

KAPIL RAJESH KAVITHA

First-Year Master's Student @ MBZUAI

📞 +971 542217325 📩 rkkapil2004@gmail.com 💬 kapil-rajesh-kavitha 🌐 kapilrk-04 🌐 kapilrk-04.github.io

Education

Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI) 08/2025 – Present
Master of Science (M.Sc) in Natural Language Processing | CGPA: 3.75/4

International Institute of Information Technology (IIIT Hyderabad) 11/2021 – 07/2025
Bachelor of Technology (B.Tech) in Computer Science and Engineering with Honors | CGPA: 8.24/10

Experience

Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI) 10/2025 – Present
Graduate Researcher *LLMs, vLLM, Evaluation, Bias & Safety*

- **Fairness Evaluation of LLMs in Indic Contexts** Benchmarked 20 open- and closed-weight multilingual LLMs on a subset of the Indic-Bias dataset to evaluate fairness across English and Indian regional-language scripts, analyzing bias magnitude, stereotypical response rates, and refusal-to-answer rates.

Athlyte Inc 05/2024 – 01/2025
Data Science Intern *LLMs, Python, MongoDB, AWS*

- **LLM-Powered Query Generator** Designed and implemented a system leveraging LLMs for tweet classification, entity extraction and structured query generation.
- **Robust Table Detection System** Developed a system for detecting, extracting, and analyzing tables from images using YOLO object detection and AWS Textract, achieving 80% accuracy in table extraction.

Machine Translation and NLP Lab (MT-NLP), IIIT Hyderabad 05/2023 – 05/2025
Researcher (Honors) *Transformers, PyTorch, Hugging Face, LLMs*

- **Encoder-Decoder Models for Code-Mixed MT** Developed and evaluated encoder-decoder models for code-mixed machine translation (MT), focusing on finetuning scripts, evaluation frameworks, and quality control impacts.
- **Indic LLMs for English-Hindi Code-Mixing** Conducted research on usage of Indic LLMs for English-Hindi code-mixing, including zero-shot evaluations, LoRA-based pretraining, and multiway translation approaches.

IIIT Hyderabad 07/2023 – 12/2024
Teaching Assistant

- **Advanced NLP, Introduction to Software Systems** Responsible for creating assignments, conducting tutorials, and grading exams for courses with 200+ students.

Projects

Football-QA | 🌐 | Python, SQLite, LLMs, Tool Calling, Gradio

- Built an LLM-powered QA agent that uses tool calling and multi-step planning to answer natural-language football queries over a local SQLite database.
- Implemented modular tools for data ingestion, structured queries, and statistical analysis, exposed via a Gradio UI and deployed on Hugging Face Spaces.

Early-Exit Transformers for Adaptive Computation | 🌐 | Transformers, PyTorch, Adaptive Computation, Efficient Inference

- Implemented early-exit transformer blocks with per-layer exit heads and confidence-based gating for dynamic depth.
- Designed a shared KV-cache across layers to support partial exits while avoiding redundant attention computation during decoding.

LLM Debater | 🌐 | LangGraph, Multi-Agent Systems, LLMs, Gradio

- Built a multi-agent debate system using LangGraph, where multiple LLM agents argue opposing sides over multiple turns with explicit state tracking and control flow.
- Implemented an LLM-based judge agent to evaluate arguments and declare a winner.

UpDawg | 🌐 | Software Engineering, Design Patterns, Python, Flask, LLMs

- Collaborated in a team of 5 to develop UpDawg, a prototype Personal Assistant that aggregates and summarizes user data from services like Slack and Outlook, tailored to user preferences using LLMs.
- Employed architectural and design patterns, focusing on performance and modifiability while addressing stakeholder concerns through comparative analysis of architectural tactics.

AS2 – Answer Sentence Selection | 🌐 | Python, PyTorch, Transformers

- Implemented and fine-tuned models to identify the best candidate answers for a question. Utilized a 2-step transformer fine-tuning approach - first on a general QA dataset, followed by fine-tuning on a task-specific dataset - achieving results close to SOTA.

Publications

Lee, A. H., Semnani, S. J., Castillo-López, G., de Chalendar, G., Choudhury, M., Dua, A., **Kavitha, K. R.**, et al. (2024). *Benchmarks Underestimate the Readiness of Multilingual Dialogue Agents*.

Asapu, L., Kodali, P., **Kavitha, K. R.**, Dua, A., & Shrivastava, M. (2024). *Bridging Laughter Across Languages: Generation of Hindi-English Code-Mixed Puns*.

Technical Skills

Programming Languages: Python, C++, JavaScript, SQL

ML / NLP: PyTorch, Transformers, Hugging Face, vLLM, WandB

Data / Infrastructure: Docker, MongoDB, Pandas, NumPy

Relevant Coursework: Advanced NLP, Introduction to Deep Learning, Machine Learning, Probabilistic and Statistical Inference

Soft Skills: Effective Communication, Teamwork, Reliability