

Supplementary Information

The AtomicLimits ALD/E Database: Unlocking the Future of ALD/E with Large Language Models

Eleni Poupaki^{1*}, Sameer Sadruddin², Jennifer D'Souza², Alex Watkins³, Bora Karasulu³, Sören Auer², Adrie Mackus¹, and Erwin Kessels¹

¹ Eindhoven University of Technology, Netherlands

² TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

³ University of Warwick, United Kingdom

*e.poupaki@tue.nl

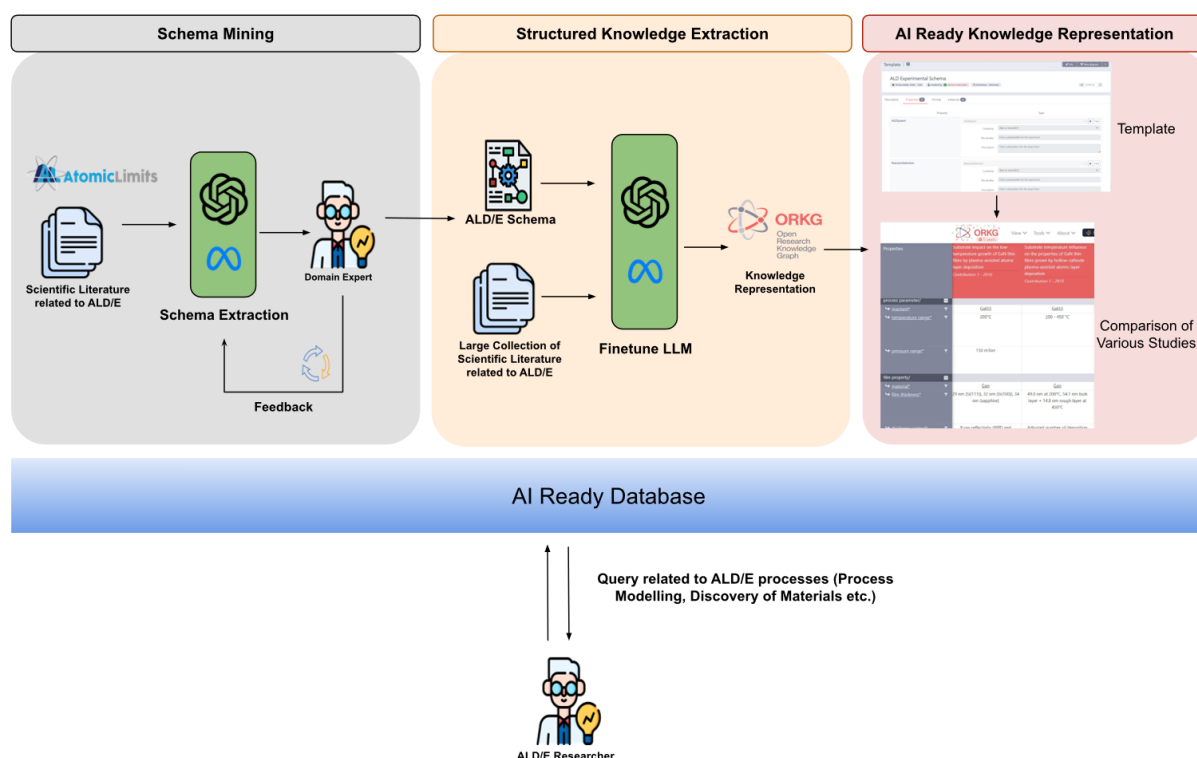


Figure 1 : Overview of the workflow for transforming unstructured ALD/E literature to structured data. The process starts with schema mining using LLMs and domain-expert feedback. These schemas are then used as a foundation for structured knowledge extraction, where LLMs process large collections of ALD/E related scientific texts. The extracted knowledge is then represented as a Knowledge Graph (KG), specifically utilizing the Open Research Knowledge Graph (ORKG). The final outcome is a database with extensive ALD/E data that enables accelerated material science discovery in ALD/E research.