

Large Language Models in Labor Market Research Data Management: Potentials and Limitations

Thursday 18 September 2025 10:10 (20 minutes)

This contribution explores the application of large language models (LLMs) in labour market research data management, particularly in occupational data analysis. Based on our empirical studies of the automated classification of job titles and critical evaluations of AI-assisted text interpretation, we contend that, although LLMs present promising opportunities to improve research processes, such as providing query assistance, offering annotation support, and facilitating preliminary content structuring, they are inadequate for consistent data management, reliable analysis, and interpretative depth. Our findings suggest that, while LLMs can support research workflows as interactive tools, they cannot replace methodological approaches in data-driven social science research. Our aim is to contribute to the discussion at the workshop on the scope and boundaries of LLM-based tools in research data management.

The increasing availability of large language models (LLMs) creates new opportunities and challenges for research data management (RDM), especially when dealing with complex and diverse data sources, such as labour market information. We build on documents from the German labour market archive containing data on vocational education and training (VET) and continuing VET (CVET). The archival form of these regulations, primarily unstructured or semi-structured scanned documents, poses challenges for digital accessibility, analysis and integration with contemporary data systems, as described in \cite{reiser2024towards}. However, the digitisation of archival material provides an opportunity to preserve, structure and analyse regulatory knowledge in a form that is compatible with semantic linking, machine learning and long-term data curation, as discussed in our previous work \cite{reiser2024analyzing,reiser2024learning}. This system incorporates a web-based information system \cite{reiser2025is} and a data warehouse backend \cite{hein2024linked}, along with various analysis pipelines.

Our research examines the integration of LLMs into two distinct areas of labour market studies: (1) the automated classification of job titles \cite{reiser2025ecai} and occupational data to ontologies like the GLMO \cite{dorpinghaus2023towards}; and (2) the analysis of texts related to labour, education and social discourse within a given hermeneutical framework \cite{hermen}. We draw on empirical studies using annotated survey data, synonym datasets, online job advertisements and vocational training records, as well as comparative experiments assessing the interpretative capabilities of LLMs across various text genres.

Our findings suggest that, although LLMs can be effective interactive tools that assist with tasks such as content summarisation, query reformulation and preliminary data exploration, their performance in core data analysis tasks is inconsistent and unreliable, making them unusable for most scientific purposes. Specifically, LLMs fail to deliver reproducible results in classification tasks; for instance, at fine-grained levels of occupational coding \cite{2025dorau,reiser2025ecai}. In hermeneutic contexts, models are highly sensitive to prompt design, language and model architecture, which undermines their suitability for structured analysis or theory-driven interpretation \cite{hermen}. These limitations emphasise the risk of overestimating LLMs' capabilities in domains requiring specific methods, domain knowledge and theoretical grounding. Nevertheless, this contradicts some recent literature which asserts that LLMs “can analyze nearly any textual statement.” \cite{tornberg2023use} The experimental results obtained in this study appear to support the arguments of researchers who claim that LLMs are incapable of performing even basic logic-based tasks, such as counting and identifying general substructures in graphs, see, for example \cite{fu2024large,nguyen2024evaluating}. Therefore, it might be debatable whether LLMs could offer any technical assistance with textual analysis. This could be against, for example, \cite{tai2024examination}.

We argue that the primary value of LLMs in RDM for labour market research lies in their potential to improve interaction between researchers and data by supporting hypothesis generation, assisting with data annotation and facilitating interdisciplinary dialogue, rather than automating analytical processes or replacing established methods, as LLMs have been shown to have difficulty understanding text \cite{saba2024llms}, and they also appear to lack an understanding of context, intentionality, and reader-writer dynamics. This distinction is crucial for the responsible integration of AI tools into research workflows without compromising scientific standards.

In conclusion, we advocate for the cautious and differentiated use of LLMs in labour market research. Rather than viewing LLMs as a universal solution for data analysis, we suggest positioning them as supportive tools that complement human expertise during the exploratory and communicative phases of research. We invite

discussion on how LLMs might augment methodological rigour in data-intensive research fields, rather than substitute it, and on the development of evaluation criteria for their responsible application in RDM.

Authors: DÖRPINGHAUS, Jens (University of Koblenz, Federal Institute for Vocational Education and Training (BIBB)); Mr TIEMANN, Michael (University of Koblenz, Federal Institute for Vocational Education and Training (BIBB))

Presenters: DÖRPINGHAUS, Jens (University of Koblenz, Federal Institute for Vocational Education and Training (BIBB)); Mr TIEMANN, Michael (University of Koblenz, Federal Institute for Vocational Education and Training (BIBB))