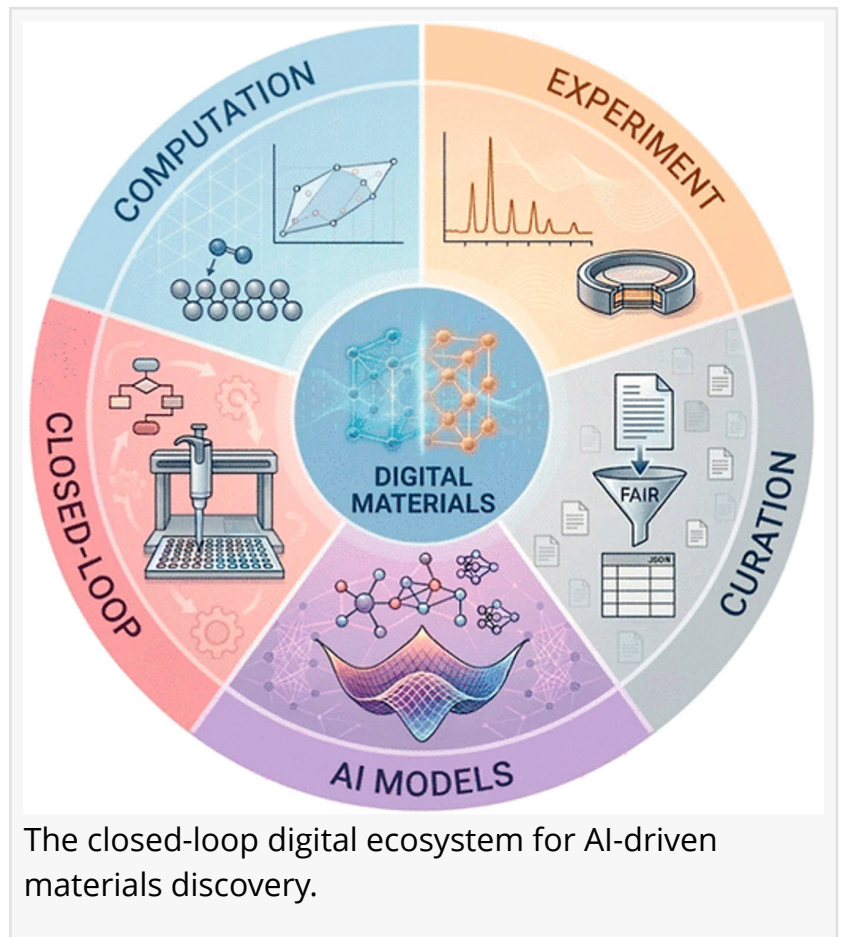


From “dark data” to smart labs: A roadmap for reliable AI in materials science

GA, UNITED STATES, May 6, 2026 /EINPresswire.com/ -- The shift from slow, trial-and-error experimentation to data-driven discovery is reshaping how we develop energy materials. A new Perspective argues that the design of materials databases—how they ingest, curate and share information—directly determines the trustworthiness of modern [artificial intelligence](#) (AI) models. The authors map the current landscape of computational and experimental databases, from bulk properties to surface chemistry and battery performance. They then lay out a roadmap linking databases to graph neural networks, machine learning potentials, and AI Agents. Without smarter data infrastructure, even the most advanced algorithms risk learning from biased or incomplete information.



For decades, materials science relied on physical intuition and slow laboratory cycles. Even with the rise of high-throughput computation, databases have often remained fragmented, using incompatible formats and missing critical context such as synthesis conditions or failed experiments. This “silo effect” and the widespread underreporting of negative results mean that many artificial intelligence (AI) models train on polished success stories only. Such biases limit generalization and can lead to overoptimistic predictions. Based on these challenges, the authors call for a deeper, systematic rethinking of database architecture to support reliable and reproducible AI-driven materials discovery.

A team led by researchers at North China Electric Power University and Tohoku University,

publishing (DOI: [10.1021/prechem.5c00449](https://doi.org/10.1021/prechem.5c00449)) in Precision Chemistry on March 27, 2026, provides a comprehensive analysis of how materials databases can be built to power the next generation of artificial intelligence in materials science. Their Perspective bridges computational repositories, experimental data platforms and emerging AI tools, offering a practical roadmap for turning raw information into autonomous discovery.

The analysis breaks materials databases into two major families. Computational repositories, such as bulk property and surface-interface databases, offer quantum mechanical data at scale but often suffer from “idealization bias,” modeling perfect crystals at zero Kelvin while ignoring real-world complexity. Experimental databases, covering crystal structures, catalysis, batteries and hydrogen storage, provide high-fidelity references but remain costly to generate and plagued by inconsistent reporting.

Importantly, the authors highlight integrated platforms that connect computed descriptors with experimental evidence. They showcase emerging AI tools, including a multi-agent workflow called DIVE that extracts hydrogen-storage data from old literature graphs with up to 30% better accuracy than standard models. For batteries, the Dynamic Database of Solid-State Electrolytes (DDSE) already links processing variables to performance metrics. The Perspective also stresses that negative results—failed syntheses or poor performance—must be treated as first-class records. Without such “dark data,” AI models cannot learn where exploration is scientifically futile, leading to wasted experimental cycles.

The authors emphasized that databases have evolved beyond passive storage, now fundamentally defining the learning boundaries of AI models. They warned that feeding algorithms solely with successful experiments and idealized crystal structures essentially trains them to be overconfident and blind to real-world complexities. Consequently, the team called for a paradigm shift: treating negative results as valuable assets, storing uncertainty estimates alongside numerical data, and designing databases that allow machines to fully comprehend data provenance. They concluded that only through such measures can AI agents transform from unpredictable “black boxes” into reliable partners within the laboratory. This roadmap points toward self-driving laboratories where AI Agents plan experiments, retrieve contextual data and automate synthesis with minimal human oversight. For energy applications—better battery electrolytes, hydrogen storage materials or catalysts for clean fuel production—such closed-loop systems could compress discovery timelines from years to weeks. Federated learning offers a way to train models across institutional boundaries without exposing proprietary data. Ultimately, the shift from “big data” to “smart data” could make materials research faster, cheaper and far more reproducible, accelerating the transition to sustainable energy technologies.

DOI
[10.1021/prechem.5c00449](https://doi.org/10.1021/prechem.5c00449)

Original Source URL

<https://doi.org/10.1021/prechem.5c00449>

Funding information

H.L. and D.Z. thank the support by JSPS KAKENHI (Nos.JP25H01508, JP25K01737, and JP25K17991).

Lucy Wang

BioDesign Research

[email us here](#)

This press release can be viewed online at: <https://www.einpresswire.com/article/910793503>

EIN Presswire's priority is source transparency. We do not allow opaque clients, and our editors try to be careful about weeding out false and misleading content. As a user, if you see something we have missed, please do bring it to our attention. Your help is welcome. EIN Presswire, Everyone's Internet News Presswire™, tries to define some of the boundaries that are reasonable in today's world. Please see our Editorial Guidelines for more information.

© 1995-2026 Newsmatics Inc. All Right Reserved.