

Ouael Ben Amara

+1 (734) 720 3785 | benamara@umich.edu | ouaelbenamara.com | [waelbenamara](https://www.linkedin.com/in/ouael-ben-amara-870234161/) |

<https://www.linkedin.com/in/ouael-ben-amara-870234161/>

PhD candidate specializing in **probabilistic databases** and **query execution engines**, with publications at premier A*/A-tier data management conferences including **SIGMOD 2024** (A* conference, awarded all three ACM reproducibility badges) and **EDBT 2022**. Current submission to **SIGMOD 2026**. Research focuses on relational probabilistic programming, **knowledge compilation**, query learning, and machine learning integration. With extensive expertise in applied machine learning systems, seeking full-time research positions focused on **fundamental machine learning research, optimization problems, low-level ML algorithms**, and **probabilistic programming**.

Experience

Research Assistant, The University of Michigan-Dearborn, USA

2020 - Present

- **StarfishDB Project:** Development of a query execution engine for relational probabilistic programming published at **SIGMOD 2024 (A* conference)**, earning all three ACM reproducibility badges (Available, Evaluated and Reusable, Reproducible). [Lab](#) | [GitHub](#)
- Designed and implemented novel algorithms for probabilistic query execution, achieving significant performance improvements over existing probabilistic programming systems.
- Developed knowledge compilation techniques for efficient probabilistic inference in relational databases, enabling scalable query processing over complex probabilistic models.
- Contributed to the definition and implementation of Gamma-Probabilistic databases for learning from query-answers (**EDBT 2022**, A-tier conference).
- Developed query learning methods using JIT-compilation techniques for probabilistic inference.
- Currently have a paper under review at SIGMOD 2026 on advanced query learning approaches and knowledge compilation for probabilistic databases.
- **Research Proposals:** Principally developed two research proposals addressing fundamental challenges in AI and database systems.
- Proposed "Explainable AI through Declarative Probabilistic Programming" framework combining De Finetti Logic with Datalog constraints to create transparent, mathematically rigorous explanations of ML decisions, addressing the recursive explainability problem in current XAI methods like LIME and SHAP.
- Proposed "Learning Adaptive Hierarchical Partitions for Efficient Large-Scale Vector Retrieval" using learned hierarchical tree indices with lightweight generative partitioning models, achieving 95%+ recall with logarithmic query complexity while enabling exact search and metadata filtering.
- **Independent Project:** Developed open-source Bayesian Point Cloud Registration exploring conjugacy and MCMC methods for 3D rigid transformations, achieving superior performance compared to standard ICP methods with uncertainty quantification. [GitHub](#)
- Conducted research on statistical assumptions underlying probabilistic models including LDA, HMM, and Word2Vec.
- Investigating integration of symbolic and subsymbolic reasoning for transparent and explainable AI systems.

NLP Engineer, Benten Technologies, Virginia, USA

May 2022 - Aug 2022

- Designed and developed a chatbot based on RASA framework.
- Developed NLU and Intent classifiers for Chatbot.
- Implemented Speech-to-text and text-to-speech modules.
- Participated in Mobile app development and embedded system design for smart speakers.
- Incorporated Blender Bot with conditional generation using HuggingFace.

Founder, DoVisual Memorality, Paris France

June 2023 - June 2025

- Co-founded a computer vision startup developing DoVisual, a platform enabling users to build machine learning pipelines using predefined nodes maintained and verified by expert agents and physical maintainers
- Specialized in deepfake detection, pose estimation, scene understanding, and object tracking
- Advanced image/media manipulation detection and unsupervised image clustering with vectorization and semantic search capabilities
- Won Arab Bank competition and MENA Chamber of Commerce pitching competition
- Selected for Station F founders program and participated in GAIA Saudi Arabia

Teacher Assistant, University of Michigan-Dearborn, MI, USA

Sept 2021 - Present

- **Main Lecturer for C++ Course:** Independently taught an entire C++ programming course, serving as the primary instructor with full responsibility for course delivery and content creation.
- Designed and developed comprehensive course materials including lecture slides, programming assignments, and practice problems.
- Created and graded exams, assessments, and project assignments for the C++ course.
- Graded database management system and software engineering courses, and taught software engineering tools lab.

Fintech Project Founder, KernelSnap, Tunis, Tunisia

June 2019 - 2022

- Developed automated identity verification process for Fintech companies using Keras, OpenCV.
- Integrated live processing OCR, Face Recognition and anomaly detection.
- Handled marketing and business development including pitching and client acquisition.

Machine Learning Intern, Slide Money, Tunis, Tunisia

June 2019 - Sept 2019

- Implemented Hidden Markov Model as a start model.
- Developed an asset trading bot based on RNN with integration of various metrics.

Publications and Contributions

January 2022	Gamma Probabilistic Databases: Learning from Exchangeable Query-Answers , Niccolò Meneghetti, Ouael Ben Amara EDBT 2022 (A-Tier Conference)	United States
2020	Speech Recognition for COVID-19 Keywords Using Machine Learning , Wael Ben Amara, Amani Touihri, Salma Hamza, International Journal of Scientific Research in Computer Science and Engineering, Vol.8, Issue.4, pp.51-57	Tunis
2020	Scikit Learn Bug Fix , Contribution to scikit-learn	United States
	StarfishDB: A Query Execution Engine for Relational Probabilistic Programming , Ouael Ben Amara , Sami Hadouaj, Niccolò Meneghetti SIGMOD 2024 (A* Conference) Awarded All Three ACM Reproducibility Badges GitHub	United States
2024	Learning From Query-Answers Using JIT-Compilation , Poster Presentation at Northeastern Database Day	Boston, USA
2026	Advanced Query Learning Methods for Probabilistic Databases, Under Review SIGMOD 2026	United States
2025	Reproducibility Report: StarfishDB: A Query Execution Engine for Relational Probabilistic Programming , Independent Validation ACM SIGMOD Reproducibility 2025	United States

Skills

Python, C, C++, Java, JavaScript, Rust

Proficiency in Programming for Research Applications

- Adept in **Python, C**, and **C++** for algorithmic solutions and low-level system programming.
- Engaged with Java and JavaScript for application development.
- Learning **Rust** focused on systems programming and memory safety.
- Applied **Pytorch, Keras, TensorFlow, OpenCV, sklearn, Scipy, PyMC3**, and Pyro in machine learning and data analysis.
- Experience with **CUDA programming** and parallel computing techniques.

Algorithmic and ML Techniques

Machine Learning and Statistical Techniques

- Experience with **Latent Dirichlet Allocation, Support Vector Machines, KNN, Decision Trees**, and **Reinforcement Learning**.
- Knowledge in **Variational Inference, Monte Carlo Markov Chain** (Gibbs sampling), Stochastic Processes, and Markovian Processes.
- Deep engagement with **probabilistic databases** and generative models like **Hidden Markov Models** and **Bayesian Networks**.

Software Engineering and Development

Foundational Skills in Software Engineering

- Involvement in **software refactoring**, CI-CD, testing, and version control using **GIT**.
- Competence in low-level debugging with **GDB** and **Valgrind**.
- Usage of cloud platforms like **GCP** and **AWS** for research computing.
- Experience in server/socket programming and foundational web/mobile app development.

NLP and AI Technologies

Engagement in NLP and AI for Research Purposes

- Application of RASA and NLTK in NLP research projects.
- Use of Generative AI Techniques and Large Language Models for innovative research.
- Implementation of speech-to-text and text-to-speech modules in AI systems.
- Work on natural language generation, including fine-tuning BlenderBot using **HuggingFace**.

Research in AI, Computer Vision, and Audio Processing

Involvement in Research and Development in Technology

- Participation in data analysis, AI problem-solving, and **computer vision** projects.
- Involvement in **audio processing**, statistical analysis, and empirical research methods.
- Contributions to open-source AI libraries and frameworks, including a bug fix in **Scikit Learn**.

LLM, RAG, Vector Databases, Semantic Search

Expertise in Large Language Models and Information Retrieval

- In-depth understanding of **Large Language Models** (LLMs) for various applications including natural language understanding and generation.
- Proficiency in **Retrieval-Augmented Generation** (RAG) for enhancing LLMs with external knowledge sources.
- Experience with **vector databases** for efficient storage and retrieval of high-dimensional data like embeddings.
- Engagement in **semantic search** technologies to improve information retrieval effectiveness by understanding query and document meaning.
- Implementation of **embedding generation** techniques for textual data, enhancing machine learning models' understanding of language nuances.
- Skilled in orchestration and re-ranking strategies for information retrieval, optimizing the relevance and precision of search results.

Education

3.9/4.0 CGPA	Masters of Science , <i>The University of Michigan - Dearborn</i> Dearborn, USA	Winter 2023
Expected 2026	Ph.D. Candidate in Computer Science , <i>The University of Michigan - Dearborn</i> Dearborn, USA Focus: Probabilistic Databases	2023 - Present
3.48/4.0 CGPA	Software Engineering Degree , <i>Mediterranean Institute of Technology</i> Tunisia	December 2022

Achievements: Awarded the Honors List in Fall 2017, Spring 2019, and Spring 2020; Dean's List in Spring 2018, Fall 2018, and Fall 2019.