📍 Abu Dhabi, UAE

⌨ BodaSadalla98

in Boda Sadallah

🌐 Blog

PROFESSIONAL EXPERIENCE

Research & Development NLP Engineer
Oct '21 - Jul '23

Applied Innovation Center - Ministry of Communications and Information Technology
Cairo,Egypt

- Built Audio Normalization module, to normalize the output audio from the TTS system.
- Built and configured TorchServe server, to deploy our TTS pipeline. This decreased the latency by half and could decrease it more if we scaled up the number of workers.
- Developed a chatbot using Rasa Framework. The chatbot is designed to help farmers by knowing information about crops and giving automatic diagnoses for crop issues, and their solutions.
- Developed Distributed Training script for Tactoron2 in Pytorch, which works in multi-node, which helped accelerate the training process to the limit of our hardware capabilities.
- Trained a spell correction model to detect and correct the input text's spelling mistakes.
- Constructed a TTS preprocessing pipeline, starting from parsing numbers, dates, times, and honorifics, into text, to be ready for the speech synthesis system.
- Trained Univnet model on the LJ-Speech dataset, then fine-tuned it on our custom Arabic speech dataset. Switched from using Parallel-WaveGan vocoder to using Univnet Vocoder, which has better sound quality and 10x less inference time.
- Built a web demo to show and test our TTS system. Using Vue.JS as the front-end framework, and using Flask for the backend.

EDUCATION

Natural Language Processing, MSc
Aug '23 - Present

Mohamed Bin Zayed University of Artificial Intelligence
Abudhabi,UAE

Cum. GPA: 4.00

Bachelor of Engineering
Jul '16 - Jul '21

Ain Shams University
Cairo,Egypt

Major: Computer & Systems Engineering

Accumulative Grade: **Very Good with honors (81%)**

Graduation Project Grade: **Excellent**

PUBLICATIONS

- Are LLMs Good Cryptic Crossword Solvers? (Submitted for review at ACL 2024)
- ArabicMMLU: Assessing Massive Multitask Language Understanding in Arabic (Submitted for review at ACL 2024) Arxiv
- ANER: Arabic and Arabizi Named Entity Recognition using Transformer-Based Approach - *2023 Intelligent Methods, Systems, and Applications (IMSA)* Paper

PROJECTS

ANER Project (Graduation Project)
- Trained Named Entity Recognition Model for Arabic, which can recognize 50 different entities. It's a Fine-tuned model over Arabert (Arabic Bert Model trained for Masked Language task) for the NER task.
- ANER model is published on the HuggingFace Models repository.
- ANER Paper

Deep Learning Framework
- Participated In building a simple Deep Learning Framework from scratch using Python, that has the essential blocks to build deep learning models from layers, activation functions, loss functions, and evaluation metrics.

Inverted-Index
- Implemented an inverted index for a database of documents, so that we can get the content of these documents efficiently. To achieve high efficiency in both indexing and querying, I saved the index in a Trie, so that it would take O(N) to insert a new word into the Trie or to search for a word.

Mancala Game
- Participated in implementing a CLI Mancala game in Python.
- The game is human players vs AI-player. The AI player chooses the best move according to the Minimax algorithm given a certain difficulty.

* References can be provided upon request.