

# TAYYAB M.

📍 Nancy, France    📞 +33 7 52 92 75 42    ✉️ [hello@tayyab.io](mailto:hello@tayyab.io)  
🌐 [tayyab.io](https://tayyab.io)    🔗 [tayyab.io/in](https://tayyab.io/in)    🐙 [tayyab.io/github](https://tayyab.io/github)

## Profile

Final-year Master's (M2) student in NLP at Université de Lorraine building practical, reliable, and measurable LLM-powered systems. Experience in **LLM fine-tuning**, **RAG pipelines**, **HITL annotation workflows**, and **model interpretability** across the full ML lifecycle. Seeking a **6-month End-of-Studies Internship** (stage de fin d'études) in Applied Science / NLP Engineering. **Available early March 2026.**

## Education

**Université de Lorraine (IDMC)** Nancy, France  
*MSc in Natural Language Processing (M2)* 2024 – 2026

– Deep Learning for NLP, Generative AI (LLMs), Prompt Engineering, Information Retrieval

**University of the People** Pasadena, CA, USA  
*BSc in Computer Science – CGPA: 3.80/4.00* 2020 – 2023

– Focus: Data Science, Machine Learning, Software Engineering

## Experience

**Researchlytic** Remote  
*Creator & Full Stack Engineer* 2022 – 2024

- [Live Demo](#) | [Video Demo](#) | [Source Code](#)
- **Built a bibliometric analytics platform on top of OpenAlex** (250M+ works, 90M+ authors, 100K+ institutions) to explore research trends across institutions, countries, and time.
- **Achieved fast UX and analytics performance:** Lighthouse score **96**, **0.7s FCP**, **0.8s LCP**; dashboard queries compute metrics and render results in under **3 seconds**.
- Developed ranking pipelines combining **publication volume**, **citation impact (SCR/SCRIP)**, and **open-access signals** to identify emerging high-impact research domains.

## Projects

**AnnotaLoop** — LLM-Driven Document Annotation with HITL | *M2 Software Project*

- [Project Website](#) | [Video Demo](#) | [Source Code](#)
- Reduced manual annotation time from **approx. 30 min to approx. 5 min per document (83% reduction)** using **under 1 minute** LLM-generated suggestions with human-in-the-loop review, combining LLM efficiency with human validation to produce **structured labeled data at scale**.
- Built a **cross-platform desktop app** (macOS, Windows, Linux) integrating **Mistral, OpenAI, Anthropic, Gemini, Ollama, and LM Studio** with batch processing, offline-first privacy, and exports to **JSON, CSV, and annotated PDFs**.

**Intrinsic Evaluation of French Word Embeddings** | *M2 Supervised Research*

- [Live Demo](#) | [Video Demo](#) | [Source Code](#)
- **Analyzed grammatical gender encoding across 33K+ French nouns and adjectives** using FlauBERT embeddings and attribution methods (**SHAP, LIME, RF, EKA**).
- **Demonstrated distributed (non-localized) encoding:** achieved **49–85% accuracy with only 1% of dimensions** and enabled up to **75% dimensionality reduction with <5% accuracy loss**.
- **Ran 16 configurations** (4 models × 2 classifiers × 2 tasks) with peak **99% adjective accuracy** and strong within-model convergence in **15/16 configurations at 5% threshold**.

📁 **View portfolio:** Project overviews, live demos and video demos at [tayyab.io](https://tayyab.io)

## Technical Skills

**GenAI & LLMs:** LLM Fine-tuning, RAG Pipelines, LangChain, Prompt Engineering, BERT, LLaMA, GPT APIs  
**ML/NLP:** PyTorch, Hugging Face, scikit-learn, spaCy, SHAP, LIME  
**Engineering:** Python, SQL, PHP, JavaScript, Docker, Git/GitHub, AWS (S3), Linux  
**Vector DBs:** FAISS, Pinecone, Chroma  
**Languages:** English (Fluent), Urdu/Punjabi (Native), French (Basic)