Two (2) PhD student positions in the overarching research domain "Scalable Agentic Entity Resolution via Randomized Algorithms and Local LLM Inference," International Hellenic University, Thermi, Greece

The specific trajectory of the research for each successful candidate will be highly individualized. The core theme will be dynamically adapted and tailored to best align with the unique technical background, strengths, and specific academic interests of each successful candidate.

The International Hellenic University (IHU) is Greece's first public university where programmes are taught exclusively in English. The School of Science and Technology (SST) has been established to provide excellence in Technology Education and Research. Its activities cover research and postgraduate programmes on Information and Communication Technology (ICT) Systems and Energy Systems. The School's philosophy is to provide state-of-the-art multi-disciplinary education.

More info for the university: www.ihu.edu.gr

While Large Language Models (LLMs) have revolutionized semantic representations, their integration into foundational data management tasks like Entity Resolution (ER) is severely bottlenecked by computational overhead. Traditional scaling algorithms fail to scale when processing massive, high-dimensional vector spaces, disrupting high-throughput and streaming execution. This PhD research will investigate how sophisticated randomized algorithms, specialized data structures, and approximation schemes can bridge this gap. The core objective is to design scalable, resource-efficient ER architectures that natively index and match embeddings on commodity hardware without compromising theoretical distortion bounds.

Primary Research Axes:

- Semantic and Lexical Blocking/Meta-Blocking via Randomized Data Structures: Design dynamic randomized data structures that natively index quantized bi-encoder embeddings. The candidate will explore ways to minimize the memory footprint required for high-throughput execution.
- Active Learning coupled with LLM Inference: Develop active learning frameworks where local LLM inference engines (such as 4GB-optimized Gemma or Qwen deployments) act as weak oracles.

- Streaming Entity Resolution: Extend randomized blocking structures to handle continuous data streams. By mapping high-dimensional vector spaces into dynamic hash tables, the candidate will address real-time clustering and similarity search on evolving datasets without relying on static partitioning.
- Agentic ER: Investigate the deployment of autonomous LLM-driven agents to orchestrate and optimize the ER pipeline. The candidate will design multi-agent workflows where specialized, resource-efficient local agents autonomously reason over ambiguous records, dynamically adapt matching rules, and resolve complex edge cases without relying on monolithic, compute-heavy foundation models.

Methodology and Expected Deliverables:
The candidate will implement these pipelines using and advancing modern indexing structures and advanced vector quantization techniques. The final dissertation will deliver theoretical statistical bounds for the randomized algorithms alongside a unified, open-source framework validating the trade-off between semantic matching accuracy and execution scalability in low-resource environments.

The successful candidate is expected to produce high-impact, rigorous research output throughout the duration of their doctoral studies. This work should yield publications in top-tier peer-reviewed journals, including IEEE Transactions on Knowledge and Data Engineering, ACM Transactions on Knowledge Discovery from Data, Elsevier Information Systems, Elsevier Knowledge-Based Systems, and Springer Knowledge and Information Systems. Additionally, the candidate is expected to present their findings at premier international conferences in the data management and data mining communities, such as PVLDB, SIGMOD, ICDE, ICDM, EDBT, and CIKM.

Related academic references:

- A. Zeakis, G. Papadakis, D. Skoutas, M. Koubarakis, Pre-trained embeddings for entity resolution: an experimental analysis, PVLDB, 2023
- G. Papadakis, G. Koutrika, T. Palpanas, W. Nejdl, Meta-blocking: Taking entity resolutionto the next level, TKDE, 2013
- D. Karapiperis, C. Tjortjis, V. Verykios, Resource-Efficient Blocking: Optimizing the Trade-off between Effectiveness and Scalability in Entity Resolution, Information Systems, 2026.
- D. Karapiperis, L. Akritidis, P. Bozanis, The Impact of Fine-Tuning on Entity Resolution: An Experimental Evaluation, Knowledge-Based Systems, 2026
- D. Karapiperis, L. Akritidis, P. Bozanis, V. Verykios, ALER: An Active Learning Hybrid System for Efficient Entity Resolution, PVLDB, 2026
- D. Karapiperis, G. Papadakis, M. Koubarakis, V. Verykios, SPER: Accelerating Progressive Entity Resolution via Stochastic Bipartite Maximization, CIKM, 2026

Requirements:

The candidate should have:

a) Degree in one of the following: electrical or computer engineering, computer science/ informatics, science (maths, physics).
b) Master of Science in Data Science, Artificial Intelligence, Computer Science, Information Systems or similar.
c) Very good knowledge of written and spoken English (min C1 level) is a requirement for the position.

The School of Science and Technology will not charge fees for doctoral studies.

Applications: The application should include

a) Application form providing your personal details and the proposed subject of the doctoral thesis
b) Full research proposal giving the motivation, the objectives and the expected outcomes for the PhD research
c) A motivation letter with your research interests and the reasons for applying for this PhD position and experience (max. 2 pages). Please indicate also your earliest possible date for starting.
d) Copy of undergraduate and Master's degrees, along with transcripts.
e) Full CV, including information on the candidate's studies, research or professional activity, his/her scientific papers and interests
f) Very good knowledge of the English language (level C1)
g) Relevant publications (if any)
h) A copy of your MSc thesis (or a summary in case the thesis is ongoing).
i) Two reference letters

Applications should be sent via email to: infotech@ihu.edu.gr.

Information:

For more information about the post please contact Ass. Prof. Dimitrios Karapiperis: dkarapiperis@ihu.edu.gr

Starting: 01 Jan 2027 or as otherwise agreed.

Please submit your application no later than 15 Dec 2026.