Research Highlight | Open Access

AI Agent

Hong *et al. AI Agent* **2026**, *2*, 2 **DOI:**10.20517/aiagent.2026.04

DIVE-to-design: how a multi-agent workflow converts figure-centric literature into an ai-native hydrogen storage discovery engine

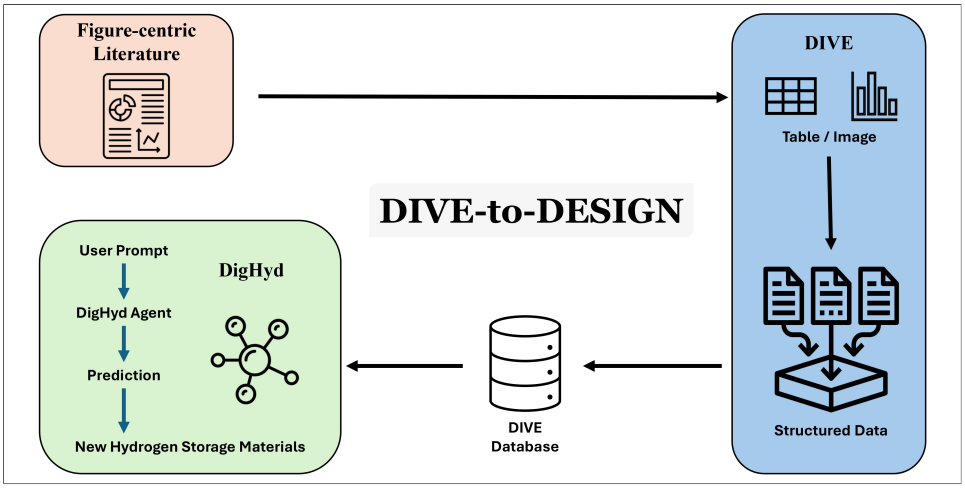
Yuyang Hong, Xin Mao

Keywords:
AI-driven materials discovery, solid-state hydrogen storage

Citation: Hong, Y.; Mao, X. DIVE-to-design: how a multi-agent workflow converts figure-centric literature into an ai-native hydrogen storage discovery engine. *AI Agent* **2026**, *2*, 2. https://dx.doi.org/10.20517/aiagent.2026.04

Received: 25 Feb 2026
First Decision: 2 Mar 2026
Revised: 6 Mar 2026
Accepted: 12 Mar 2026
Published: 27 Mar 2026

Academic Editor:
Weijie Yang
Copy Editor:
Xing-Yue Zhang
Production Editor:
Xing-Yue Zhang



Abstract

A major bottleneck in artificial intelligence (AI)-driven materials discovery is not model architecture, but limited data accessibility: critical experimental knowledge remains locked in figures, heterogeneous reporting formats, and unstructured PDFs. A recent study by Li *et al.* addresses this challenge by introducing DIVE (Descriptive Interpretation of Visual Expression), a multi-agent extraction framework that transforms figure-centric scientific content into structured, machine-actionable data. Applied to solid-state hydrogen storage materials, DIVE demonstrates substantial extraction gains over conventional direct large language model (LLM) parsing, then scales to mine 4,053 publications (1972-2025) and build a > 30,000-entry database that powers a downstream inverse-design agent, DigHyd. This work offers a practical blueprint for moving from "LLM-assisted reading" to "AI-enabled discovery infrastructure", linking literature mining, quality scoring, database construction, and target-driven candidate generation in a single workflow.

Check for updates

School of Chemistry and Physics and Centre for Material Science, Faculty of Science, Queensland University of Technology (QUT), Brisbane 4000, Australia.

Correspondence to: Dr. Xin Mao, School of Chemistry and Physics and Centre for Material Science, Faculty of Science, Queensland University of Technology (QUT), Brisbane 4000, Australia. E-mail: x2.mao@qut.edu.au

MAINTEXT

In data-driven materials science, one of the most persistent barriers is not the lack of models, but the form in which scientific knowledge is stored. For solid-state hydrogen storage in particular, many of the most useful experimental signals are embedded in figures rather than in machine-readable text. Pressure-composition-temperature (PCT) curves, temperature-dependent desorption profiles, and electrochemical trends often carry the central evidence, while key metadata remain scattered across captions, methods, and surrounding discussion. As a result, conventional text-first extraction tends to produce incomplete records, especially when visual information is decisive.

The work highlighted here[1] addresses that practical bottleneck by introducing a multi-agent workflow, DIVE (Descriptive Interpretation of Visual Expression) [Figure 1]. It demonstrates that improvements in extraction architecture can directly translate into enhanced discovery capability.

The central idea behind DIVE is conceptually simple but technically important. Instead of asking a single model to read a full paper and directly output structured data, the workflow first isolates figure-centric evidence, converts visual information into explicit descriptive language, and then performs structured extraction on this enriched representation. In other words, visual content is externalized into an intermediate semantic layer before the final parsing stage. This design choice reduces ambiguity in downstream extraction, because the model no longer has to infer quantitative meaning from raw image context in a single step. The paper's framing of "descriptive interpretation" represents more than a naming convention; it is a mechanism that makes figure-derived information computationally tractable in a stable way.

Several literature-mining tools have been developed for chemistry and materials science, most of which focus on extracting information from text. For instance, systems such as ChemDataExtractor are effective at identifying chemical entities and structured relationships described in written form[2]. However, many key experimental results in materials research are presented mainly in figures rather than in exact literature statements. The performance curves, phase diagrams, and desorption profiles often contain the most informative quantitative evidence. These signals are difficult to recover with text-centered extraction alone. DIVE addresses this gap by treating figures as a primary source of information. It converts visual content into descriptive representations before structured extraction. This design allows graphical evidence to be incorporated into machine-readable datasets and complements existing text-oriented literature mining tools.

A notable strength of the study is that it does not evaluate performance through a simplistic exact-match lens. Scientific data extraction quality is usually graded rather than binary: a record may capture the right field with imperfect normalization, or preserve numerical trends while missing experimental conditions. The authors respond to this reality with a scoring framework that jointly measures accuracy and completeness after standardization and semantic alignment. This provides a more realistic basis for model comparison and helps explain why apparently similar extraction outputs can differ greatly in downstream value.

On a manually curated benchmark of 100 papers, DIVE consistently outperforms direct extraction settings across both commercial and open-source combinations. Reported gains are substantial rather than marginal, with improvements typically in the 10%-15% range for leading commercial configurations and 15%-30% for open-source settings. The best overall result is achieved by Gemini 2.5 Flash paired with DeepSeek R1, while the combination of Gemini 2.5 Flash with DeepSeek-Qwen3-8B provides a particularly practical balance between quality, speed, and cost. These comparisons are important because they indicate that the gain is architectural rather than tied to a single proprietary model stack. Although the workflow itself contributes to the improvement, the performance still depends on model versions, prompting methods, and figure quality.

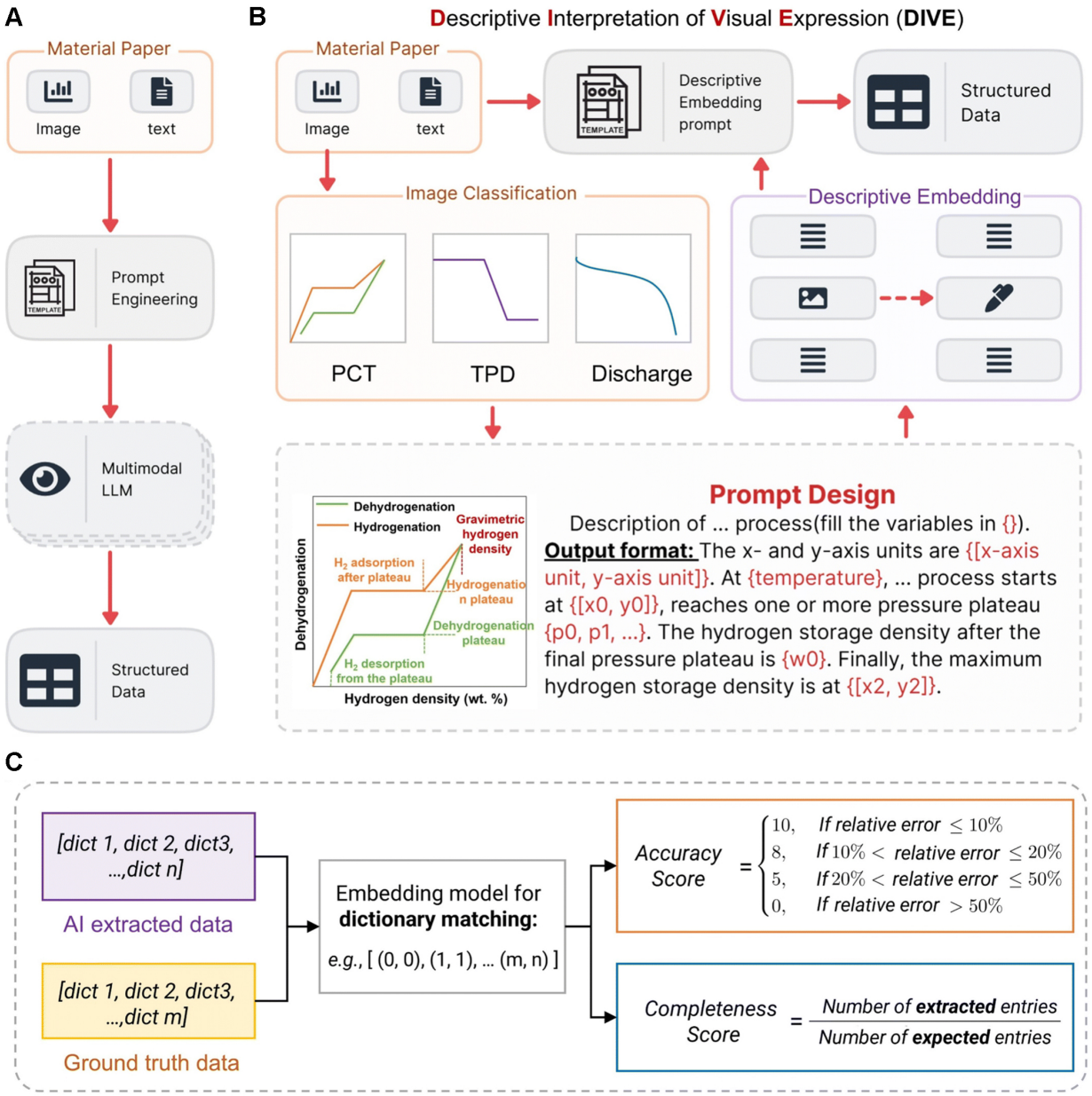


Figure 1. Overview of the DIVE workflow and its evaluation. (A) Conventional single-model pipeline for multimodal extraction; (B) DIVE pipeline: key figure information is first converted into descriptive text prompts, then used for structured data extraction. PCT refers to pressure-composition-temperature curves, and TPD refers to temperature-programmed desorption curves; (C) Batch-evaluation strategy for extraction quality. AI outputs and human annotations are both represented as lists of dictionaries; a shared embedding model aligns matched fields, and the aligned numeric entries are used to compute accuracy and completeness. DIVE: Descriptive Interpretation of Visual Expression; PCT: pressure-composition-temperature; TPD: temperature-programmed desorption; AI: artificial intelligence.

What makes this study especially relevant for the artificial intelligence (AI) Agent community is that it does not stop at benchmark performance. The authors scale DIVE to 4,053 publications spanning 1972 to 2025 and build a structured DigHyd (Digital Hydrogen Platform, www.dighyd.org) resource containing more than 30,000 entries. At that stage, extraction effectively becomes infrastructure. The contribution shifts from "a better parser" to "a usable scientific substrate" that can support reasoning, model training, and design workflows at the field scale. This transition is often missing in AI-for-science studies, where promising extraction metrics are not followed by deployable data systems. Here, the bridge is explicit and operational. To the best of our knowledge, the DigHyd platform is, so far, among the largest experimental solid-state hydrogen storage materials databases reported to date.

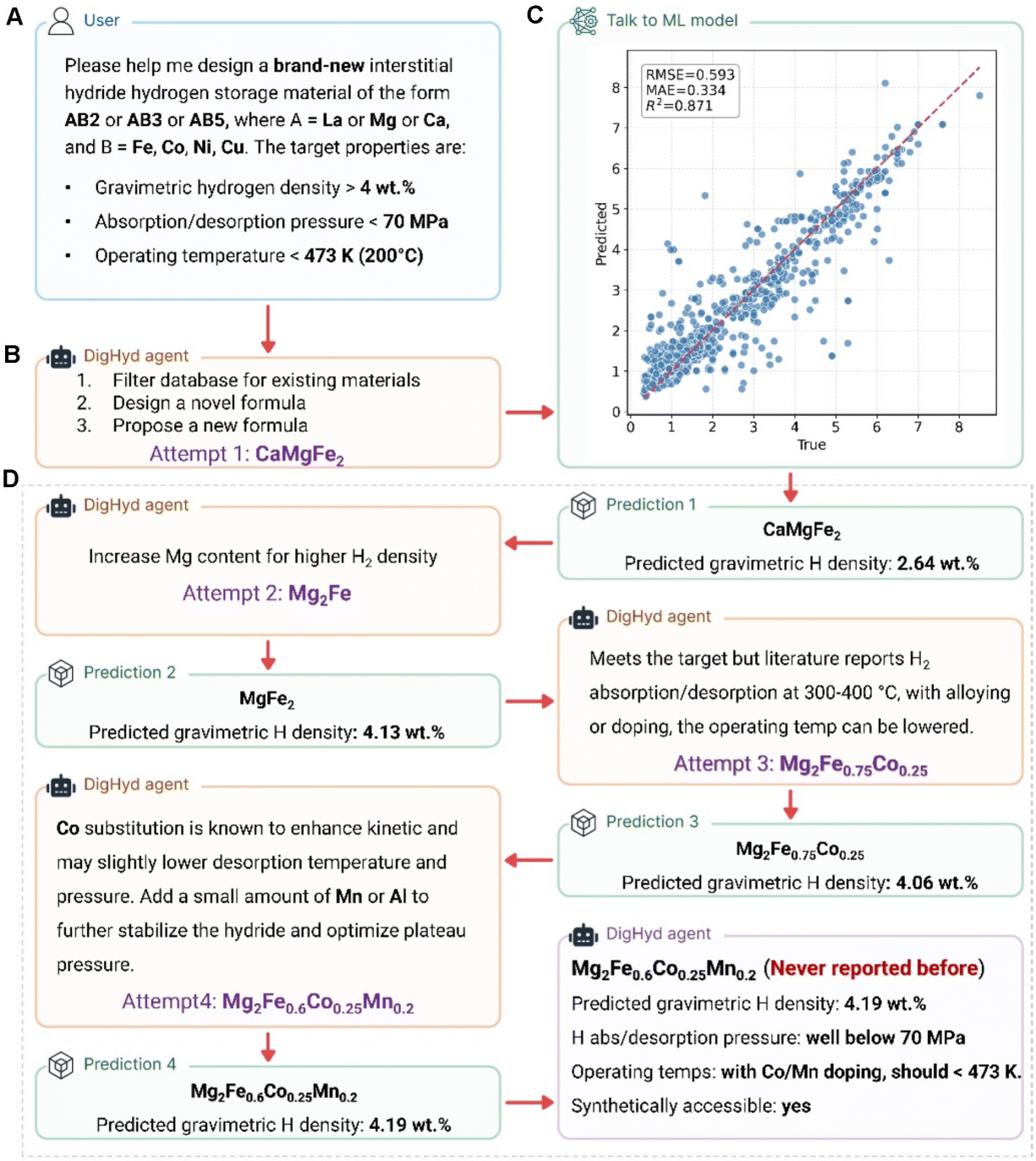


Figure 2. AI agent-driven workflow for discovering new hydrogen storage materials. (A) The researcher defines design constraints, including material class, elemental composition, and target performance; (B) Using knowledge extracted from > 4,000 publications, the DigHyd agent generates initial candidate compositions; (C) A pretrained machine-learning model then predicts the gravimetric hydrogen capacity of each candidate; (D) The agent iteratively designs, evaluates, and refines candidates within minutes to satisfy user-defined objectives, outputting final compositions together with suggested reaction conditions and a synthetic-feasibility assessment. AI: Artificial intelligence.

The most forward-looking aspect of the work is the coupling of this database layer to an inverse-design agentic interface [Figure 2]. DigHyd is presented not only as a repository, but as an interactive design environment in which user-defined constraints can be translated into candidate generation and iterative screening[3]. That closed path, which goes from literature to structured records to model-assisted proposal, marks a meaningful shift in emphasis: AI is used not merely to summarize what is known, but to propose potential solutions under explicit targets. In this sense, the study offers a concrete example of how multi-agent systems can function as scientific workflow organizers rather than standalone chat interfaces.

The broader implications of this work are clear. Figure-centric knowledge is common across catalysis, batteries, hydrogen materials, and many adjacent domains. A method that systematically converts visual evidence into extraction-ready text, while preserving evaluation rigor and downstream usability, is likely to transfer well beyond the immediate case study. The paper also remains realistic about unresolved issues, including heterogeneous figure quality, incomplete contextual reporting, and residual hallucination risks under sparse evidence. Nevertheless, these limitations do not diminish the core achievement. They define the next engineering frontier: tighter uncertainty control, improved provenance handling, and deeper integration with automated or high-throughput experimentation.

Overall, this work demonstrates that progress in AI-driven materials discovery can come as much from workflow design as from model scaling. By connecting visual interpretation, structured extraction, quality-aware scoring, and inverse design within one coherent system, DIVE and DigHyd provide a credible template for building AI-native discovery pipelines grounded in the real texture of scientific literature.

DECLARATIONS

Authors' contributions

Literature investigation: Hong, Y.; Mao, X.
Original manuscript preparation: Hong, Y.; Mao, X.
Review and editing of the manuscript: Hong, Y.; Mao, X.

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

The authors acknowledge financial support from the Australian Research Council (Grant No. DE260101596).

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

REFERENCES

1. Zhang, D.; Jia, X.; Tran, H. B.; et al. "DIVE" into hydrogen storage materials discovery with AI agents. *Chem. Sci.* **2026**, *17*, 3031-42. DOI PubMed PMC
2. Mavračić, J.; Court, C. J.; Isazawa, T.; Elliott, S. R.; Cole, J. M. ChemDataExtractor 2.0: autopopulated ontologies for materials science. *J. Chem. Inf. Model.* **2021**, *61*, 4280-9. DOI PubMed
3. Jang, S. H.; Zhang, D.; Tran, H. B.; et al. Physically interpretable descriptors drive the materials design of metal hydrides for hydrogen storage. *Chem. Sci.* **2025**, *16*, 23111-20. DOI PubMed PMC

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.