

Title Upsk-AI-lling: Mapping skills in higher and further education with AI

Lead Supervisor Denis Newman-Griffis, d.r.newman-griffis@sheffield.ac.uk
School of Computer Science, University of Sheffield

Project goals and context

Higher and further education are vital pathways for skill development, but sector-level understanding of the skills provided by the enormous variety of courses available is difficult to achieve. This project, funded by a British Academy Innovation Fellowship and a QR-Policy Support Fund award from the University of Sheffield, is developing a Large Language Model (LLM)-driven platform for analysing at scale the skills described in higher education, further education, and apprenticeship training courses, and mapping these to professional occupations and industrial sectors.

The overarching goals of the project are to:

- 1) Enable national analysis and monitoring of key skills addressed in subject-oriented training, beginning with AI-related skills;
- 2) Facilitate on-demand insight into the skills offered by specific courses and the UK tertiary education sectors as a whole.

Computational aspects and need for HPC resource

We have developed an initial pipeline for automated skill extraction and mapping from course descriptions, built around the Google Gemini API. However, reliance on Gemini has several significant limitations, including the need to restrict analysis to public data (thus excluding internal course documents), the environmental and financial cost of cloud-based GenAI services, and the slow speed of network-based prompting.

The role of the N8 CIR intern will be to re-engineer LLM usage within the Python-based project to enable use of locally-hosted LLMs that can be run on Tier 2 or Tier 3 HPC resources with both public and private data. This will also enable researchers to experiment with the empirical effects of different prompting strategies and compare different pre-trained and fine-tuned LLMs for skills analysis, making the project both more rigorous and more adaptable. These improvements are necessary for our planned work to scale up the analysis from a current dataset of 650 HE courses, 88 FE courses, and 61 apprenticeships; to a broader sector-level analysis of over 30,000 HE courses and 700 apprenticeships.

We are already working with the University of Sheffield RSE team (with Edwin Brown as lead contact) on a web interface for interactive exploration of the output of the skills analysis, between April-June. The N8 CIR intern will therefore be able to learn from the work of an established RSE on this project, whilst developing their own skills and contributing to a different part of the software package.

Schedule for N8 CIR intern

Week 1: Familiarisation with existing code base. Week 2: Implementing local LLM wrapper for skill extraction and mapping. Week 3: Implement abstraction layer for configurable use of web API or local LLM wrapper. Week 4: Performance profiling of web API vs local LLM models. Week 5: Add user interface to existing Flask project, for adjusting skill extraction and analysis LLM prompts. Week 6: Improving provenance tracking of analysis outputs based on LLM model selection and user-configured prompts. Week 7-8: Add user interface for batch processing new collections of course data, using existing data models.