Wenyu Li
www.linkedin.com/in/wenyuli020303
Email : wenyu.li.6753@student.uu.se
Mobile : +46 734801520

Education

- **Uppsala University** — Uppsala, Sweden
  Master's degree, Language Technology; Uppsala University International Scholarship — *Sept. 2024 – Jun. 2026*
- **Guangdong University of Foreign Studies** — Guangzhou, China
  Bachelor's degree, Urdu; GPA: 3.79/4.00 — *Sept. 2020 – Jun. 2024*

Experience

- **Shandong Keli Digital Intelligence Technology Co., Ltd.** — Zibo, China
  NLP Engineer Intern — *Jun. 2025 - Aug. 2025*
  - Built data preprocessing pipelines (cleaning, normalization, sampling) that improved downstream model stability by 15% fewer invalid samples.
  - Fine-tuned BERT-based classification models, achieving measurable accuracy improvements for AI-generated text detection tasks.
  - Designed evaluation workflows (F1, A/B testing) to validate algorithm performance in real-world settings.
  - Processed large cross-domain corpora (law, literature, engineering), producing high-quality datasets for model training.
  - Proposed model optimization strategies after reviewing SOTA literature, resulting in more robust detection performance.
- **Kuaishou Technology** — Guangzhou, China
  International Project Intern — *Sept. 2022 - Mar. 2023*
  - Collaborated with multimodal R&D teams to optimize annotation guidelines for OCR, speech, and image understanding tasks.
  - Executed systematic data quality checks using sampling & Python scripting, reducing annotation errors by 20%+.
  - Supported data pipelines by reviewing edge cases and identifying patterns affecting model performance.

Projects

- **Gradient-based vs. Example-based Explainability in Text Classification**
  Individual project
  - Compared Integrated Gradients, LIME, SHAP, Attention on BERT sentiment classification.
  - Implemented token-level contribution aggregation & designed faithfulness tests (deletion metrics).
  - Delivered explanation visualizations, evaluation tables, and reproducible code.
- **Multilingual Information Retrieval & Dense Vector Analysis (SBERT vs. BM25)**
  Group project
  - Conducted a comparative study of sparse (BM25, TF-IDF) and dense (SBERT) retrieval algorithms on a Croatian dataset (2,100+ docs).
  - Built a CLASSLA preprocessing pipeline for morphologically rich text, gaining insights applicable to complex document chunking.
  - Assessed performance using ranking metrics, establishing a strong foundation for defining Recall/Precision in RAG systems.

Programming Skills

- **Programming & Tools**: Python (**NumPy**, **Pandas**, **scikit-learn**, **PyTorch**), Linux, Pycharm, Jupyter
- **NLP & LLMs**: Text preprocessing, multilingual processing, **HuggingFace Transformers**, **BERT**, SBERT, RAG concepts, embedding models, QA systems, explainability (IG, LIME, SHAP), API-based LLM interaction
- **Information Retrieval**: **BM25**, TF-IDF, vector embeddings, ranking evaluation (Recall@k, Precision, F1), document indexing, chunking strategies