

## Abdelrahman Abdallah

(+02) 01111371734 - [abdelrahmanelsayed@aun.edu.eg](mailto:abdelrahmanelsayed@aun.edu.eg) – [abdelrahman.abdallah@uibk.ac.at](mailto:abdelrahman.abdallah@uibk.ac.at) – [abdoelsayed2016@gmail.com](mailto:abdoelsayed2016@gmail.com)

Profile Links: [Personal website](#), [Github](#), [Assiut University](#), [University of Innsbruck](#) and [Linkedin](#),

---

### PROFESSIONAL SUMMARY

A machine learning engineer passionate about cutting-edge technology and solving real-world problems, with previous experience in computer vision and natural language processing, leading a lean team, and developing new products.

---

### RESEARCH INTERSETS

- Natural Language Processing:
  - Large Language Models (LLM): Extensive experience in training, fine-tuning, and applying LLMs for various NLP tasks, including working with Arabic data.
  - Information Retrieval: Leveraging LLMs for document reranking and improving search relevancy.
  - Open-Domain Question Answering: Developing and enhancing QA systems using state-of-the-art LLMs.
  - Keyword Information Extraction
  - Text Generation: Implementing LLMs for high-quality text generation in different domains, including creative and technical writing.
- Computer Vision:
  - Handwritten
  - OCR
  - Object detection
  - Generative Adversarial Network
  - Image Retrieval
  - Image Processing
  - Image segmentation

---

### SKILLS

- Data Science: Machine and Deep Learning, Computer Vision, NLP
- Programming Language: Python, PHP, Java
- Package: Tensorflow, Scikit-learn, Pytorch, Keras
- Web: Laravel, HTML & CSS, JQuery, JavaScript
- Tools: Pycharm, Anaconda, Jupyter notebook

---

### OPEN-SOURCE CONTRIBUTIONS

#### Rankify Library | Creator and Maintainer | 2024-Present

- Developed and maintain the Rankify library (<https://github.com/DataScienceUIBK/Rankify>) - a comprehensive Python package for information retrieval and reranking evaluation
- 500+ GitHub stars demonstrating significant community adoption and impact
- Supports evaluation of retrieval, reranking, and RAG systems with both automated metrics and human feedback
- Integrated with the RankArena platform for standardized IR evaluation

#### RankArena Platform | Lead Developer | 2024-Present

- Created and deployed RankArena (<https://rankarena.ngrok.io/>) - a unified evaluation platform for retrieval and reranking systems
- Accepted at CIKM 2025 as "Rankarena: A unified platform for evaluating retrieval, reranking and rag with human and llm feedback"
- Enables researchers to benchmark their models against state-of-the-art baselines with standardized evaluation protocols

---

### EXPERIENCE

- 10/2022 to now **Researcher Assistant, Digital Science Center ([DiSC](#))**  
**Universität Innsbruck– Innsbruck, Austria**
- knowledge extraction and information retrieval from unstructured text documents
  - Methods for Natural Language Processing and Information Retrieval
  - application of text mining methods to the field of digital history
- 08/2021 to 05/2025 **Machine Learning Engineering**  
**DISCO App– Cairo, Egypt**
- Working on Receipt extraction, OCR system, and NLP.
  - Built and improve the accuracy of the OCR system for receipt and extracting the information and classifying them.
  - The application is live on [Google Store](#).
- 01/2022 to 10/2022 **Machine Learning Researcher**  
**Università Ca' Foscari – Venezia, Italy**
- Preliminary comparative analysis between graph neural network and random forest models for climate change tasks.
  - Review paper on ML/AI models for risk assessments.
  - Survey on graph neural network for spatial-temporal data.
- 08/2021 to 06/2022 **Machine Learning Engineering**  
**KMG Engineering– Nur-Sultan, Kazakhstan**
- Working in inpainting images using GAN.
  - English Grammar Correction using deep learning.
  - Curve detection and track.
- 11/2019 to 06/2021 **Machine Learning Researcher**  
**National Open Research Laboratory for Information and Space Technologies – Satbayev University, Almaty, Kazakhstan**
- built the handwritten Kazakh, Russian database. The database can serve as a basis for research in handwriting recognition.
  - tried to describe various approaches and achievements of recent years in the development of handwritten recognition models about Cyrillic characters.
  - built the table detection database. The database can be used for machine learning and deep learning model.
  - building a new model to achieve a good result in table detection and classification.
- 07/2016 to 06/2019 **Research and Teaching Assistant**  
**Assiut University, Faculty of Computers and Information - Assuit, Egypt**
- Represented the team at meetings with executives and discussed project goals and milestones.
  - Kept abreast of emerging technologies, software, and trends and applied them to projects.
  - Teach classes, work with students in laboratories, grade papers, and projects or work directly for a professor.
- 06/2019 to 11/2019 **Software Developer**  
**Ccc at limkokwing university, - Cyberjaya, Malaysia**
- Developed and implemented a scanning component using MySQL, PHP.
  - Built databases and table structures following n-tier architecture methodology for web applications.
  - Designed and developed Using Expression Engine Framework in php.
- 01/2016 to 08/2017 **Web Developer**  
**FastKood Company – Cario, Egypt**
- Converted mockups into HTML, JavaScript, AJAX and JSON.

- Represented the team at meetings with executives and discussed project goals and milestones.
- Kept abreast of emerging technologies, software, and trends and applied them to projects.
- Hands-on experience using UNIX and Apache web servers.
- Developed data architecture design to enable analysts to perform targeted customer analysis.
- Developed work-flow charts and diagrams to ensure production team compliance with client deadlines.

06/2015 to 01/2016

## Software Developer

### ***Overcoffeesolutions, Assiut, Egypt - Assuit, Assuit***

- Developed object-oriented software.
- Designed intuitive graphical user interfaces using knowledge of serial communications and database design.
- Developed and implemented a scanning component using MySQL.
- Built databases and table structures following n-tier architecture methodology for web applications.
- Designed and developed 8 computer software web applications.

---

## EDUCATION

2022 to now

PhD of Science: Computer Science

Department of Computer Science –Innsbruck University

2019 to 2021

Master of Science: Data Science and Machine Learning

Faculty of Information and Telecommunication Technologies –Satbayev University

2011 to 2015

Bachelor of Science: Computer Science

Faculty of Computer and Information – Assuit University

2016 to 2017

Pre-Master of Science: Computer Science

Faculty of Computer and Information – Assuit University

---

## PUBLICATIONS

- DeAR: Dual-Stage Document Reranking with Reasoning Agents via LLM Distillation. [Accepted at EMNLP 2025.](#)
- How Good are LLM-based Rerankers? An Empirical Analysis of State-of-the-Art Reranking Models. [Accepted at EMNLP 2025.](#)
- ComplexTempQA: A Large-Scale Dataset for Complex Temporal Question Answering. [Accepted at EMNLP 2025.](#)
- Rankarena: A unified platform for evaluating retrieval, reranking and rag with human and llm feedback. [Accepted at CIKM 2025.](#)
- MultiOCR-QA: Dataset for Evaluating Robustness of LLMs in Question Answering on Multilingual OCR Texts. [Accepted at CIKM 2025.](#)
- Wrong Answers Can Also Be Useful: PlausibleQA - A Large-Scale QA Dataset with Answer Plausibility Scores. . [Accepted at SIGIR 2025](#)
- DynRank: Improving Passage Retrieval with Dynamic Zero-Shot Prompting Based on Question Classification. [Accepted at Coling 2025.](#)
- ASRank: Zero-Shot Re-Ranking with Answer Scent for Document Retrieval. [Accepted at NAACL.](#)
- CascadePLS-ViT: Cascade With Patch-Level Self-Supervised Vision Transformers for Breast Cancer Classification in Mammography. [Accepted at ISBI 2025](#)
- From Retrieval to Generation: Evaluating the Best Approach. [Under Review.](#)

- TempDPR: Temporal Dense Passage Retrieval for Explicit Temporal Questions. [Under Review](#).
- A Survey of Recent Approaches to Form Understanding in Scanned Documents. [Accepted Artificial Intelligence Review](#).
- Evaluating Temporal Robustness of Large Language Models. [Accepted at ACL 2025](#).
- HiGenQA: Exploring Hint Generation Approaches for Open Domain Question Answering. [Accepted EMNLP 2024](#).
- Detecting Temporal Ambiguity in Questions. [Accepted EMNLP 2024](#).
- IHRRB-DINO: Identifying High-Risk Regions of Breast Masses in Mammogram Images Using Data-Driven Instance Noise (DINO). [Accept at MICCAI2024](#).
- Arabicaqa: A comprehensive dataset for Arabic question answering. [Accepted at SIGIR 2024](#).
- Exploring the state of the art in legal QA systems. [J Big Data](#).

---

## LANGUAGE

- Arabic: Native
- English: Fluent ([Duolingo: 140](#))

---

## AWARDS

1. Scholarship to study for a master's at Satbayev University