Methodology: Pipeline Design and Formal Definitions

This section describes the six components of our computational meta-analysis pipeline: literature retrieval, canonical deduplication, LLM-based assertion extraction, probabilistic knowledge graph construction, hypothesis scoring, and end-to-end orchestration. The pipeline extends the systematic literature analysis approach of Knight et al. knight2022fep—which combined manual annotation with ontology-based automated analysis—by substituting manual coding with fully automated, LLM-driven assertion extraction and citation-weighted hypothesis scoring.

Multi-Source Literature Retrieval

We retrieve papers from three complementary academic databases to maximize coverage and enable cross-source deduplication:

arXiv. We query the arXiv Atom API using the phrase-matched search `all:"active inference" OR all:"free energy principle"`. The `all:` prefix searches titles, abstracts, and full text; phrase matching reduces contamination from unrelated physics papers that mention "free energy" in thermodynamic contexts.