Himanshu Maurya

in ⊕ ⋂ ≣

📞 +44-7496537947

✉ lordzuko.research@gmail.com

EDUCATION

- **University of Edinburgh** — Edinburgh, UK
  Master of Science in Speech & Language Processing — *Sept 2022 – Aug 2023*

- **IIITDM Jabalpur** — Madhya Pradesh, IN
  Bachelor of Technology in Computer Science & Engineering — *Aug 2013 – Aug 2017*

EXPERIENCE

- **Quantiphi Inc. (Applied Science Team)** — London, UK
  MLE - Part Time — *Oct 2022 – Jan 2023*
  - Developed a Python package using Google-OR tools to optimize vehicle routing with capacity, time, and distance constraints.

- **iSchoolConnect Technologies Pvt. Ltd.** — Mumbai, IN
  MLE, Sr. MLE, Lead MLE — *Dec 2018 – Aug 2022*
  - Led a team of machine learning engineers, data engineers and annotators shaping the company's AI product offerings.
  - Spearheaded the ideation and development of the Writing Mentor Tool and online Exam Proctoring Engine, resulting in acquisition by Turnitin.
  - Trained and fine-tuned language models like ULMFiT and BigBird transformer for automatic essay scoring.
  - Fine-tuned a T5 and GPT-2 model to generate essay feedback, resulting in 20% faster turnaround time and 37% reduction in editing service cost to company.
  - Trained adapter transformers to optimize low-resource model training, resulting in $1/8^{th}$ reduction in model size.
  - Developed logistic regression model with conformal predictions to flag suspicious exams for prioritized human review.
  - Optimized the orchestration of 6 ML & DL models to proctor over 0.7 million exams/month at less than \$0.1/exam.
  - Established data pipelines with MongoDB and Apache Kafka. Implemented MLFlow and DVC workflows for tracking.
  - Registered patent *MAURYA, Himanshu et al.* Machine Learning System for Analysing the Quality and Efficacy of Essays for Higher Education Admissions. US Application No. 17/453,416

- **Staqu Techonologies Pvt. Ltd. (R&D Team)** — Gurugram, IN
  Jr. Research Engineer — *Jun 2017 - Oct 2018*
  - Implemented a News Intelligence Engine for Indian languages, and similar news clustering and machine translation modules. These modules enabled semantic search capabilities in PINE, an advanced AI data aggregation and search engine.

- **Oracle India Pvt. Ltd.** — Bengaluru, IN
  Software Development Engineering Intern — *Jun 2016 - Nov 2016*
  - Implemented a Python module for scheduled password updates as part of automation for PaaS setup using Oracle Cloud.

PROJECTS

- **Masters Thesis**: Improving the Appropriateness in Cross-Text Prosody Transfer using Human Supervision, supervised by Atli Sigurgeirsson and Prof. Simon King. Implemented a novel approach for enhancing naturalness and prosodic appropriateness in cross-text Prosody Transfer (PT) through efficient Human-in-the-Loop (HitL) supervision.
  - Trained Daft-Exprt with HiFi-GAN on Blizzard 2013 dataset for expressive prosody modelling.
  - Developed a speech editing application for word-level acoustic feature modification, leading a speech editing experiment with 33 participants, resulting in a novel Speech Editing Corpus (SEC).
  - Conducted subjective evaluations of SEC with 68 participants, yielding a 70% improvement in perceived PT appropriateness. Introduced a quantitative edit effectiveness metric, providing valuable HitL system design insights.

- **Professional Matchmaking Engine**: Match professionals with common interests and aligned objectives.
  - Created a knowledge graph containing 0.2 million users using Neo4j and used Node2Vec and Transformers to embed interaction and profile data into user embeddings. Implemented a stable roommate algorithm for relative candidate scoring and matchmaking based on user preferences. Leveraged FAISS to decrease candidate scoring time by 23%.
  - Improved engagement on relevant users page by 31% resulting in a 7% increase in matches.

- **Stance Detection**: Automated detection of author's political stance in news articles using a transformer model with a classification head for stance prediction.

TECHNICAL SKILLS

- **Frameworks**: PyTorch, TensorFlow, Hugging Face Transformers, LangChain
- **MLOps & Data Engineering**: MLFlow, DVC, Ray, Apache Kafka
- **Languages**: Python, Java, C++
- **Application Servers**: Tornado, Flask
- **Database**: MongoDB, PostgreSQL, Neo4j
- **Developer Tools**: Git, Docker, Jenkins (CI/CD)
- **Cloud Platforms**: GCP (Vertex AI, BigQuery), AWS (Sagemaker), HPC Cluster (Cirrus) with SLURM Job Scheduler