

Yulin SHI

✉ eliane.shiyulin@gmail.com ☎ 0766082632 📍 Paris 🌐 yulin-shi 🔗 shi-yulin.com

EDUCATION

- 09/2022 – 09/2024
Paris, France
- Master's Degree in Computer Science (DAC), Sorbonne University**
- Specialized in Data, Learning, and Knowledge.
 - Courses: Databases, Statistical Learning, Machine Learning, Deep Learning, Reinforcement Learning, Optimization, Symbolic Learning
- 09/2019 – 06/2022
Paris, France
- Bachelor's Degree in Computer Science, Sorbonne University**

EXPERIENCE

- 02/2025 – present
Paris, France
- NLP engineer, Freelance**
- Develop a news aggregation site (Feed news [🔗](#))
- Multi-agent: Utilize multiple specialized agents to gather news from diverse sources and cluster articles by topic to generate coherent summaries.
 - Generate daily summaries tailored to users' preferences.
- LLM, Multi-agent systems, Information retrieval, Web development, Recommendation algorithms*
- 09/2024 – 12/2024
Paris, France
- NLP engineering, Freelance**
- Developed an advanced search platform for African legal documents (Legomnia [🔗](#))
 - Hybrid Search: Integrated hybrid search capabilities for efficient information retrieval.
 - Metadata Extraction: Automated extraction of document metadata for filtering purposes.
- Embeddings Models, Information retriever, Hybrid search, Metadatas extractions, Web development*
- 03/2024 – 08/2024
Paris, France
- Internship in AI/NLP, RTE**
- Detection of Equipment Failures from Operational Reports
 - Designed a complete NLP pipeline: data acquisition, model training, performance evaluation, and deployment.
 - Benchmarked models (CamemBERT, Llama, Vigotral, etc), fine-tuning for domain-specific tasks.
 - Created a user-friendly web interface for end-user interaction and report analysis.
- LLM, QLoRA, Finetuning, classification*
- 02/2023 – 05/2023
Paris, France
- NLP engineering, Freelance**
- Implemented methods from the article "Webpage text extraction using long short-term memory network based deep learning method" to extract the main content from HTML files.
 - Evaluated text length, number of punctuations, and content of each tag to identify the main content.
 - Achieved an F1 score of 87% using the LSTM model with FastText embedding.
- Embedding Model, LSTM, Web*
- 11/2022 – 02/2023
Paris, France
- Computer Vision Engineer, Freelance**
- Developed the a high-performance solution for recognizing album covers.
 - Utilized the SIFT algorithm to extract essential features from images. Integrated Faiss for rapid indexing and similarity search, enabling fast retrieval from large databases.
 - Achieved 95% accuracy with an inference time of just 0.03 seconds.

ACADEMIC PROJECTS

01/2023 – 06/2023	Reinforcement Learning in POMDPs, Sorbonne University This project explored the effect of LSTM neural networks in reinforcement learning, focusing on POMDP environments. Adjusted DQN and PPO with LSTMs, then evaluated them in a variant of the Cartpole game. <i>Reinforcement Learning, DQN, PPO, LSTM</i>
02/2022 – 06/2022	Next-generation blockchain PoC implementation - Implementation of a traditional blockchain and development of a new proposal where different types of information can coexist within the same blockchain in Java. - Link to the paper ↗

HACKATHONS

12/2024	4th Prize Winner, NASA Breath Diagnostics Challenge - Developed a classification model to analyze data from NASA E-Nose for distinguishing between COVID-positive and COVID-negative breath samples. - Applied advanced data preparation and AI techniques to optimize performance despite limited sample size. Link to the competition ↗ <i>Machine learning, Classification</i>
11/2024 Paris	Finalist in the Huawei Tech Arena Challenge 2024 – Optimization of a Fiber Optic Network - Designed an algorithmic solution in Java to optimize a fiber optic network, leveraging advanced graph theory concepts. - Developed algorithms to minimize costs and maximize installation efficiency in a simulated environment. <i>Java, Graph Theory, Algorithmic Optimization</i>
06/2024	First Prize Winner, Cryptocurrency Price Prediction Challenge - Conducted data analysis and preprocessing to ensure quality and relevance. - Optimized prediction accuracy through model selection and concatenation and developed a high-performance predictive model to capture market trends. <i>Machine learning, Decision tree</i>
05/2024 Paris	Finalist in the Mistral AI Paris Hackathon 2024: Development of a guitar Tablature Generation - We developed a model that generates guitar tablature and MIDI files from initial measures, desired style, and key. Using an open-source dataset of guitar recordings, we fine-tuned the Mistral-7B model for coherent musical continuations. This work was presented at the Mistral AI Paris Hackathon and Mistral Online Hackathon. - Link to our solution ↗ <i>LLM, Fine-Tuning, QLoRA, Quantization</i>

SKILLS

Python (PyTorch, langchain, tensorflow, sklearn, numpy, pandas), **Java**, **C++**, **SQL**, **LaTeX**, **Shell**, **Git**, **Cloud Platforms** (GCP), **Web development**

LANGUAGES

French	● ● ● ● ●	English	● ● ● ● ●
Chinese	● ● ● ● ●		