

## Contact

Aalborg, Denmark.  
+45 52701019 (Mobile)  
ibraheemtaha@yahoo.com

[www.linkedin.com/in/ibraheemtaha](https://www.linkedin.com/in/ibraheemtaha)  
(LinkedIn)

## Top Skills

Machine Learning  
Python (Programming Language)  
C++

## Languages

English (Full Professional)  
Arabic (Native or Bilingual)  
German (Elementary)  
Danish (Elementary)

## Certifications

Cisco Certified Network Associate Security (CCNA)  
Cisco Certified Network Professional Routing and Switching (CCNP-RS)  
DEDS Winter School  
4th ACM Europe Summer School on Data Science  
Eleventh European Big Data Management & Analytics Summer School (eBISS 2023)

## Honors-Awards

Erasmus Mundus Exchange  
Erasmus Mundus Joint Masters Scholarship  
Erasmus Mundus Joint Masters Scholarship  
DEDS Scholarship

# Ibraheem Taha

Ph.D.AI & Data | Engineering | Science | Machine Learning | GenAI | AIOps.

Aalborg, North Denmark Region, Denmark

## Summary

+ .

AI/ML Engineer and Ph.D. candidate in Data Engineering (Aalborg University & University of Athens), specializing in data-intensive LLM applications, RAG architectures, and production-scale deep learning. My work blends research innovation with engineering rigor, delivering outcomes like 80x faster ML pipelines, 18x smaller vector indexes, and retrieval systems that power intelligent, context-aware applications.

I am into the intersection of AI, data engineering, ML engineering, and applied research, turning messy, high-volume data into reliable foundations for AI products. Whether it's building lakehouse/warehouse-ready pipelines, designing real-time feature and retrieval layers, or shipping LLM-powered agentic workflows, I focus on systems that are technically excellent and business-critical.

- Architect end-to-end data platforms and pipelines for analytics + ML (batch/streaming, reliability, observability)
- Design and deploy LLM-powered systems (RAG, embeddings, vector databases, agents)
- Build production ML pipelines with MLOps best practices (MLflow, Airflow, model monitoring)
- Optimize retrieval and inference systems for latency, cost, and scale
- Lead technical delivery from experimentation to production, architecture, evaluation, iteration

Python · SQL · PyTorch · MLflow · Airflow · Docker · SQL/NoSQL · AWS · Git

6+ years across research and industry (OpenAIRE, Orange Labs, Cadence Design Systems). My Ph.D. research on vector retrieval systems and intelligent table search directly supports modern LLM workflows and data discovery. Awards: 3× Erasmus Mundus scholarships + Marie Skłodowska-Curie fellowship.

Originally from Palestine, I've lived and worked across Europe and visited 30+ countries. I believe in building technology that matters and teams where people come first.

,

Open to Data Engineering / AI Platform / ML Engineering roles where I can combine strong data foundations with LLM-driven product impact, especially in LLM applications, retrieval, recommender systems, and personalization at scale.

ibraheemtaha@yahoo.com  
<https://ibraheemtaha.github.io/>

---

## Experience

"Athena" Research and Innovation Center  
Researcher in Data Engineering for Data Science  
February 2022 - Present (4 years)  
Athens, Attiki, Greece

✓ Researched data engineering and provided subject matter expertise through all aspects of data science while under investigation.

OpenAIRE AMKE  
Data Engineer  
February 2024 - May 2024 (4 months)

✓ Collect and curate datasets related to scholarly publications and their affiliations to academic and research institutions.

✓ Collaborate with the data science team to identify and close gaps in the knowledge graph, ensuring comprehensive coverage of global research output.

✓ Participate in the evaluation and adoption of new technologies and methodologies to improve data collection and processing workflows.

✓ Contribute to the development of tools and applications that utilize the knowledge graph to provide insights and analytics to the research community.

## Orange

Machine Learning Scientist

March 2020 - September 2020 (7 months)

France

- ✓ Implemented Smart Non-Linear Equalizers for Big Data Long-Haul Fiber Optics Transmission Systems; Acted as a go-to resource with data engineering/analysis, AI, ML, and deep learning to optimize performance and eliminate issues.

## Cadence Design Systems

Product Validation & Verification Engineer.

December 2016 - August 2018 (1 year 9 months)

- ✓ Noted for vital input to software development of the cutting-edge Xcelium parallel simulator. Received the best possible performance reviews and ranked among the top of all engineers.

## Google

Organizer (Google Developer Groups Nablus)

January 2015 - August 2016 (1 year 8 months)

- ✓ Advocating Google Technologies and its applications.
- ✓ Organizing events and workshops.

## IEEE

Student Member

April 2014 - August 2016 (2 years 5 months)

- ✓ IEEE Member.
- ✓ Computer Society Member.

## Microsoft

MSP - Microsoft Student Partner.

March 2015 - December 2015 (10 months)

- ✓ Organized sessions and workshops.
- ✓ Developed Windows apps.

## Isra' Software and Computer

.NET Developer Intern

May 2015 - August 2015 (4 months)

Nablus, West Bank

- ✓ Built best-in-class websites using ASP, increasing knowledge with C# and SQL databases.

---

## Education

Aalborg University

Doctor of Philosophy - PhD, Data Engineering · (June 2022 - June 2026)

National & Kapodistrian University of Athens

Doctor of Philosophy - PhD, Data Engineering · (February 2022 - February 2027)

Institut Polytechnique de Paris

Master's degree, Big Data Analytics & 5G - Optical Networks and Photonics Systems · (October 2018 - September 2020)

National & Kapodistrian University of Athens

Master's degree, Big Data Analytics & 5G · (October 2018 - September 2020)

An Najah National University

Bachelor's Degree, Computer Engineering · (June 2011 - August 2016)