Emad Gohari

Senior Research Engineer — NLP & Machine Learning

E.Gohari.B@gmail.com · github.com/emadg · linkedin.com/in/emad-gohari

EXPERIENCE

Senior Research Engineer — Thomson Reuters   Jul 2021 - Present
Toronto, Canada

Delivering software components and ML Pipelines for experimentation & deployment of AI Models (ML/DL/LLM-based models) to improve customers' flagship products, and add new AI features related to NLP, search, knowledge graph, and recommendation problems.

NLP & Machine Learning Engineer — INAGO INC.   Jul 2018 - Jun 2021
Toronto, Canada

Deep neural network development, managing datasets, training, and deployment for Language Understanding Engines and Automated Text Generation Models. [Python, PyTorch, Word Embedding, AWS EC2, SpaCy, NLTK, BERT, T5 models]. Managing collaborative research projects with universities related to Automated Text Generation and Linguistics.

EDUCATION

M.Sc. in Computer Science — York University   2014 - 2018
Toronto, Canada · GPA: 8.17 / 9
Thesis: Interactive Question Answering Using Frame-based Knowledge Representation

B.Sc. in Computer Engineering — Amirkabir University of Technology   2010 - 2014
Tehran, Iran · GPA: 17.18 / 20

PROJECTS

GenAI Summary for Business Entity Reports   2024
Using LLMs prompting to generate summaries with customer-focused insights to facilitate the use of long business reports. Improving the current data extraction pipeline from reports.
Python · OpenAI GPT-4 · Claude Sonnet

Design and Building New Entity Matching System   2022 - 2024
Worked on new generation of flagship product for Due Diligence and Risk Management.
Spark · Python · Java · Scikit-Learn · AWS RDS · Opensearch · EMR · SageMaker

New Search Improvement Feature to Promote More Relevant Documents   2021
Delivered new search functionality to detect more general type of queries and questions and promoting the more suitable content for these queries for a Tax Research web application (in-production).
Python · Java · NLTK · Sentence Embedding · Elasticsearch · AWS

Improving Automated Question Generation from Documents   2020
Fine-tuning T5 Transformer Language Model, experimenting for improving model input, using BLEURT evaluation. Auto-generated questions resulted in 40% reduction in manual data curation efforts.
PyTorch · SpaCy · HuggingFace Transformers · Python

Improved Language Understanding Engine Models   2019
Training domain-specific Word2Vec word embeddings and adding detailed model testing for LSTMs to improve NLU model interpretability.
PyTorch · Python · Word2Vec · LSTM

Conversational Question Answering   2018
Creating a question answering dialogue system powered by syntactic and semantic analysis of documents and ontology generation. Domain-specific question answering using a dialogue interface. Part of a collaborative project with an industrial partner.
SpaCy · Python

PUBLICATIONS

Question-worthy sentence selection for question generation
Canadian AI 2020, 2020 · co-authored

Interactive Question Answering Using Frame-based Knowledge Representation
York University M.Sc. Thesis, 2018

Time aware topic based recommender system
Big Data & Information Analytics, 2016 · co-authored

A study on prediction of user's tendency toward purchases in websites based on behavior models
Information and Knowledge Technology (IKT), 6th Conference, IEEE, 2014 · co-authored

TECHNICAL SKILLS

Python · PyTorch · Scikit-Learn · Numpy · Pandas · SpaCy · NLTK · Transformers · Large Language Models · NLP · Machine Learning · Deep Learning · AWS · Docker · PostgreSQL · Elasticsearch · Opensearch · Java · Scala · Bash · Agile Development