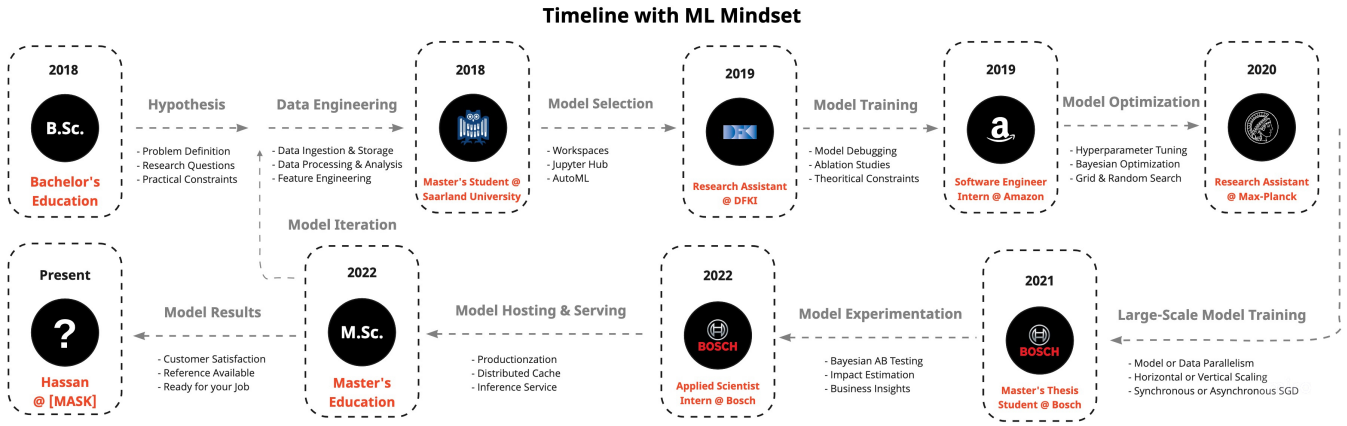
Hassan Soliman

Master's (M.Sc.) in CS · Saarbrücken, Germany

(+49) 1782554815 | hassanmahmoudsd@gmail.com | Web Page | Github | LinkedIn | Google Scholar

"Teaching machines to teach themselves"

Summary



Main Interests: Machine Learning, Deep Learning, Natural Language Processing, Natural Language Understanding, Information Extraction

Experience

Bosch Center for AI (BCAI)

Renningen, Germany

Applied Scientist Intern — *May. 2022 - Aug. 2022*

- Joined NLP & Semantic Reasoning group, and Utilized my master thesis output in Neural Entity Linking for a real use case in a top project.
- Refactored, tested & documented my code, and worked on training & evaluating the developed ML system on large-scale domain-specific data.
- Model performance is evaluated by comparing with the existing classical system, and achieved a **77%** end-to-end recall at the top 3 entities.

Master's Thesis Student — *Jun. 2021 - Jan. 2022*

- Joined NLP & Semantic Reasoning group, collaborated with research scientists & engineers, and worked on Cross-domain Neural Entity Linking.
- Utilized a state-of-the-art model to make it easy-to-extend-domain, and experimented which data & information sources are best for fine-tuning.
- Achieved an Average Precision gain of **9%** for the top-1 entity, and submitted a paper in RepL4NLP workshop at ACL 2022 venue.

Max Planck Institut for Informatics

Saarbrücken, Germany

Research Assistant — *Nov. 2020 - May. 2021*

- Joined Database & Information Systems group, and created a prototype for a model that finds diverse peer groups for an entity.
- Implemented a baseline model on the set expansion for an entity that exploits Wiki lists as a source of knowledge.
- Accomplished a **3x** faster runtime of the underlying algorithm using efficient sparse matrix multiplication approach.

Amazon

Luxembourg, Luxembourg

Software Development Engineer Intern — *Aug. 2019 - Feb. 2020*

- Joined Fulfillment Acceleration team, and owned a web application simulation tool by working on it as a Full-stack, using **AWS** cloud platform.
- Worked in an Agile environment as a system admin to maintain the server hosting team's tools, and acted as technical support for the team.
- Developed a tool for data visualization on maps, and provided workshops for the Business Analysts & Intelligent team members.

German Research Center for Artificial Intelligence (DFKI)

Saarbrücken, Germany

Research Assistant — *Jan. 2019 - Aug. 2019*

- Focused on a (QA) model, which formalizes answer planning to support fair dialogues under mixed motives. Model was already built & deployed.
- Contributed to the evaluation of the QA model, and implemented a model to detect fair dialogues using ensemble methods of decision trees.
- Helped to show quantitatively that **69.2%** of the dialogues by the QA model are perceived as fair, and reported the results in a draft paper.

Education

Saarland Informatics Campus, Saarland University

Saarbrücken, Germany

Master of Science in Computer Science (AI and Data Science Track) — *Oct. 2018 - Present*

- **Total GPA: 1.40 / 1.00** (Excellent) | Thesis: Cross-Domain Neural Entity Linking | Supervision: Prof. Dr. Dietrich Klakow — Transcript

Faculty of Engineering, Alexandria University

Alexandria, Egypt

Bachelor of Science in Computer and Communication Engineering — *Sep. 2013 - Sep. 2018*

- **Total GPA: 3.96 / 4.00** (Excellent) | Thesis: Egyptian Car License Plate Information Detection | Final Batch Ranking: 1st — Transcript

Publications

ACL 2022

Dublin, Ireland

The 7th Workshop on Representation Learning for NLP (RepL4NLP-2022) — *Video Presentation*

- **Hassan Soliman**, Heike Adel, Mohamed H. Gad-Elrab, Dragan Milchevski, and Jannik Strötgen. 2022. A Study on Entity Linking Across Domains: Which Data is Best for Fine-Tuning?. In Proceedings of the 7th Workshop on Representation Learning for NLP, ACL, 184–190, Dublin, Ireland

Skills

Selected Fields	Software Engineering, Machine Learning, Deep Learning, Natural Language Processing, Large-scale Language Modelling, Natural Language Understanding, Neural Information Extraction, Neural Machine Translation, Transformer-based Models
Technologies	Scikit-Learn, Keras, Pytorch, Tensorflow, Pandas, NumPy, Leaflet, Django, Streamlit, Flask, Jira, Git, Airflow, Kubernetes
Programming	Python, R, Java, C++/C (languages ordered from the most used to the least) | Linux Familiarity (Unix Operating Systems)
Web & DBs	PHP, Laravel, Java Spring, Javascript, jQuery, Bootstrap, Angular.js | SQL, MySQL, Redshift, DynamoDB
Languages	Native Arabic, Fluent English, Pre-Intermediate German

Preprints

Effective General-Domain Data Inclusion for Machine Translation by Vanilla Transformers

Python, Numpy

Master's College Project — *Paper*

- Built and trained a Transformer model architecture from scratch on the German-English translation task applications of WMT'13.
- Utilized a general-domain dataset from IWSLT'16 TED talks to help improve performance of the Transformer model, achieving a 25.8 BLEU score.

Offensive Language Detection & Classification on Twitter

Python, Classical ML, Sklearn

Master's College Project — *Paper*

- Trained a classifier to detect offensive tweets from Twitter using Support Vector Machines, after performing iterative experiments & analysis.
- Achieved a Binary Accuracy of 74% in classifying offensive tweets, and received the highest score among all participant teams.

Data Augmentation using Feature Generation for Volumetric Medical Images

Python, Tensorflow, Matplotlib

Master's College Project — *Paper*

- Proposed using U-net and ACGAN as a learning framework for feature generation of medical images of two complex types of brain tumors.
- Deployed a classifier pipeline to test & validate the quality of the generated features.

Theses

Cross-Domain Neural Entity Linking

Python, Pytorch, Transformer

Master's Thesis Project — *Thesis*

- Contributed to a single system that enables simultaneous linking of named entities to a general-domain KB (Wikipedia) & a domain-specific KB.
- Utilized Semantic Search by learning a joint vector space for these KBs from different domains using contextual-aware embeddings by BERT.
- Experimented with four different domain-specific KBs, and achieved an increase of up to 20% in Mean Average Precision for the top-10 entities.
- Helped in writing an invention report for my thesis contribution, which is submitted as a patent application, and received an Incentive-Prämie.

Egyptian Car License Plate Information Detection

Python, Tensorflow, CNN, Keras

Bachelor's Thesis Project — *Thesis*

- Implemented an application which extracts license information from car images in Egypt, using product life cycle stages.
- Collected datasets for various kinds of car plates in Egypt in various conditions, and applied different Data Transformation (ETL) techniques.
- Utilized pre-trained CNN models for fine-tuning for Object Detection, Localization, Semantic Segmentation, and OCR for the letters & numbers.

Projects

Better Diet to fight COVID-19

Python, R, Data Visualization

Master's College Project — *Report*

- Analysed food consumption from all countries to investigate a relationship between country food culture & their recovery rate, using Analytics.
- Implemented three Random Forest models and benchmarked the results to report the best one, and displayed figures using Seaborn library.

TF-IDF-based Information Retrieval

Python, TF-IDF, NLTK

Master's College Project — *Report*

- Built a system that takes an query as input and returns a certain number of relevant documents, using various statistics.
- Implemented a second step to retrieve and rank relevant sentences from the relevant documents, achieving a Mean Reciprocal Rank of 0.393.

Udacity Deep Learning Nanodegree

Python, Tensorflow, Keras, Pandas

Program Passing Projects — *Certificate*

- Studied and worked on 5 main scientific modules:- Vanilla NN, CNN, RNN, GAN, and Reinforcement Learning.
- Submitted related projects to each module with an excellent impact, which can be found on my online portfolio under the "Projects" section.

Accomplishments

2018 **Udacity Foundation**, Deep Learning Nanodegree — *Global Institute*

2018 **Alexandria University**, First Class Honor Degree — *Egypt*

Activities

Microsoft Student Partner Technical Club

Alexandria, Egypt

Technical & Media Member — *Dec. 2015 - Oct. 2018*

- Provided technical support & acted as a mentor during events e.g. Hackathons, and involved in marketing & media plans for events in the club.