SIWAR JENDOUBI

SENIOR AI RESEARCHER/DATA SCIENTIST, NLP, GEN AI

2 Avenue Florentine
92700, Colombes
+33 6 80 60 54 76
Jendoubi.siwar@yahoo.fr

PhD-level AI Data Scientist and Researcher with a decade of experience specializing in automated document intelligence and information extraction. Expert in designing advanced NLP systems, including RAG applications that automate the analysis and summarization of complex documents. Proven ability to transform unstructured text into actionable insights, demonstrated by developing a novel clustering algorithm that increased data processing efficiency by 50%. Proficient in Python, PyTorch, and FastAPI, with a strong publication record (20+ papers) and a focus on delivering practical solutions.

EXPERIENCES

Thesis Supervisor, Upgrad & GGU university, Remote, Part-time Job   *June2025-Now*

- Mentored Doctor of Business Administration (DBA) candidates through the complete dissertation development process, providing guidance on research methodology, academic writing, and business plan formulation.

Senior AI Researcher/Data Scientist, PhD, Upskills R&D, Paris   *Aug 2019-June2025*

- Designed and implemented a Retrieval-Augmented Generation (RAG) application to automate the extraction and summarization of key clauses from complex financial contracts.
- Deployed the solution as an API (FastAPI, vector database) with a React frontend, enabling business teams to query contracts in natural language.
- Developed and implemented a new LLM guided clustering algorithm that uses insights and recommendations provided by an LLM.
- Developed and implemented Natural Language Processing (NLP) and Deep Learning models for email classification, executing end-to-end projects from planning to package delivery with a success rate greater than 90% and achieving customer satisfaction.
- Mentored and managed interns to analyze customer emails using active learning and generative AI, improve email preprocessing using NLP, and finetune deep learning models through transfer learning.
- Mentored and managed a PhD student on a Generative AI project, he defended his thesis in January.
- Developed and implemented a clustering solution for trading transactions data to help the client making the Murex trading system reconciliation more efficient, resulting in increased reconciliation efficiency by about 50%.
- Planning project tasks, carrying out tasks, evaluating, testing, and validating solutions.
- Write detailed reports about accomplishments, present solutions/results, and write CIR reports.
- Write and publish scientific papers and participate to scientific conferences.
- Animation of internal training about AI and machine learning.

Postdoctoral researcher in Machine Learning, LISTIC Lab, Annecy   *Dec 2017 to May 2019*

- Build a multimodal tree species recognition approach from leaves and barks
- Increase the recognition rate by about 20% compared to the existing solutions
- Write and publish scientific papers and participate to scientific conferences.

PhD student researcher, IRISA Lab, LARODEC Lab, and CERT, Lannion and Tunis   *Jan 2013 to Dec 2016*

- Build two influence maximization models to select a small number of influencers who can maximize the propagation of a given marketing campaign through Twitter.
- Improve the quality of detected influencers compared to existing solutions.
- Write and publish scientific papers and participate to scientific conferences.

EDUCATION

PhD	**Computer Science**, joint supervision between University of Tunis and University of Rennes I and CERT company	*2013-2016*
	Dissertation: "Influencers characterization in a social network for viral marketing perspectives"	
MS	**Business Analysis**, Higher Institute of Management of Tunis	*2010-2012*
BS	**Business Analysis**, Higher Institute of Management of Tunis	*2006-2010*

SKILLS

Machine/Deep learning: Scikit-learn, PyTorch, Pandas, Langchain, Ollama, Haggingface
Natural language processing: NLTK, spaCy, and Gensim
Data visualization: plotly, matplotlib, seaborn
Model interpretation: LIME, SHAP
Programming languages: Python, Jupyter Notebooks, MATLAB and Java
Data Bases: SQL, PostgreSQL
Version Control: Git, Bibucket
Proficiency in Microsoft Office: Word, Excel, PowerPoint
CI/CD & MLOps: Bitbucket, Jira, Confluence
Cloud: AWS, Azure
Experience working in an agile environment.

LANGUAGES

English: Advanced levels in listening, speaking, reading, and writing
French: Distinguished levels in Listening, Speaking, Reading, and Writing
Arabic: Native Language

PUBLICATIONS & SOCIAL MEDIA

Published more than twenty research and peer reviewed papers and participated to different scientific conferences like IPMU, DaWaK, Dexa, etc. All publication references are available in this link: https://scholar.google.fr/citations?user=-oNKimAAAAAJ&hl=fr

Thesis: https://ged.univ-rennes1.fr/nuxeo/site/esupversions/7c413286-41e7-47ca-b6fe-9a115fa1852b?inline

LinkedIn: https://www.linkedin.com/in/siwar-jendoubi/

HOBBIES

- Travel and discover new cultures.
- Play Oud (Arabic fretless string music instrument).