

Title: Large Language Models for Automated Metadata Extraction in Social Science Research

Abstract

This study examines the use of large language models (LLMs) for automated extraction of structured metadata from academic research papers. Using a corpus of peer-reviewed social science articles, we evaluate whether schema-constrained prompting can reliably identify research objectives, data sources, methodological approaches, and key findings. Results indicate that LLM-based extraction achieves high precision for explicitly stated sections such as data and methods, while recall is lower for implicit limitations. The findings suggest that LLMs can support scalable literature review workflows when used as extraction tools rather than interpretive agents.