LitCurate: An AI-Powered Package for Scientific Literature Curation, Demonstrated on Lower Mantle Equation-of-State Data

Wilson Bishop Samuels[1,2], Abin Shakya[2], and Renata M. Wentzcovitch[2]

[1] Tuskegee University, [2] Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York

Seismic interpretations of the lower mantle depend on a small set of equation-of-state (EOS) parameters — bulk modulus, its pressure derivative, and unit-cell volume — that describe how each candidate mineral compresses and stiffens under the pressures and temperatures found 700 to 2,900 km beneath Earth's surface. These parameters are reported across decades of experimental and computational literature, each paper covering a different mineral, phase, or pressure range, in a different format. The field's standard reference has long been a hand-compiled table that requires someone to manually read each new paper and re-enter its values, a process that does not scale as the literature grows, and no existing tool combines automated literature harvesting, structured data extraction, and semantic search for this class of mineral-physics data. To address this, we built LitCurate, a thirteen-stage pipeline that harvests candidate papers by DOI through CrossRef, converts and cleans full-text PDFs, and uses LLM-based schema extraction to pull EOS parameters into a structured, searchable database. Of 1,855 candidate papers identified, 918 were downloaded and 314 passed full-text filtering, yielding 208 papers successfully extracted into 1,350 structured EOS parameter rows, an overall search-to-extraction yield of about 11%. By pairing this extraction pipeline with a semantic search layer and an interactive dashboard, LitCurate lets researchers query the literature by meaning rather than exact keywords, cutting what would normally take weeks of manual review down to minutes. Implications extend beyond lower-mantle mineral physics to any geoscience subfield whose foundational data remains scattered across a large and growing body of literature.