

Soulaima Sellemi

AI Engineering Student – LLM Agents & Deep Learning | Seeking a PFE (end-of-studies) internship

📞 (+216) 29 732 985 📍 Manouba, Tunisia ✉️ soulaima.sellemi@ensi-uma.tn

🌐 [linkedin.com/in/sellemi-soulaima](https://www.linkedin.com/in/sellemi-soulaima) 🐙 github.com/soulaimasellemi

SUMMARY

Final-year software engineering student at ENSI, driven by turning promising AI research into tools people can actually trust. Two R&D internships at the Larodec lab: developed a multi-agent LLM pipeline for untangling commits and a hybrid CNN-Mamba model for skin lesion segmentation, both evaluated on real-world data. Comfortable from model to delivery: LangChain/LangGraph, RAG, Docker and Jenkins CI/CD..

EDUCATION

National School of Computer Science (ENSI), University of Manouba – Engineering Degree, Software Engineering 2024 – 2027
Preparatory Institute for Engineering Studies of Bizerte (IPEIB) – Mathematics & Physics Preparatory Classes 2022 – 2024

EXPERIENCE

AI R&D Intern – Larodec Laboratory, ENSI Summer 2026

Toward AI-Assisted Decomposition of Large Pull Requests into Reviewable Changes

- Evaluated Atomizer, a published multi-agent LLM framework for untangling commits (ICSE 2026), on real Gerrit history instead of synthetic benchmarks: 65.0% accuracy.
- Mined 937,327 Gerrit changes from 7 ecosystems into 142 validated change-split cases, each traced to the developer's words.
- Built a taxonomy of why and how developers split changes, with inter-rater agreement of $\kappa = 0.896$ and 0.769.
- Diagnosed over-splitting as the main cost driver and built a taxonomy-routed pipeline that reached 72.4% accuracy.

Deep Learning Intern – Larodec Laboratory, ENSI Summer 2025

AC-MambaSeg: Hybrid CNN-Mamba Model for Skin Lesion Segmentation

- Reproduced and extended AC-MambaSeg (CNN + Mamba state-space blocks, CBAM, attention gates) in PyTorch Lightning.
- Reached 97.0% Dice / 94.2% IoU on PH2 (CNN baseline: 95.0% / 91.0%) and 90.5% Dice on ISIC 2018, ahead of SM-UNet.
- Hair-removal preprocessing, AdamW and fine-tuning raised ISIC 2018 Dice from 85.3% to 90.5%.

PROJECTS

Commit History Q&A – Ask Questions About Any Git Repository 2026

Python, PyDriller, Tree-sitter, ONNX Runtime, rank-bm25, Groq, Streamlit, Jenkins

- Built a RAG system that extracts repository history with PyDriller and splits commits into function-level units with Tree-sitter.
- Labels unclear commits with gpt-oss-120b (Groq); hybrid search fuses all-MiniLM-L6-v2 embeddings (ONNX Runtime) with BM25.
- Writes answers that cite commit hashes, served in a Streamlit app delivered through a Jenkins CI/CD pipeline.

Explainable CV-Job Matching with Bias Audit 2025 – 2026

Llama 3.3 70B (Groq), GLiNER, spaCy, NetworkX, Angular, MongoDB

- Designed a 5-phase pipeline: job skill graph (GLiNER + LLM), evidence-based CV audit, rule-based scoring, LLM bias auditor, report.
- Compared Gemma 3 4B, Llama 3.1 8B and Llama 3.3 70B via LLM-as-a-judge on up to 500 CV-job pairs: 70B cut error from 25.7 to 8.2 points.
- Delivered as a web app with an Angular front end and MongoDB storage.

SKILLS

- **LLMs & Agents:** LangChain, LangGraph, multi-agent pipelines, RAG, hybrid search (BM25 + embeddings), LLM-as-a-judge, Llama 3, Groq, Ollama
- **Deep Learning & NLP:** PyTorch, PyTorch Lightning, TensorFlow, Transformers, ONNX Runtime, spaCy, GLiNER, OpenCV
- **Programming & Web:** Python, Java, C/C++, SQL, Streamlit, MySQL
- **DevOps & Tools:** Git, Docker, Linux, Jenkins, SonarQube, Maven, CI/CD pipelines, Tree-sitter, PyDriller, LaTeX
- **Languages:** Arabic (Native), French (C1), English (B2)

CERTIFICATIONS

NVIDIA – Fundamentals of Deep Learning; Computer Vision for Industry (2026) | Talan – Building GenAI-based Chatbots (2025)
Microsoft – AI Conduit VS Code (2025) | DeepLearning.AI – Advanced Learning Algorithms (Coursera)

ACTIVITIES

Treasurer, Executive Board – ENSI Robotics Association 2025 – 2026
Member – IEEE ENSI Student Branch, ENSI Junior Enterprise, ENSI Competitive Programming Club 2024 – 2025