Commentary | Open Access

AI Agent

Yang. *AI Agent* **2026**, *2*, 16 **DOI:**10.20517/aiagent.2026.29

Battery material databases in the age of AI agents

Menghao Yang

Citation: Yang, M. Battery material databases in the age of AI agents. *AI Agent* **2026**, *2*, 16. https://dx.doi.org/10.20517/aiagent.2026.29

Received: 17 Jun 2026 **First Decision:** 10 Jul 2026 **Revised:** 14 Jul 2026 **Accepted:** 22 Jul 2026 **Published:** 31 Jul 2026

Academic Editor: Weijie Yang **Copy Editor:** Tong Wang **Production Editor:** Tong Wang

MAIN TEXT

With the integration of artificial intelligence (AI) and materials science, the material database has changed from a passive data warehouse to an active tool that can support reasoning and new material discovery. The most realistic achievement of this development trend is the solid-state electrolyte dynamic database (DDSE), which has been renamed DigBat recently and can be accessed through www.digbat.org. DDSE organizes the relevant properties of various solid electrolytes into a platform with regular structure and open to the outside world. Researchers can directly query and analyze data; the exported format can be directly used for research related to AI agents[1,2] [Figure 1]. It keeps updating data, adopts standardized data structures, and is designed to meet the needs of users, just in line with the requirements of AI-driven battery material research on data semantic richness. Currently, DDSE/DigBat provides solid-state electrolyte data warehouses that are structured and machine-readable. As for the fully independent closed-loop discovery function for AI agents to form research hypotheses and plan experiments, DigBat has not yet been realized, which is a goal to be achieved in the future. There are still many practical key problems that need to be solved first. For example, different measurement conditions and different report formats make data quality control very complicated. In addition, if FAIR (findable, accessible, interoperable, and reusable) principles can be implemented and unified data entry templates can be used, data reliability can also be improved.

The importance of this kind of database is far more than that of a single platform itself. The whole research field has also done a lot of complementary work to build the basic support for AI-driven material discovery[3]. For example, some people have done the research of automatic text mining, built the solid electrolyte ionic conductivity database, and sorted the extracted conductivity values and typical structure descriptors into specific formats; it is specifically used for literature-based data extraction and database reconstruction[4]. The underlying thinking of these projects is the same as that of DDSE: database is no longer just a tool for storing data, but a constantly updated knowledge system that allows algorithms to directly call material-related knowledge. Now, we have systematically built a complete data

College of Smart Energy, Shanghai Jiao Tong University, Shanghai 200240, China.

Correspondence to: Dr. Menghao Yang, College of Smart Energy, Shanghai Jiao Tong University, Shanghai 200240, China. E-mail: menghaoyang@sjtu.edu.cn

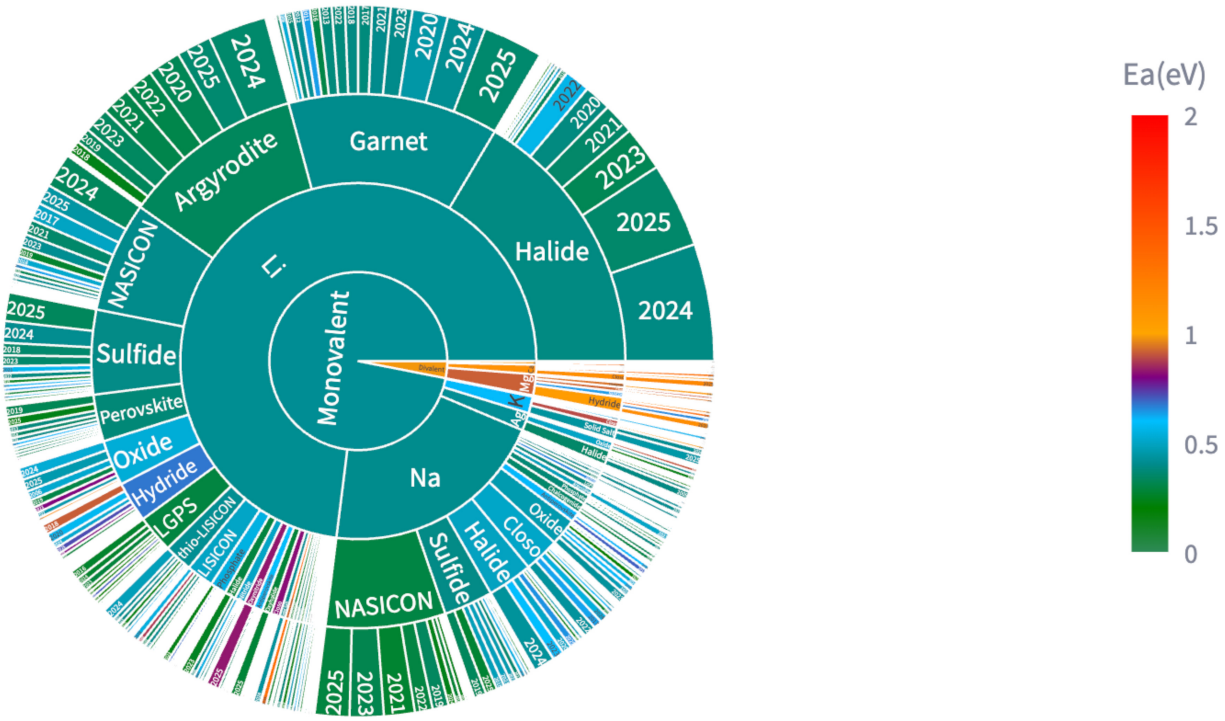

Figure 1. The interface of the DDSE, now rebranded as DigBat. Reproduced from the DigBat website (https://www.digbat.org/). DDSE: Dynamic Database of Solid-State Electrolyte; NASICON: Na Super Ionic CONductor; LGPS: lithium germanium phosphorus sulfide.

structure and filled it with audited measurement data; it is equivalent to building the most basic support for more complex AI tools. The whole system can be divided into three layers: the database provides basic support with structure and clear semantics; AI models (such as graph neural network and large language model) learn the relationship between material structure and performance from these data; AI agents combine model, retrieval and reasoning functions to complete such multi-step tasks as literature mining and candidate material screening.

However, the development from database to intelligent AI agent cannot be achieved naturally, and there are also many difficulties in the process. The database provides the AI agent with structured basic data, but the agent must have a specially designed architecture to convert the data into practical operations. With the support of structured databases, AI agents can accurately extract performance values of materials, retrieve corresponding materials according to performance indicators, screen candidate objects, summarize development trends, and perform clustering analysis on materials. The new results can also be compared with previously published studies. Therefore, the AI agent with database support is a very practical interactive auxiliary tool, which cannot completely replace researchers to do research independently. Recent advances in AI agents in the field of solid-state battery research have clearly reflected the relationship between the database and agents. A representative study integrates a comprehensive sulfide solid electrolyte database into a framework that combines a large language model and *ab initio* dynamics simulation; it accelerates the discovery speed of the hydrogen solid electrolyte. This kind of materials have a brand-new two-step migration mechanism, and the activation energy is as low as 0.62 eV[5]. Another breakthrough achievement is the research agent AutoSEE, which can independently summarize the design principles of solid electrolyte, relying on the use of GPT-4o to coordinate literature mining and unsupervised clustering work; the design principle summarized by the agent has also been verified by experiments[6]. In these two cases, the database is not passive background information; it is the core cognitive basis supporting the realization of independent discovery. Now studies have proved that the wrong data in the database will

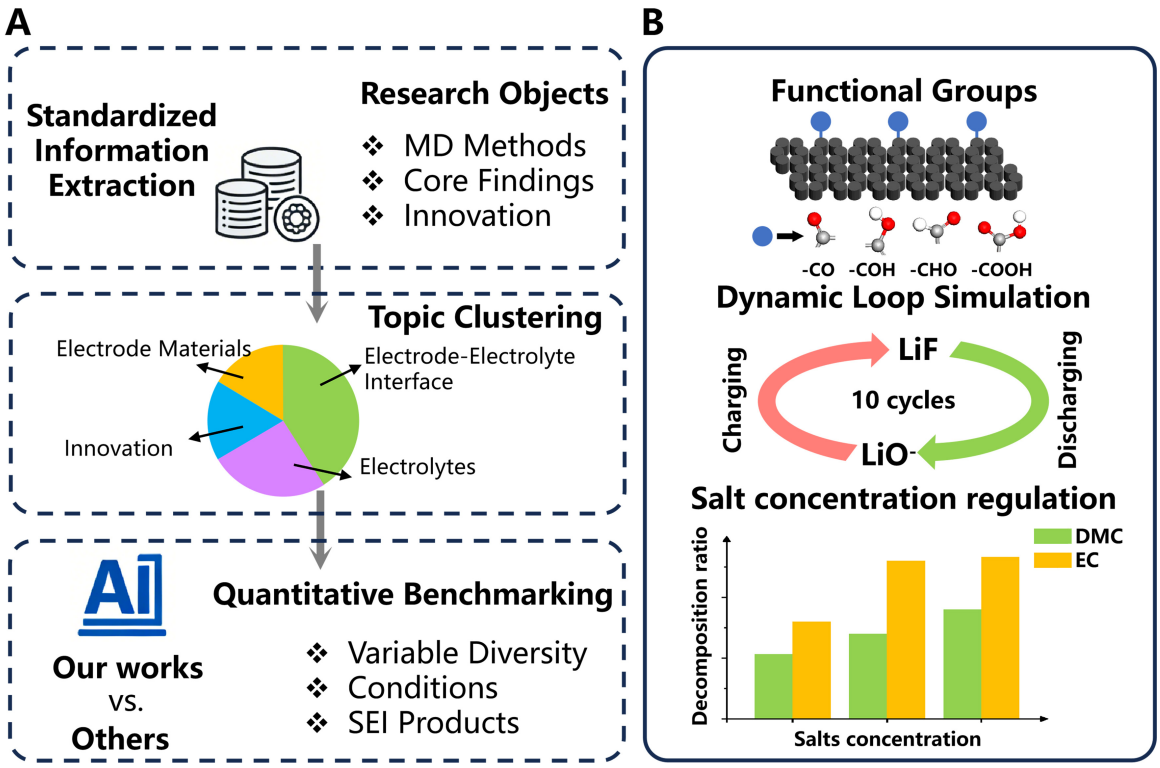

Figure 2. AI-driven comparative analysis framework for lithium-ion batteries. (A) MD research; (B) multi-dimensional innovation of this research. (A) builds a knowledge base based on MD simulation research: AI agent identifies the shortcomings of current research, and performs the dynamic cycle research combined with graphite functional groups; (B) shows the experimental components that can directly make up for these deficiencies. MD: Molecular dynamics; SEI: solid-state electrolyte interface; DMC: dimethyl carbonate; EC: ethylene carbonate.

reduce the accuracy of the model. For example, the collected ionic conductivity data are measured at different temperatures without normalization, which will lead to extremely optimistic prediction results. Repeated data entries and incorrectly labeled components can also mislead the classifier. If the AI agent does not evaluate the uncertainty and feels that the reliability of all data is the same, then its reasoning chain is based on errors from the beginning; this also means that we need to do a better job in data review and develop a proxy architecture that can consider data uncertainty.

Recently, there is a study on the simulation of lithium-ion battery (LIB) solid-state electrolyte interface (SEI), which shows the operation process from database to agent more intuitively. This study built a knowledge base containing thousands of high-impact molecular dynamics (MD) studies, using AI agents for standardized information extraction, topic clustering and quantitative comparison [Figure 2]. In the end, they not only clearly explained the innovation points of their own research, but also verified a more universal thinking: in the AI agent era, the structure is regular, and databases with rich semantics are the essential foundation for scientific research intelligence[7]. There are still deeper problems unsolved. The problem of data deviation is very common: the proportion of chemical systems and materials with better performance that have been studied more in the database is too high, and the failed experimental data are basically not included, this will cause the forecast result to be too optimistic. The universality of the model is also uncertain: if the agent performs well on the lithium-ion conductor, it may be impossible to use it if it is changed to the sodium-based or magnesium-based system. The combination of big language model proxy and simulation engine results in high computing costs, which are not taken into account in many proof-of-concept studies. To solve these problems, we must develop a proxy architecture that can adapt to changes in data distribution, evaluate uncertainties, and achieve higher computing efficiency.

From this SEI study, the three functions of this kind of database can be clearly summarized. First, databases can implement semantic standardization. The reason why AI agents can make meaningful comparisons across different studies is that they have a preset data structure, which can convert scattered expressions in scientific research papers into electrode materials, force field types, dynamic/static conditions and SEI product integrity are comparable and machine-readable fields[8]. Without this structured basis, the content output by the agent will be superficial or fictitious. With this basis, the agent can reliably divide the relevant documents into electrodes, electrolytes, interfaces, method innovation; we can also find that interface SEI research is the most important innovation direction now. If combined with multi-modal processing capability, it can not only process text descriptions, but also process images characterized by materials, and the analysis capability of agents can be further improved; information across different expressions can be extracted[8].

Second, the database can achieve real quantitative comparison. The agent compares the technical characteristics proposed by the control of graphite functional group (FG), with the structured entries in the knowledge base. Through this method, three specific shortcomings of the most advanced simulation research are directly found: only a single FG model can be used; static simulation is used instead of dynamic cyclic simulation. There is no correlation between macro concentration effect and the formation mechanism of atomic scale SEI. Therefore, the database is equivalent to a computing framework, which can compare the new research with the previous work and quantitatively evaluate the degree of innovation. Although the examples given here are all related to solid electrolyte and SEI, these ideas are also applicable to the databases related to positive electrode, negative electrode, liquid electrolyte and battery interface that are being developed in the whole field.

Third, databases can support large-scale and multi-dimensional knowledge integration. By integrating relevant multi-scale data from the database, the AI agent can uncover the underlying mechanistic associations. For example, high concentration electrolyte will destroy the tightness of SEI[9]. The ability to process massive and multi-parameter datasets cannot be realized by the human research team. Platforms such as Materials Project, Novel Materials Discovery (NOMAD) and Automatic-FLOW (AFLOW) have become practical standards in the field of Materials and data, and are now being continuously optimized to develop towards the semantic richness required by AI agents. The battery database and AI agent framework jointly built by everyone in the field, including the functions of automatic document extraction, multi-step reasoning and experimental coordination, have laid a solid foundation. In this overall ecosystem, our work (DDSE/DigBat, AutoSEE, MatImageAgent, and SEI-related research) is only a reference case, not an isolated or optimal solution.

We can get a very clear conclusion: Now AI agents play an increasingly important role in proposing research hypotheses and designing experiments. The quality of their reasoning results is more determined by the depth and structure of the queried database, while the advanced degree of the algorithm is secondary. We have invested resources to build a battery material database that is jointly maintained by the field, can trace data sources, and has rich semantics. At the same time, we have included successful experimental results and valuable failure cases; it will directly determine the speed of energy material discovery. The database redesigned for the proxy era can turn the knowledge of the whole field into the content that machines can call directly, so that the idea of AI-driven scientific research can be truly implemented and the development of the next generation of battery technology can be promoted.

DECLARATIONS

Authors' contributions

The author contributed solely to the article.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

This work was supported by the National Natural Science Foundation of China (Grants 52302302 and 52572257).

Conflicts of interest

The author declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

REFERENCES

1. Yang, F.; Campos Dos Santos, E.; Jia, X.; et al. A dynamic database of solid-state electrolyte (DDSE) picturing all-solid-state batteries. *Nano. Mater. Sci.* **2024**, *6*, 256-62. DOI
2. Yang, F.; Wang, Q.; Cheng, E. J.; Zhang, D.; Li, H. User instructions for the Dynamic Database of Solid-State Electrolyte 2.0 (DDSE 2.0). *Comput. Mater. Contin.* **2024**, *81*, 3413-9. DOI
3. Hargreaves, C.; Gaultois, M. W.; Daniels, L. M.; et al. A database of experimentally measured lithium solid electrolyte conductivities evaluated with machine learning. *npj. Comput. Mater.* **2023**, *9*, 9. DOI
4. Shon, Y. J.; Min, K. Extracting chemical information from scientific literature using text mining: building an ionic conductivity database for solid-state electrolytes. *ACS. Omega.* **2023**, *8*, 18122-7. DOI PubMed PMC
5. Wang, Q.; Yang, F.; Wang, Y.; et al. Unraveling the complexity of divalent hydride electrolytes in solid-state batteries via a data-driven framework with large language model. *Angew. Chem. Int. Ed. Engl.* **2025**, *64*, e202506573. DOI PubMed PMC
6. Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **2018**, *120*, 145301. DOI PubMed
7. Jiang, M.; Zhou, J.; An, Y.; Lin, Z.; Yang, M. Artificial intelligence for energy materials research: from classical machine learning to large models. *ENG. Energy.* **2026**, *20*, 10535. DOI
8. Chen, Y.; Fan, Y.; Gong, Y.; et al. Atomistic insights into dynamic evolution of solid electrolyte interface. *J. Energy. Chem.* **2025**, *111*, 401-11. DOI
9. Li, M.; Lu, D.; Wang, J.; et al. Solvation sheath reorganization enables fast ion transfer kinetics in lithium-ion battery. *Nat. Commun.* **2026**, *17*, 3953. DOI PubMed PMC