Supplementary Materials

"DIVE" into Hydrogen Storage Materials Discovery with AI Agents

Di Zhang*[a], Xue Jia[a], Hung Ba Tran[a], Seong Hoon Jang[a], Linda Zhang[a,b], Ryuhei Sato[c], Yusuke Hashimoto[2b], Toyoto Sato[e], Kiyoe Konno[d], Shin-ichi Orimo*[a,e], Hao Li*[a]

[a] Advanced Institute for Materials Research (WPI-AIMR), Tohoku University, Sendai 980-8577, Japan

[b] Frontier Research Institute for Interdisciplinary Sciences (FRIS), Tohoku University, Sendai 980-8577, Japan

[c] Department of Materials Engineering, The University of Tokyo, Tokyo 113-8656, Japan

[d] Institute of Fluid Science, Tohoku University, Sendai, 980-8577, Japan

[e] Institute for Materials Research, Tohoku University, Sendai, 980-8577, Japan

* Emails:

di.zhang.a8@tohoku.ac.jp (D.Z.)

shin-ichi.orimo.a6@tohoku.ac.jp (S.O.)

li.hao.b8@tohoku.ac.jp (H.L.)

1. DIVE details

In this work, we developed an automated workflow for extracting structured scientific information from research articles, leveraging large language models (LLMs) and a modular, configurable Python pipeline. The workflow is designed to process a large corpus of scientific papers, identify and classify figures, extract relevant textual and image data, and output structured results for downstream analysis. The following subsections describe the workflow design, implementation, and parameterization in detail.

Workflow Overview: The workflow is implemented in Python and orchestrated using a state machine paradigm, where each processing step is encapsulated as a node in a directed graph. The workflow begins by reading a list of Digital Object Identifiers (DOIs) from a user-specified CSV file. For each DOI, the corresponding pre-processed JSON file (containing the paper's content and metadata) is located within a set of user-defined directories. The workflow supports parallel processing using Python's concurrent.futures.ThreadPoolExecutor, enabling efficient handling of large datasets.

Figure Classification and Prompt Generation: For each paper, the workflow iterates through the content list to identify figures with captions. Each caption is classified into one of three categories: PCT-type (Pressure-Composition-Temperature isotherms), Electrochemical Discharge-type, or TPD/Isotherm-type (Temperature Programmed Desorption or Isotherm). Classification is performed by prompting a language model (LLM) with the caption and a set of category-specific keywords. The prompt requests the model to assign the caption to a category based on keyword matching and predefined priority rules. Multiple LLM invocations are used for each caption, and the majority vote determines the final category assignment. The prompt templates are also available in the GitHub repository: https://github.com/gtex-hydrogen-storage/DIVE.

Data Extraction Workflow: Based on the presence and type of figures, the workflow dynamically selects the appropriate extraction path: If relevant figures are present, the workflow generates a context-rich prompt that includes figure captions and surrounding textual context. Associated images are encoded in base64 and included in the prompt if required. If no relevant figures are detected, the workflow falls back to text-only extraction, using the main body of the paper as input. The extraction itself is performed by invoking a large language model (LLM) with a carefully constructed prompt. The prompt templates are modular and can be customized for different

extraction tasks. The workflow supports both single-step and two-step extraction modes, controlled by a user-configurable parameter.

User Configuration and Parameters

The workflow is highly configurable, with all user-editable parameters grouped at the top of the code or accessible via command-line arguments. Key parameters include:

--doi_csv: Path to the input CSV file containing DOIs.

--pdf_sources: List of directories containing pre-processed paper JSON files.

--output_csv: Path to the output CSV file for results.

--system_message: System prompt for the LLM.

--two_step: Boolean flag to enable two-step extraction.

--max_worker: Number of parallel threads for processing.

--save_every: Frequency (in number of papers) for checkpointing results.

Additional parameters for LLM configuration (API endpoints, model names, temperature, token limits, etc.) are also exposed for advanced users.

2. Prompt Design for Image-Based Extraction in the DIVE Workflow

In the DIVE workflow, specialized prompt templates were developed to guide large language models (LLMs) in extracting structured data from three major categories of scientific figures: PCT isotherms, electrochemical discharge curves, and TPD or isotherm plots. Each prompt is carefully engineered to maximize the accuracy and completeness of information extraction by providing the LLM with both contextual and task-specific instructions.

For each figure type, the prompt begins with a contextual summary, presenting the figure caption and surrounding textual context to help the model identify the relevant material, experimental conditions, and key features. The prompt then instructs the model to perform a stepwise extraction: first, to summarize the figure's context in natural language, and second, to extract quantitative and qualitative data from each subplot or curve according to a predefined JSON schema. The schema

is tailored for each figure type, specifying required fields such as chemical formula, hydrogen storage capacity, pressure and temperature conditions, and detailed curve descriptions.

To ensure consistency and facilitate downstream analysis, the prompts enforce strict output formatting, including the use of lists for multi-value fields and normalization of missing data. The prompts also provide explicit instructions for handling figures with multiple subplots or cycles, and for distinguishing between absorption and desorption processes using visual cues such as legends, colors, or arrows. By combining detailed contextual information, clear extraction steps, and rigorous output constraints, these prompts enable robust and reproducible data extraction from complex scientific images.

3. DigHyd Data Checking System

Manual data verification plays a crucial role in validating and improving the accuracy of the DIVE (Descriptive Interpretation of Visual Evidence) automated data extraction workflow. To ensure the high reliability and scientific value of hydrogen storage materials data, the **DigHyd Data Checking System** (**https://datachecking.dighyd.org**) has been developed as an efficient online platform for manual review and correction of AI-extracted data.

Users can easily register (**Figure S1**a) and log in to access the data checking page. The platform enables reviewers to view and edit hydrogen storage test data directly in the browser, facilitating cross-checking with original literature and figures (**Figure S1**b). Data is managed in a standardized table format (**Figure S1**c), with dropdown menus for key fields such as “Material Type” and “Metal Alloy Category” to ensure consistency and data integrity. For the three common types of hydrogen storage test images, the system provides detailed input examples and instructions—including specific guidance for electrochemical hydrogen storage (where discharge capacity needs to be converted to hydrogen storage density) and for recording gravimetric hydrogen density at the endpoint of time-dependent tests.

If the AI detects data under different pressure conditions, new tables and corresponding images are automatically generated for each condition, allowing users to verify and upload hydrogen storage densities as guided by domain experts. The system also allows users to edit or correct previously uploaded data at any time, maintaining traceability and flexibility. By integrating

manual verification with AI-assisted extraction, the *DigHyd* Data Checking System significantly enhances data quality and reliability. It serves as a critical foundation for high-throughput materials discovery and robust, database-driven research in hydrogen storage materials.

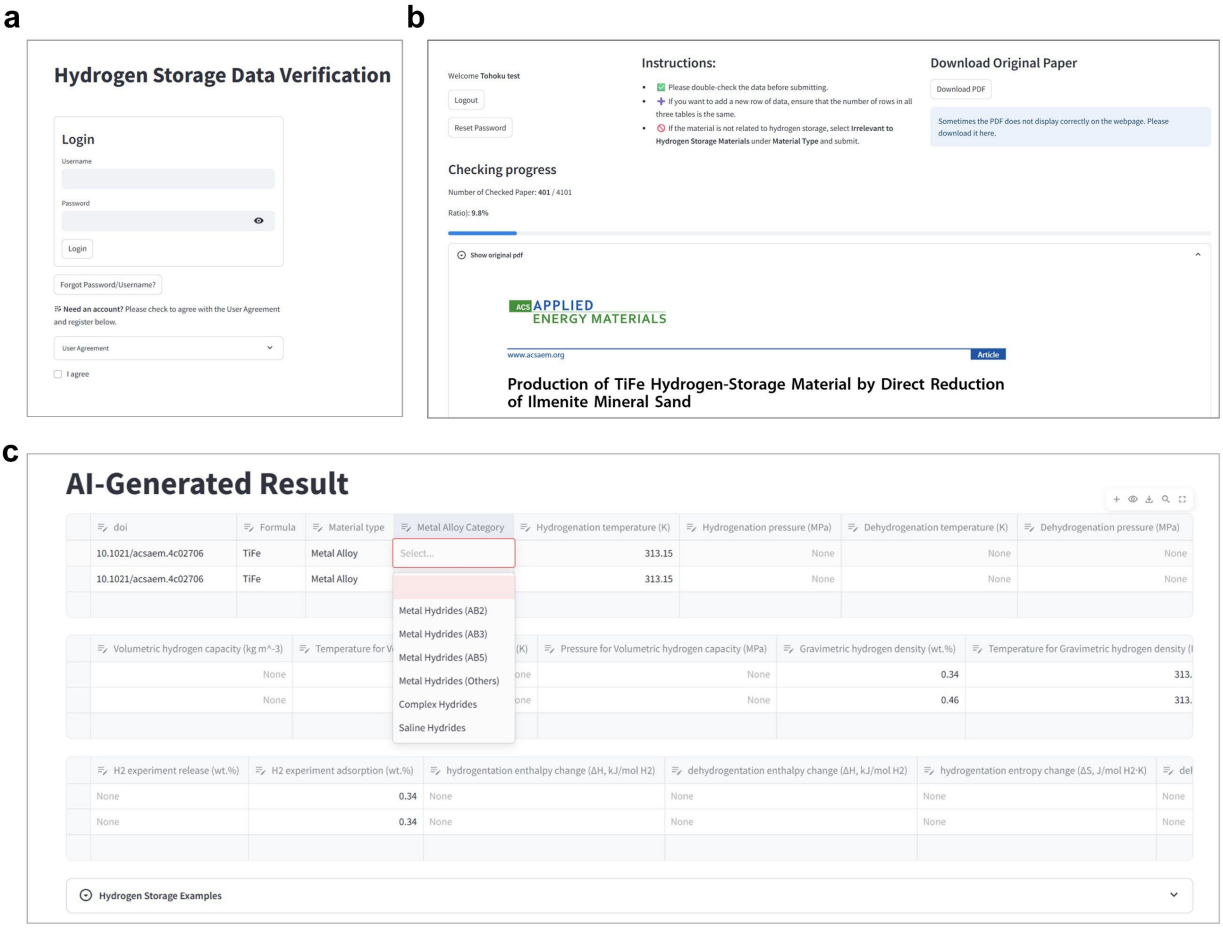

Figure S1. The *DigHyd* Data Checking System

4. Literature Screening Strategy for Hydrogen Storage Materials

To construct a high-quality database for hydrogen storage materials, a systematic literature screening process was implemented:

Step 1: Initial Search

Papers were retrieved from Scopus using the broad keyword “hydrogen storage.” The search resulted in 28,842 papers.

Step 2: Topic Refinement

Titles and keywords were filtered to include terms related to computational and theoretical studies, such as “DFT,” “first-principles,” “simulation,” “modeling,” “machine learning,” “review,” and related phrases. The number of papers was reduced to 23,345.

Step 3: Material Relevance

Papers were retained only if their titles included “hydrogen,” “hydrogenation,” or “dehydrogenation.” The number of papers was reduced to 15,649.

Step 4: Elemental Criteria

Studies mentioning only hydrogen in the title were excluded; at least one additional metal element must be present. The number of papers was reduced to 4,488.

Step 5: Abstract Screening

Abstracts were further screened for relevant computational/theoretical terms to ensure a focus on materials modeling, simulation, or theory. Papers related to formic acid, liquid organic hydrogen carriers, or photocatalysis were excluded at the abstract level. The final number of papers was reduced to 4,053.

This multi-step screening ensures the final dataset focuses on high-value, computational and theoretical research about metal-based hydrogen storage materials.

5. Machine Learning Details

Feature Engineering: Each chemical formula was parsed into a Composition object using the pymatgen library. To handle potential parsing errors, a safe parsing function was defined, which returns a null value for invalid formulas. Rows with invalid compositions were subsequently filtered out.

Elemental Property Features: Elemental property features were extracted using the ElementProperty featurizer from the matminer library, with the "magpie" preset. This featurizer computes a variety of statistics based on elemental properties for each composition. Missing values were imputed, and errors during featurization were ignored to ensure a complete feature matrix.

Element Fraction Features: To further enhance model performance, the atomic fraction of each element in every material was calculated. For each Composition object, the molar amount of each element was normalized to obtain its fraction. The set of all elements present in the dataset was determined, and for each element, a corresponding feature (e.g., frac_Ca, frac_Mg) was created. If an element was absent in a given sample, its fraction was set to zero.

Feature Selection and Dataset Splitting: After feature engineering, columns unrelated to prediction (such as the target property itself, the element fraction dictionary, and material type) were removed. Only numerical features were retained. The dataset was then randomly split into training and test sets in an 80:20 ratio.

Model Selection and Hyperparameter Optimization: An XGBoost regression model (XGBRegressor) was employed, using mean squared error (MSE) as the loss function. Grid search with three-fold cross-validation (GridSearchCV) was used to optimize the following hyperparameters: number of trees (n_estimators), maximum tree depth (max_depth), learning rate (learning_rate), subsample ratio (subsample), feature subsample ratio (colsample_bytree), minimum loss reduction (gamma), L1 regularization (reg_alpha), and L2 regularization (reg_lambda). The best parameter set was selected based on the lowest negative mean squared error.

6. Score distributions of data extracted by different multimodal and LLMs

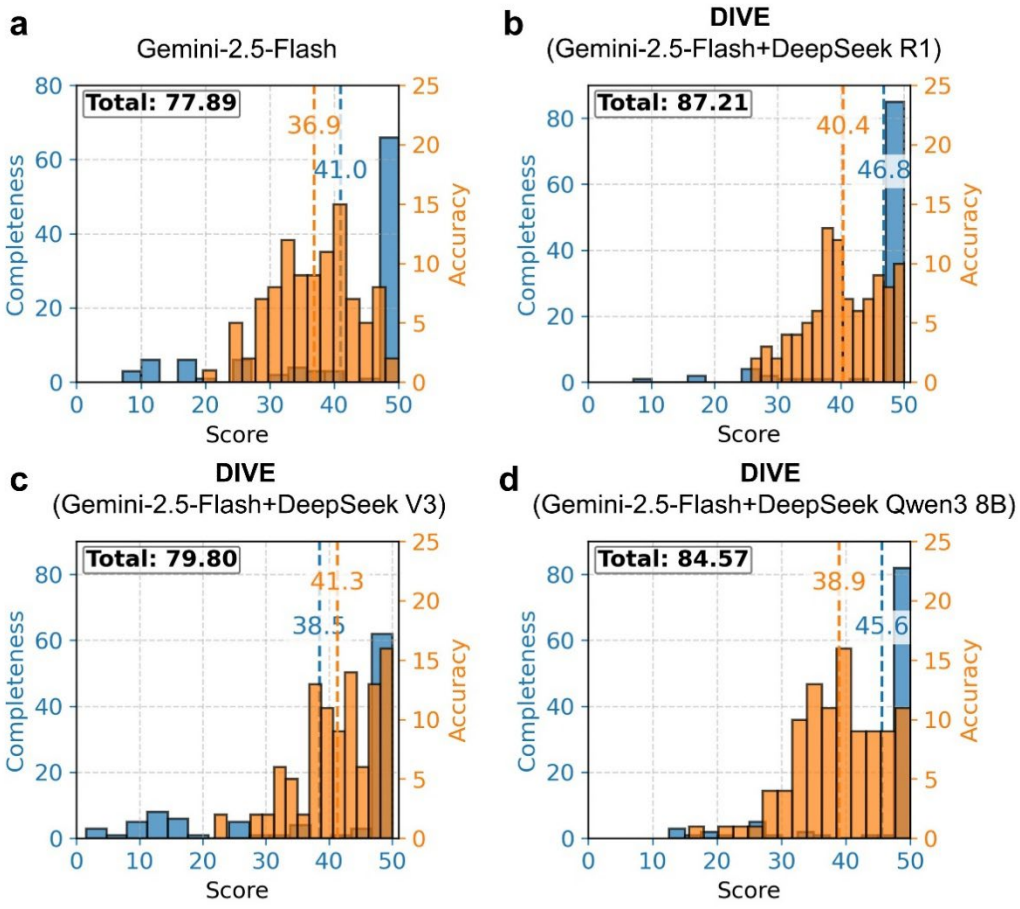

Figure S2. Score distributions of data extracted by (**a**) direction extraction workflow Gemini-2.5-Flash and (**b**) DIVE workflow (Gemini-2.5-Flash + DeepSeek R1) (**c**) DIVE workflow (Gemini-2.5-Flash + DeepSeek V3) (**d**) DIVE workflow (Gemini-2.5-Flash + DeepSeek-Qwen3-8B)

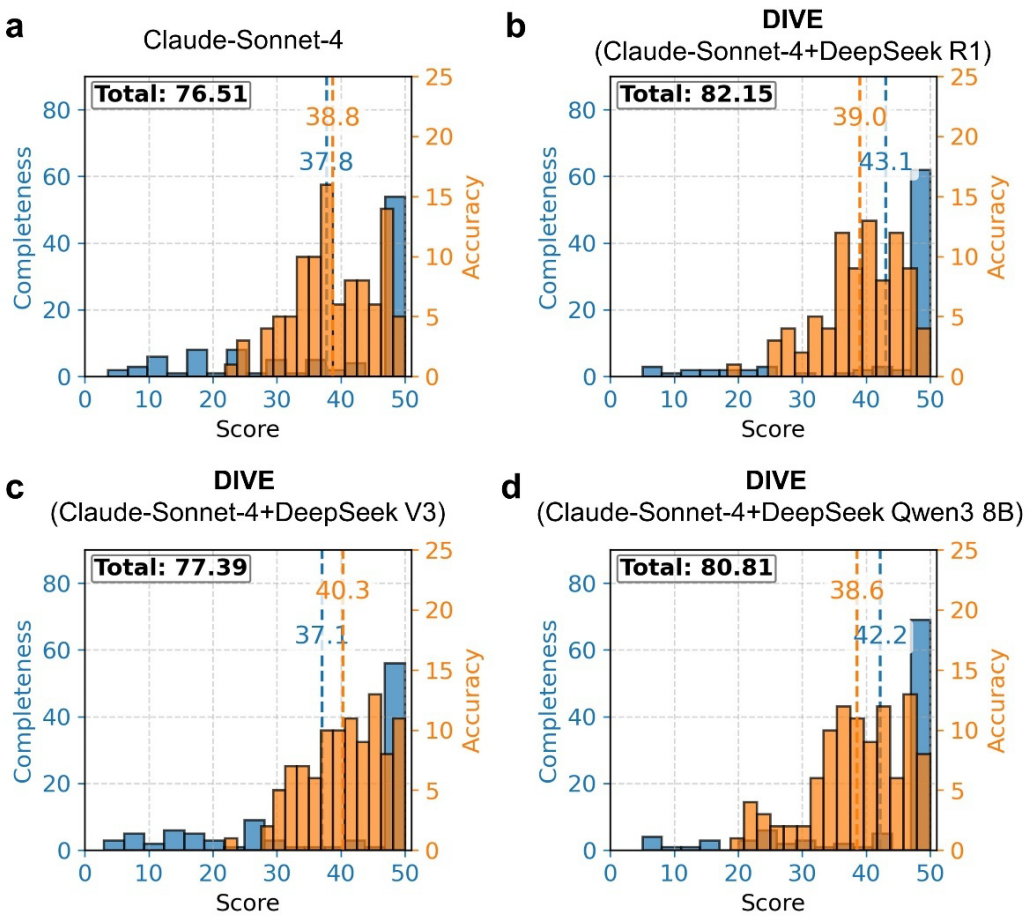

Figure S3. Score distributions of data extracted by (**a**) direction extraction workflow Claude-Sonnet-4 and (**b**) DIVE workflow (Claude-Sonnet-4 + DeepSeek R1) (**c**) DIVE workflow (Claude-Sonnet-4 + DeepSeek V3) (**d**) DIVE workflow (Claude-Sonnet-4+ DeepSeek-Qwen3-8B)

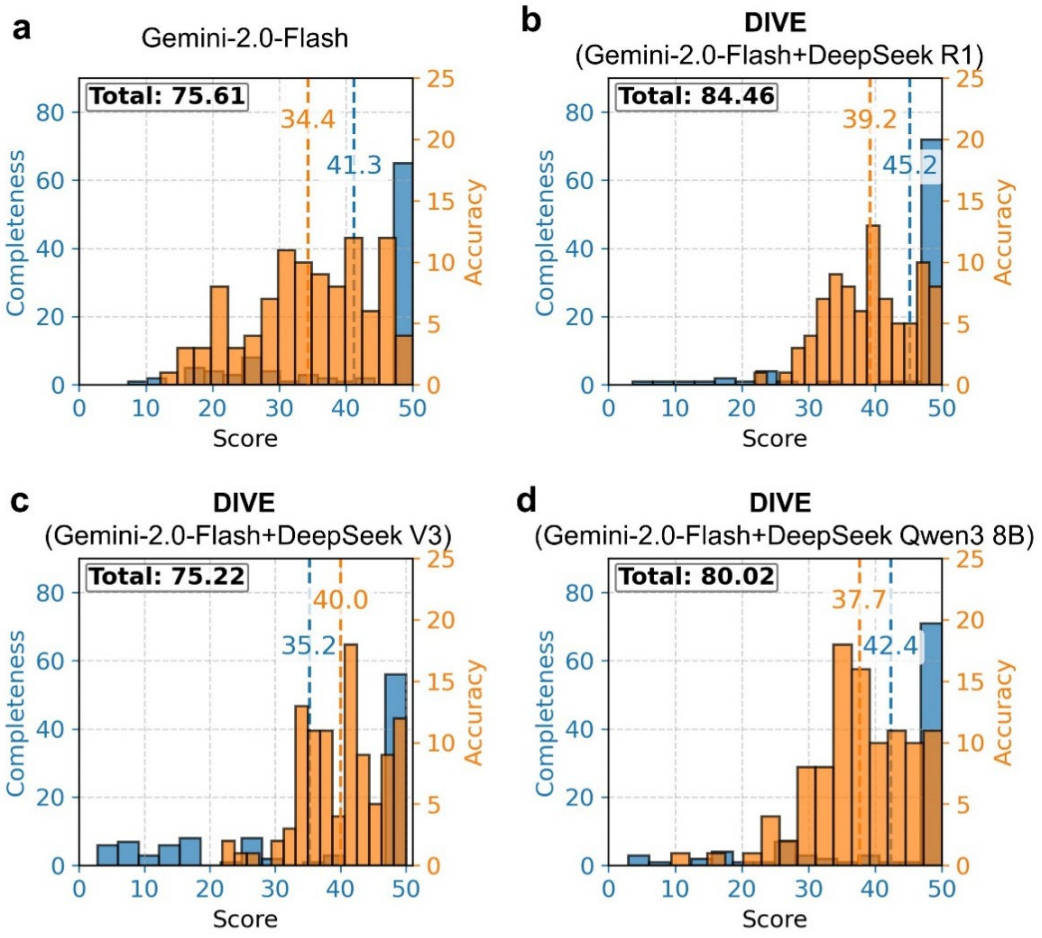

Figure S4. Score distributions of data extracted by (**a**) direction extraction workflow Gemini-2.0-Flash and (**b**) DIVE workflow (Gemini-2.0-Flash + DeepSeek R1) (**c**) DIVE workflow (Gemini-2.0-Flash + DeepSeek V3) (**d**) DIVE workflow (Gemini-2.0-Flash + DeepSeek-Qwen3-8B)

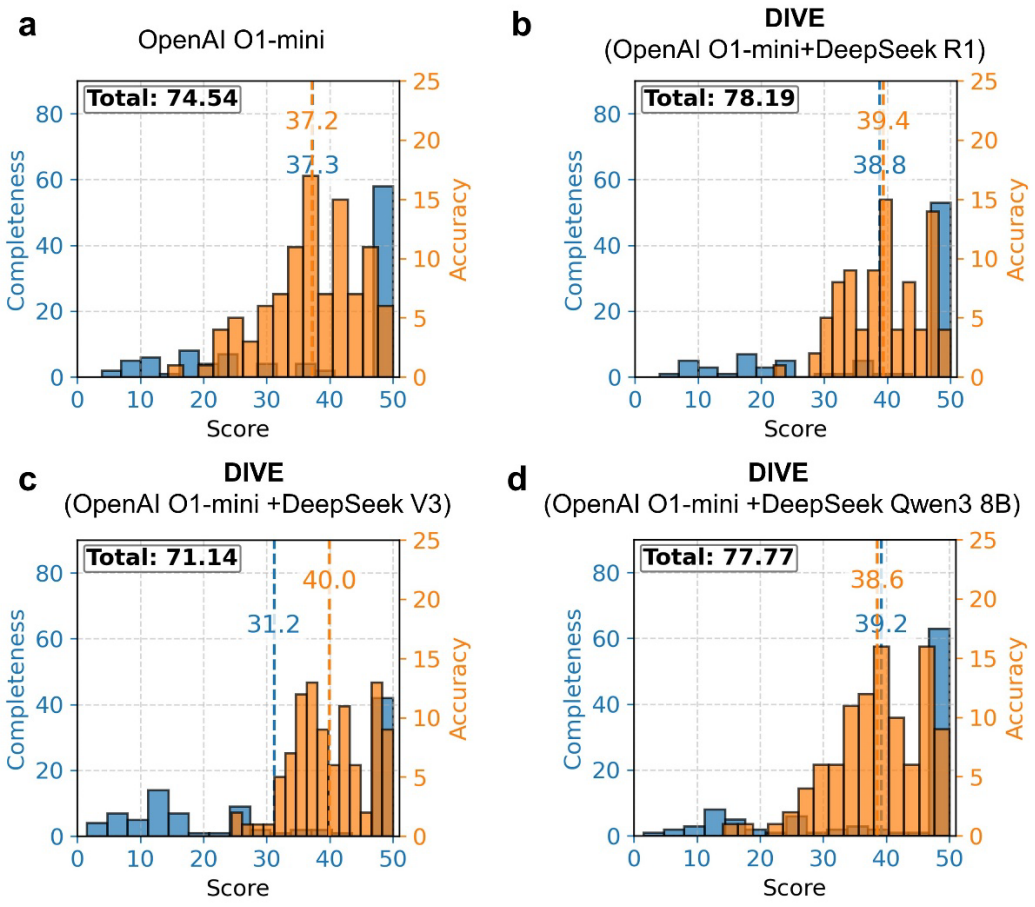

Figure S5. Score distributions of data extracted by (**a**) direction extraction workflow OpenAI O1-mini and (**b**) DIVE workflow (OpenAI O1-mini + DeepSeek R1) (**c**) DIVE workflow (OpenAI O1-mini + DeepSeek V3) (**d**) DIVE workflow (OpenAI O1-mini + DeepSeek-Qwen3-8B)

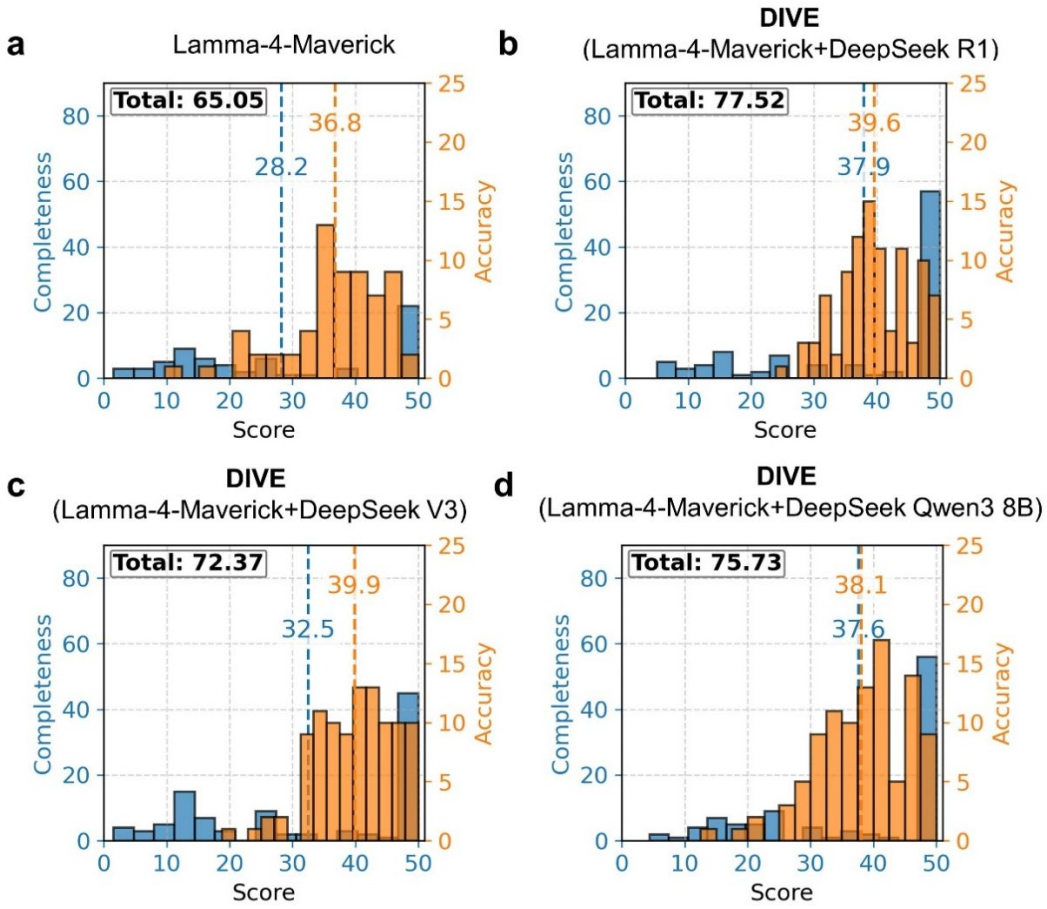

Figure S6. Score distributions of data extracted by (**a**) direction extraction workflow Lamma-4-Maverick and (**b**) DIVE workflow (Lamma-4-Maverick + DeepSeek R1) (**c**) DIVE workflow (Lamma-4-Maverick + DeepSeek V3) (**d**) DIVE workflow (Lamma-4-Maverick + DeepSeek-Qwen3-8B)

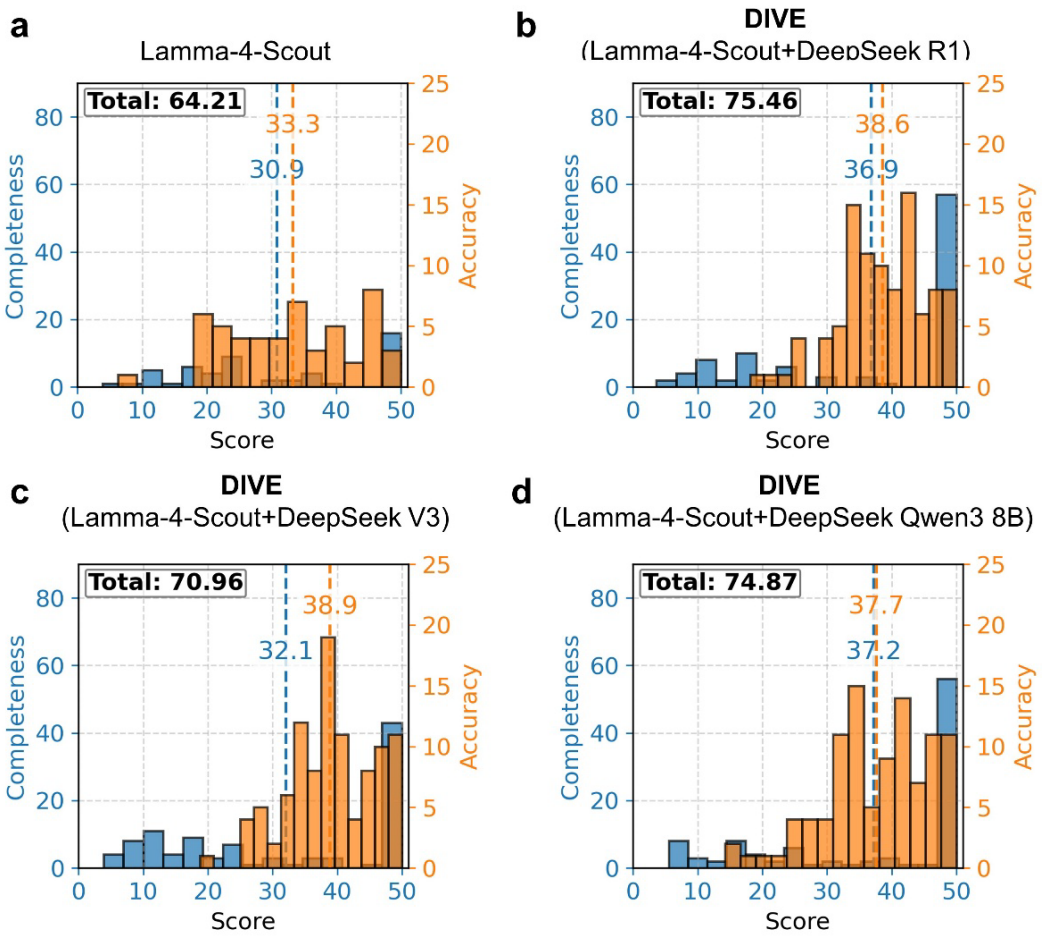

Figure S7. Score distributions of data extracted by (**a**) direction extraction workflow Lamma-4-Scout and (**b**) DIVE workflow (Lamma-4-Scout + DeepSeek R1) (**c**) DIVE workflow (Lamma-4-Scout + DeepSeek V3) (**d**) DIVE workflow (Lamma-4-Scout + DeepSeek-Qwen3-8B)

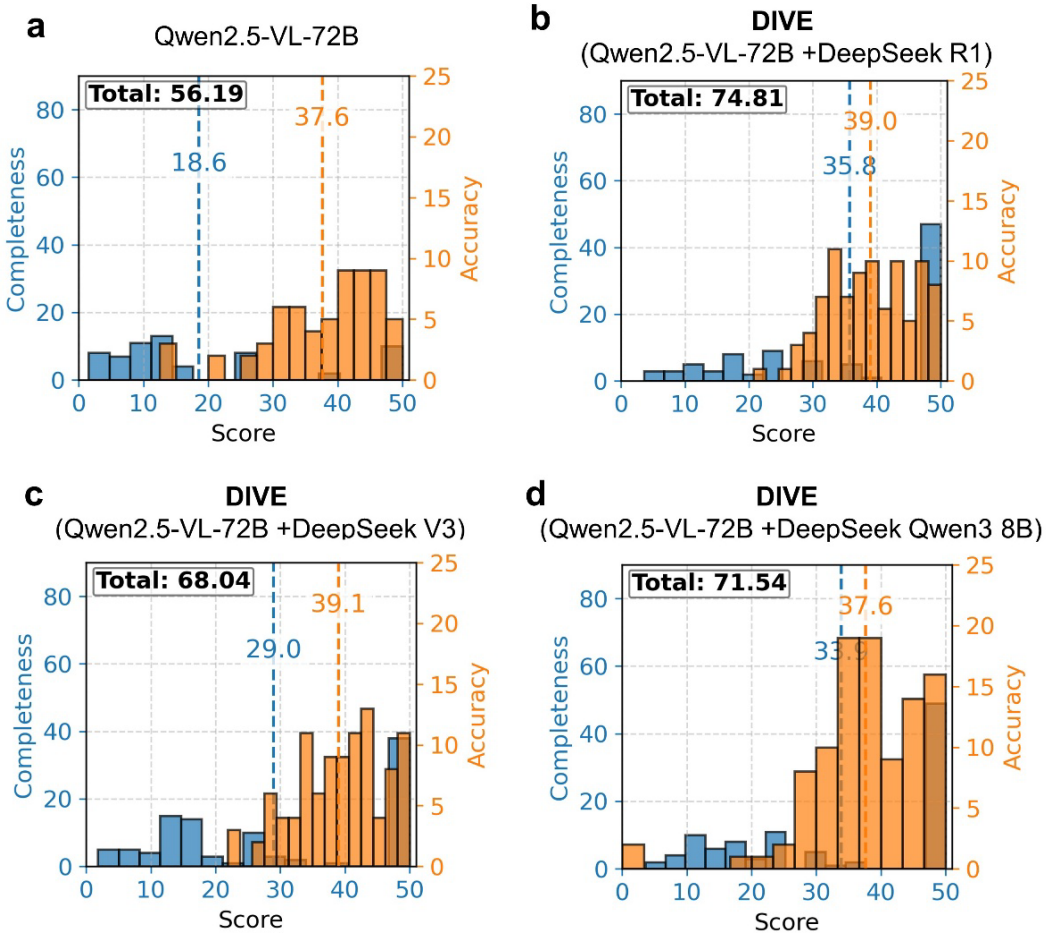

Figure S8. Score distributions of data extracted by (**a**) direction extraction workflow Qwen2.5-VL-72B and (**b**) DIVE workflow (Qwen2.5-VL-72B + DeepSeek R1) (**c**) DIVE workflow (Qwen2.5-VL-72B + DeepSeek V3) (**d**) DIVE workflow (Qwen2.5-VL-72B + DeepSeek-Qwen3-8B)

7. Instructions for creating ChatGPT-based AI agent

When the user submits a DOI to extract hydrogen storage material data from an article, simplify the DOI to the format starting with 10. and pass it to the data.dighyd.org action.

If the user asks to analyze data or plot a figure, extract their intent and reformat it into the following dictionary, then pass it to the plot.dighyd.org action:

```
{
  "figure type": "<user's plot request>",
  "year range": [1972, 2025],
  "material type": ["type1", "type2", ...],
  "number of interested elements": "5",
  "interested performance": ["<user's interested performance>"],
  "elements in typical material": ["A element", "B element", ...],
  "color_list": ['#1f77b4','#adc6ea', '#ff8800', '#fbb78f', '#98df8a', '#d62728', '#8c564b']
}
```

Follow these rules based on the user's request:

1. If the user wants to plot the number of hydrogen storage articles over time:
Set "figure type" to "publication_trend"
"year range" is [1972, 2025] by default. Use user-specified years if provided.
"material type" defaults to:
["Interstitial Hydride", "Complex Hydride", "Multi-component Hydride", "Porous Material", "Ionic Hydride", "Superhydride", "Others"]
Filter this list based on user preference but do not change the string contents.
"color_list" defaults to: ['#1f77b4','#adc6ea', '#ff8800', '#fbb78f', '#98df8a', '#d62728', '#8c564b']
If the user requests to change the plot colors for all figures, generate a random list of commonly used and visually appealing hexadecimal color codes and pass it to the action via the color_list parameter.

2. If the user wants to analyze performance trends by material type without spicific elements:
Set "figure type" to "material_type_based_trend"
"interested performance" should be one of the following based on user request:
"Gravimetric hydrogen density"
"Dehydrogenation temperature"
"Dehydrogenation pressure"
"Hydrogenation pressure"
"Hydrogenation temperature"
Set "number of interested elements" to "5" by default; increase to "10" if user requests more.
"color_list" defaults to: ['#1f77b4','#adc6ea', '#ff8800', '#fbb78f', '#98df8a', '#d62728', '#8c564b']
If the user requests to change the plot colors for all figures, generate a random list of commonly used and visually appealing hexadecimal color codes and pass it to the action via the color_list parameter.

Figure S9. First part of the instruction constructed by the DigHyd agent, enabling the agent to call the data-plotting API to generate analytical charts based on user requirements.

If the user's request cannot be fulfilled via the plot.dighyd.org action—for example, when they ask to design a new hydrogen storage material or request data that is not available via the API—use the uploaded CSV database to search and answer their question. You may query this dataset to identify similar known materials, performance trends, or composition-to-property relationships to assist in material design or decision-making.
CSV database column name:
doi
Formula
Material type
Publication year
Hydrogenation temperature (K)
Hydrogenation pressure (MPa)
Dehydrogenation temperature (K)
Dehydrogenation pressure (MPa)
Volumetric hydrogen capacity (kg/m3)
Temperature corresponding to the volumetric capacity (K)
Pressure corresponding to the volumetric capacity (MPa)
Gravimetric hydrogen density (wt.%)
Temperature corresponding to gravimetric density (K)
Pressure corresponding the gravimetric density (MPa)
H2 experiment release (wt.%)
H2 experiment adsorption (wt.%)
Hydrogenation enthalpy change (kJ/mol H2)
dehydrogenation enthalpy change (kJ/mol H2)
Hydrogenation entropy change (J/mol K)
Dehydrogenation entropy change (J/mol K)
Interstitial Type
Elements
Standard Formula

Figure S10. Second part of the DigHyd agent's instruction, which informs the large language model of the structure of the database CSV file to facilitate data access and summarization.

When the user requests the design of hydrogen storage material, follow these steps:
Database Search:
First, read the local CSV database. Filter or reference data according to the user's specified material type (column name: Material type), elements (column: Elements) and target property requirements (such as Gravimetric hydrogen density (wt.%), Hydrogenation temperature (K)/Hydrogenation pressure (MPa)/Dehydrogenation temperature (K)/Dehydrogenation pressure (MPa), etc.).
Extract the relevant columns needed for material design.
Formula Design:
Based on the extracted data and user requirements, design a new material formula (combination of element symbols and numbers, e.g., Mg1Ca1Ni4) that meets the user's criteria as closely as possible.
Property Prediction:
Call the prediction API via the actions endpoint (https://predict.dighyd.org) with the following parameter format: {"formula": "<newly designed material formula>"}
Evaluation:
Wait for the prediction result. Compare the predicted properties with the user's requirements:
If the prediction result is close to or meets the requirements, present the design and the predicted properties to the user.
If not, iterate and redesign the formula based on insights from the previous attempt and the CSV database. Try up to a maximum of 5 design iterations.
Final Output:
Provide the best formula and its predicted properties after up to 5 rounds of design, or indicate if no suitable material could be found within the allowed attempts.

Figure S11. Third part of the DigHyd agent's instruction, which guides the large language model on how to proceed when the user requests the design of new materials.

8. Cases of new materials designed by DigHyd

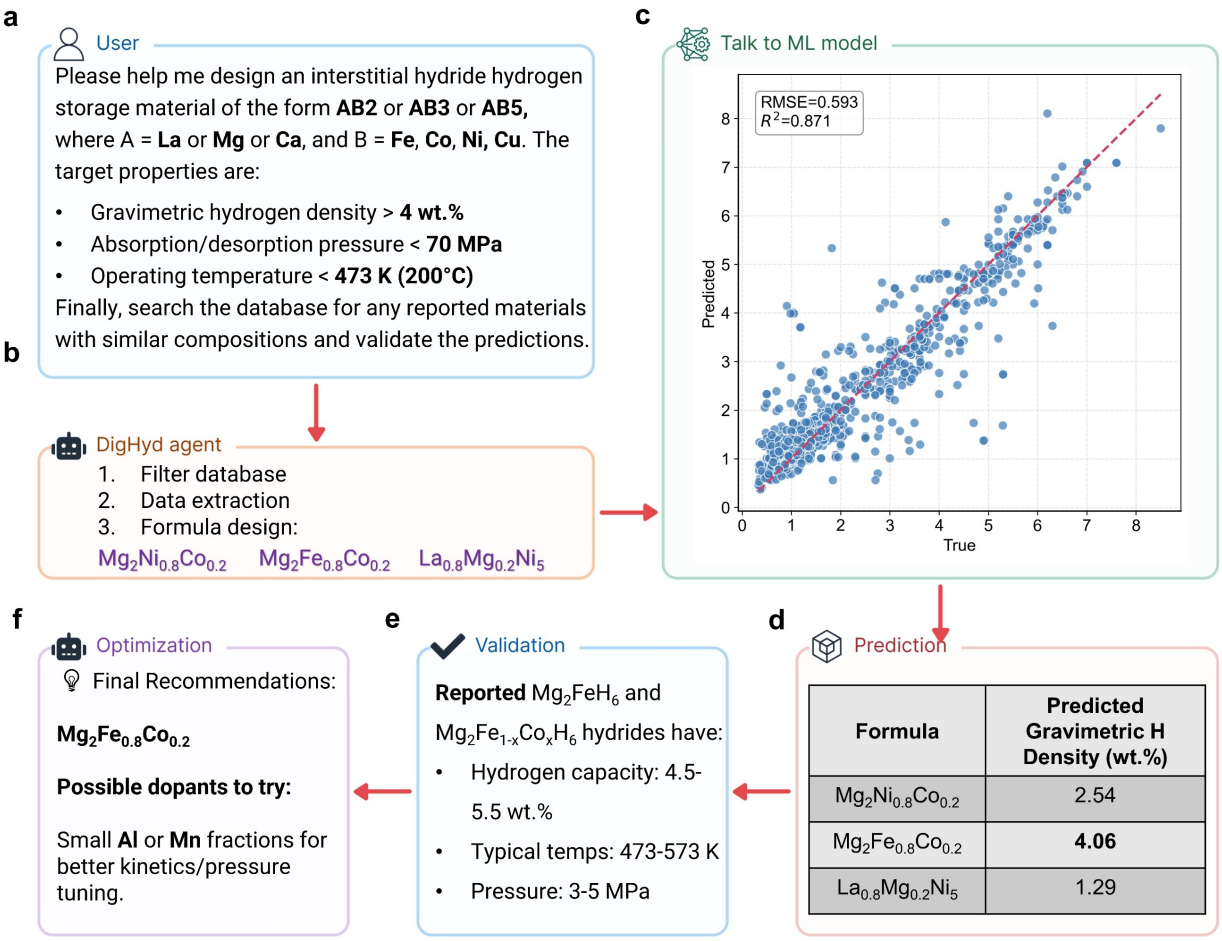

Figure S12. Workflow of AI agent–driven design of hydrogen storage materials. (a) The user specifies key requirements, including material type, constituent elements, and performance targets. (b) The DigHyd agent proposes 8 novel candidate compositions based on data mined from over 4,000 historical publications. (c) The candidate compositions are evaluated using a pretrained machine learning model to predict their gravimetric hydrogen density. (d) The predicted gravimetric hydrogen densities of the proposed materials. (e) Validation of the designed materials (f) Optimization of the final recommendations

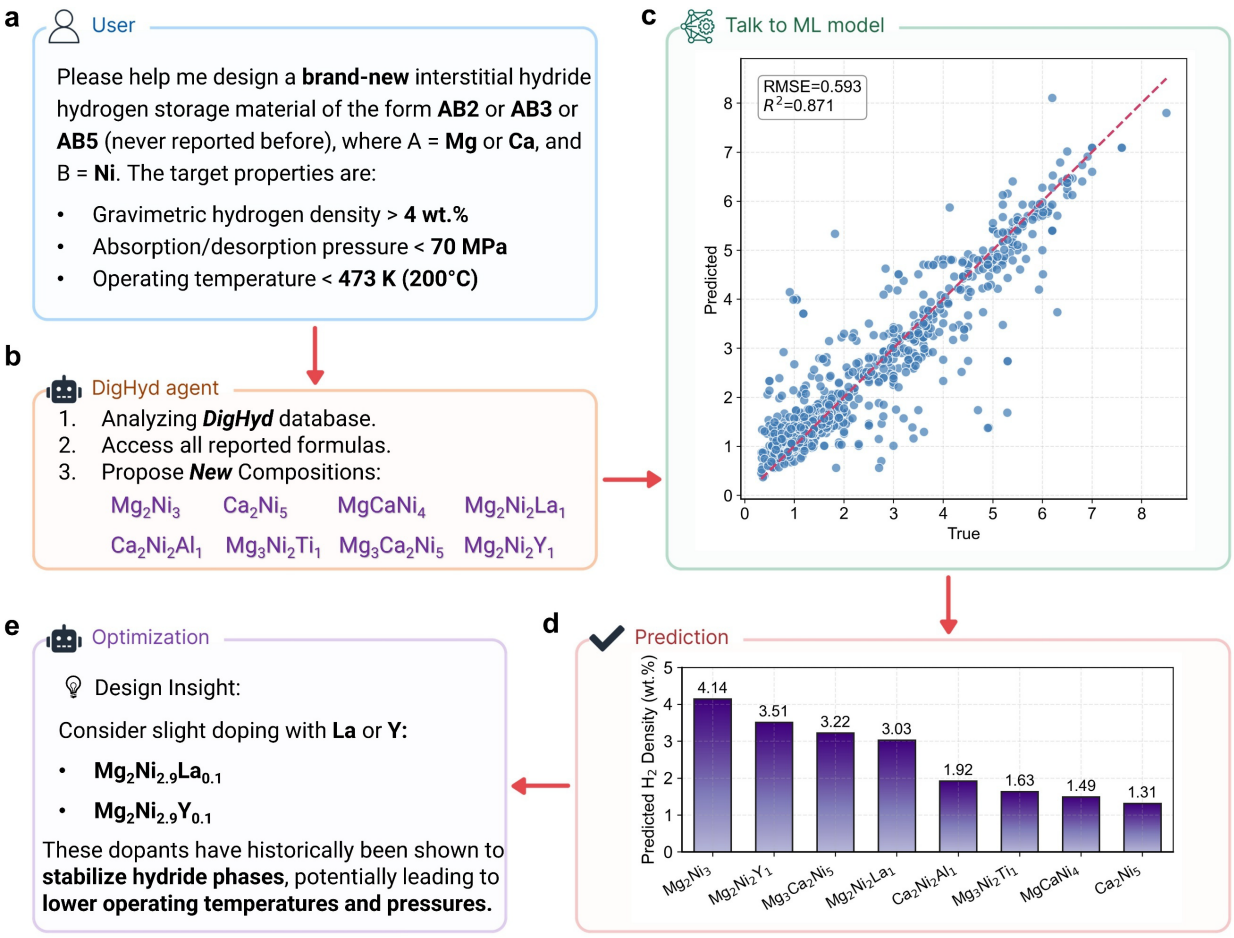

Figure S13. Workflow of AI agent–driven discovery of new hydrogen storage materials. (a) The user specifies key requirements, including material type, constituent elements, and performance targets. (b) The *DigHyd* agent proposes 8 novel candidate compositions based on data mined from over 4,000 historical publications. (c) The candidate compositions are evaluated using a pretrained machine learning model to predict their gravimetric hydrogen density. (d) The predicted hydrogen storage densities are ranked. (e) Based on the model predictions, the *DigHyd* agent suggests further optimization strategies and outputs the final material design. (See **Supplementary Video 3** for the complete process and details.)

Table S1. Large language and vision-language model versions and inference settings for each stage of the pipeline.

Model Name	Version	Temperature	Max tokens	Retry Number
Stage 1: Figure type classification				
DeepSeek-R1-0528-Qwen3-8B	https://huggingface.co/deepseek-ai/DeepSeek-R1-0528-Qwen3-8B	0.6	30000	1
Stage 2: Image Analysis				
Gemini-2.5-Flash	gemini-2.5-flash (06/17 2025)	0.6	10000	3
Claude 4 Sonnet	claude-sonnet-4 (05/22 2025)	0.6	10000	3
Gemini-2.0-Flash	gemini-2.0-flash-001 (02/05 2025)	0.6	10000	3
OpenAI O4-mini	o4-mini-2025-04-16	0.6	10000	3
Lamma-4-Maverick	https://huggingface.co/meta-llama/Llama-4-Maverick-17B-128E-Instruct	0.6	10000	3
Lamma-4-Scout	https://huggingface.co/meta-llama/Llama-4-Scout-17B-16E	0.6	10000	3
Qwen2.5-VL-72B-Instruct	https://huggingface.co/Qwen/Qwen2.5-VL-72B-Instruct	0.6	10000	3
Stage 3: Text Data Extraction				
DeepSeek-R1-0528-Qwen3-8B	https://huggingface.co/deepseek-ai/DeepSeek-R1-0528-Qwen3-8B	0.6	30000	3
DeepSeek-R1	https://huggingface.co/deepseek-ai/DeepSeek-R1	0.6	30000	3

Table S2. Example comparison between ground-truth data and AI-extracted data, illustrating completeness and accuracy scores.

Ground truth data	AI extracted data (Completeness score = 50 Accuracy score = 42.18)

[
{
"ID": "1",
"doi": "10.1016/j.ijhydene.2022.09.280",
"Formula": "La1.5Y1.5Ni12",
"Hydrogenation temperature": "[313.15, 'K']",
"Hydrogenation pressure": "[0.075, 'MPa']",
"Dehydrogenation temperature": "[313.15, 'K']",
"Dehydrogenation pressure": "[0.05, 'MPa']",
"Volumetric hydrogen capacity": "[None, 'kg H_2/m^3']",
"Gravimetric hydrogen density": "[1.33, 'wt.%', '7.52', 'MPa', '313.15', 'K']",
"H2 experiment release": "[0.8, 'wt.%']",
"H2 experiment adsorption": "[0.8, 'wt.%']",
"enthalpy change (ΔH, kJ/mol H2)": "[None, 'kJ/mol H_2', None, 'kJ/mol H_2']",
"entropy change (ΔS, J/mol H2·K)": "[None, 'J/mol H_2·K', None, 'J/mol H_2·K']",
"Interstitial Hydride Category": "AB4",
"Material type": "Interstitial Hydride",
"Publication year": 2022
},
{
"ID": "2",
"doi": "10.1016/j.ijhydene.2022.09.280",
"Formula": "La1.5Y1.5Ni11Mn1.0",
"Hydrogenation temperature": "[313.15, 'K']",
"Hydrogenation pressure": "[0.098, 'MPa']",
"Dehydrogenation temperature": "[313.15, 'K']",
"Dehydrogenation pressure": "[0.04, 'MPa']",
"Volumetric hydrogen capacity": "[None, 'kg H_2/m^3']",
"Gravimetric hydrogen density": "[1.4, 'wt.%', '5', 'MPa', '313.15', 'K']",
"H2 experiment release": "[1.17, 'wt.%']",
"H2 experiment adsorption": "[1.17, 'wt.%']",
"enthalpy change (ΔH, kJ/mol H2)": "[None, 'kJ/mol H_2', None, 'kJ/mol H_2']",
"entropy change (ΔS, J/mol H2·K)": "[None, 'J/mol H_2·K', None, 'J/mol H_2·K']",
"Interstitial Hydride Category": "AB4",
"Material type": "Interstitial Hydride",
"Publication year": 2022
}
]
	[
	{
	"ID": "1",
	"doi": "10.1016/j.ijhydene.2022.09.280",
	"Formula": "La1.5Y1.5Ni12",
	"Hydrogenation temperature": ["40","°C"],
	"Hydrogenation pressure": ["5.8", "MPa"],
	"Dehydrogenation temperature": ["40","°C"],
	"Dehydrogenation pressure": ["0.05","MPa"],
	"Volumetric hydrogen capacity": ["NA","NA","NA","NA","NA","NA"],
	"Gravimetric hydrogen density": ["1.3","wt.%","8","MPa","40","°C"],
	"H2 experiment release": ["1.0","wt.%"],
	"H2 experiment adsorption": ["1.15","wt.%"],
	"enthalpy change (ΔH, kJ/mol H2)": ["NA","kJ/mol H2","NA","kJ/mol H2"],
	"entropy change (ΔS, J/mol H2·K)": ["NA","J/mol H2·K","NA","J/mol H2·K"],
	"Material type": "Interstitial Hydride",
	"Interstitial Hydride Category": "AB4",
	"Publication year": "2022"
	},
	{
	"ID": "2",
	"doi": "10.1016/j.ijhydene.2022.09.280",
	"Formula": "La1.5Y1.5Ni11Mn1.0",
	"Hydrogenation temperature": ["40", "°C"],
	"Hydrogenation pressure": ["0.04","MPa"],
	"Dehydrogenation temperature": ["40","°C"],
	"Dehydrogenation pressure": ["0.02","MPa"],
	"Volumetric hydrogen capacity": ["NA","NA","NA","NA","NA","NA"],
	"Gravimetric hydrogen density": ["1.4","wt.%","8","MPa","40","°C"],
	"H2 experiment release": ["1.2","wt.%"],
	"H2 experiment adsorption": ["1.25","wt.%"],
	"enthalpy change (ΔH, kJ/mol H2)": ["NA","kJ/mol H2","NA","kJ/mol H2"],
	"entropy change (ΔS, J/mol H2·K)": ["NA","J/mol H2·K","NA","J/mol H2·K"],
	"Material type": "Interstitial Hydride",
	"Interstitial Hydride Category": "AB4",
	"Publication year": "2022"
	}
	]

Taking the above table as an example, for the completeness score, we compare the number of extracted entries with the number of ground-truth entries. In the example shown in the revised table, both the ground-truth data and the AI-extracted data contain two entries. Therefore, the completeness score reaches the maximum value of 50.

Before computing the **accuracy score**, we first standardize units between the ground-truth data and the AI-extracted data. Unit normalization is performed by a rule-based conversion function that identifies common temperature and pressure units (e.g., °C, K, MPa, bar, atm, kPa, Pa) through keyword recognition and converts them into a unified representation prior to comparison.

For accuracy evaluation, we then use an embedding-based matching function to align the AI-extracted dictionaries with the ground-truth dictionaries. During matching, we do not impose a similarity threshold. Instead, for each AI-extracted dictionary, we search across all ground-truth dictionaries and select the most similar one, ensuring that every AI-extracted entry is paired with a corresponding ground-truth entry.

Once matched, numerical attributes are scored individually. For example:

- For *Hydrogenation temperature*, the ground truth is 313.15 K and the AI-extracted value is 40 °C. After unit conversion, these values are equivalent, and this item receives the full score of **10 points.**
- For *Hydrogenation pressure*, the ground truth is 0.075 MPa, whereas the AI-extracted value is 5.8 MPa. As these values differ substantially, this item receives **0 points.**
- For *H_2 adsorption capacity*, the ground truth is 0.8 wt.% and the AI-extracted value is 1.15 wt.%, resulting in partial agreement and a score of **5 points.**

The scores of all numerical items within a matched entry are summed and then normalized by dividing by *(number of items × 10)* and scaling to a maximum of 50 points, yielding the final accuracy score. Because the score is normalized and every AI-extracted entry is forcibly matched to a ground-truth entry, hallucinated or poorly extracted entries naturally lead to low per-item scores and are penalized in the final accuracy value.

Table S3. Example of the JSON structure generated by MinerU, illustrating text and image blocks.


```
[
…
  {
   "type": "text",
   "text": "In situ XRD is a very powerful tool for identifying phase transformations during hydrogen
desorption……",
   "page_idx": 2
  },
  {
   "type": "text",
   "text": "The as-milled material was heated with a rate of 5 K min−1 to a final temperature of
400 °C……",
   "page_idx": 2
  },
  {
   "type": "image",
   "img_path": "images/figure_1.jpg",
   "img_caption": [
    "Fig. 1. In situ XRD measurement of LiBH4–MgH2 composites during hydrogen desorption."
   ],
   "img_footnote": [],
   "page_idx": 2
  },
  {
   "type": "text",
   "text": "At approximately 330 °C, hydrogen desorption from MgH2 occurs….",
   "page_idx": 2
  },
  {
   "type": "text",
   "text": "During the isothermal holding period, the formation of MgB2 is observed……",
   "page_idx": 2
   }
   …
]
```


After PDF parsing using MinerU, each paper is converted into a structured JSON file represented as an ordered list of blocks, preserving the original reading order of the document. Each block corresponds to either a text segment or an image (figure) and is described by a dictionary with a mandatory "type" field. Based on this JSON structure, for each image block we extract a fixed-size contextual window consisting of 5 text blocks immediately preceding and 5 text blocks immediately following the image block. These text blocks typically correspond to approximately 5 natural paragraphs before and after the figure in the original PDF. In most cases, this window provides sufficiently complete contextual information to associate the figure with the relevant materials, experimental conditions, and terminology described in the text.

```
def get_image_pct_prompt() -> PromptTemplate:
    image_response_schema = [
        ResponseSchema(
            name="ID",
            description="Sequential integer number (e.g., '1', '2'...). This integer corresponds to the
order of extracted curves from the image."
        ),
        ResponseSchema(
            name="Formula",
            description="Chemical formula of the material, including distinguishing features if the
same formula appears under different preparation conditions (e.g., 'MgH₂ (ball-milled)', 'Ti-Fe alloy
(annealed at 500°C)', 'Ni-MH (as-cast)'). Use standard element symbols and numeric subscripts where
applicable."
        ),
        ResponseSchema(
            name="Description of the hydrogenation process",
            description="Description of hydrogen absorption process (fill the variables in {}).  Output
format: The x- and y-axis units are {[x-axis unit, y-axis unit]}. At {temperature}, absorption starts
at {[x0, y0]}, reaches one or more pressure plateau {p0, p1, ...}. The hydrogen storage density after
the final pressure plateau is {w0}. Finally, the maximum hydrogen storage density is at {[x2, y2]}."
        ),
        ResponseSchema(
            name="Description of the dehydrogenation process",
            description="Description of hydrogen desorption process (fill the variables in {}). Output
format: The x- and y-axis units are {[x-axis unit, y-axis unit]}. At {temperature}, desorption starts
at {[x2, y2]}, reaches one or more pressure plateau {p0, p1, ...}. The hydrogen storage density at the
start of the first pressure plateau during the dehydrogenation process is {w1}. Finally, the minimum
hydrogen storage density is at {[x3, y3]}."
        ),
        ResponseSchema(
            name="Gravimetric hydrogen density",
            description="Maximum hydrogen storage capacity for each curve, along with the corresponding
pressure and temperature. this value is {x2}. Output format: ['Value', 'unit', 'pressure value',
'pressure unit', 'temperature value', 'temperature unit']."
        ),
        ResponseSchema(
            name="Gravimetric hydrogen density at 0.01 MPa",
            description="Hydrogen storage capacity at 0.01 MPa for each curve, with temperature. If
unavailable, record value as 'NA'. Output format: ['Value', 'unit', '0.01', 'MPa', 'temperature value',
'temperature unit']."
        ),
        ResponseSchema(
            name="Gravimetric hydrogen density at 0.1 MPa",
            description="Hydrogen storage capacity at 0.1 MPa for each curve, with temperature. If
unavailable, record value as 'NA'. Output format: ['Value', 'unit', '0.1', 'MPa', 'temperature value',
'temperature unit']."
        ),
        ResponseSchema(
            name="Gravimetric hydrogen density at 1 MPa",
            description="Hydrogen storage capacity at 1 MPa for each curve, with temperature. If
unavailable, record value as 'NA'. Output format: ['Value', 'unit', '1', 'MPa', 'temperature value',
'temperature unit']."
        ),
        ResponseSchema(
            name="Gravimetric hydrogen density at 10 MPa",
            description="Hydrogen storage capacity at 10 MPa for each curve, with temperature. If
unavailable, record value as 'NA'. Output format: ['Value', 'unit', '10', 'MPa', 'temperature value',
'temperature unit']."
        ),
        ResponseSchema(
            name="H2 experiment release",
            description="Based on the description of the dehydrogenation process, this value is {w1} -
{x3}. Output format: ['value', 'unit']."
        ),
        ResponseSchema(
            name="H2 experiment adsorption",
            description="Based on the description of the hydrogenation process, this value is {w0} -
{x0}. Output format: ['value', 'unit']."
        ),
    ]
    image_parser = StructuredOutputParser.from_response_schemas(image_response_schema)
    return PromptTemplate(
        template="""
Please extract information from the following figures step by step using the provided context.

Figure Context:
{caption_context}

For each figure, follow the exact instructions below:
Step 1: Figure Identification and Contextual Summary
Identify which figure the image corresponds to from the Figure Context.
Summarize the relevant figure context in plain text, including the material formula, temperature,
testing condition, and any keywords such as "reversible", "equilibrium pressure", "plateau", etc.
Format: a natural paragraph. Do not output JSON here.

Step 2: Curve Extraction and Grouping
If the figure contains multiple subplots, treat each subplot independently.
In each subplot, identify all absorption and desorption cycle curves. If possible, use legends, colors,
or arrow directions to distinguish absorption from desorption.

For each complete hydrogenation–dehydrogenation cycle (if only hydrogenation or dehydrogenation is
present, record the available data), extract the information as a single dictionary. Group the
corresponding absorption and desorption data together.
For each such cycle, create one dictionary following the specified JSON schema:
{parser_schema}

Important:
Define the equilibrium pressure as the average pressure in the central region of the pressure plateau
observed during hydrogen absorption or desorption in the PCT measurement.
If the y-axis of the PCT plot is shown on a logarithmic scale, first convert it to a linear scale
before identifying the pressure plateau.

The final output should be a JSON-formatted list (array) of these dictionaries, with each dictionary
representing one full absorption-desorption cycle under given conditions.
Output strictly in valid JSON format.


        """,
        input_variables=["caption_context"],
        partial_variables={"parser_schema": image_parser.get_format_instructions()}
    )
```

Figure S14. Example screenshot of the complete prompt template used for image-based PCT data extraction.

9. Failure mode analyses

We include a representative low-scoring case to illustrate typical error patterns and their impact on the final score. **Figure S14 shows the original PCT figure from the paper** 10.1016/j.ijhydene.2011.02.001**.** To facilitate error analysis, we additionally provide a side-by-side comparison table of the **AI-extracted results (left)** and the **ground-truth annotations (right)**, where **clearly incorrect extracted content is highlighted in red**.

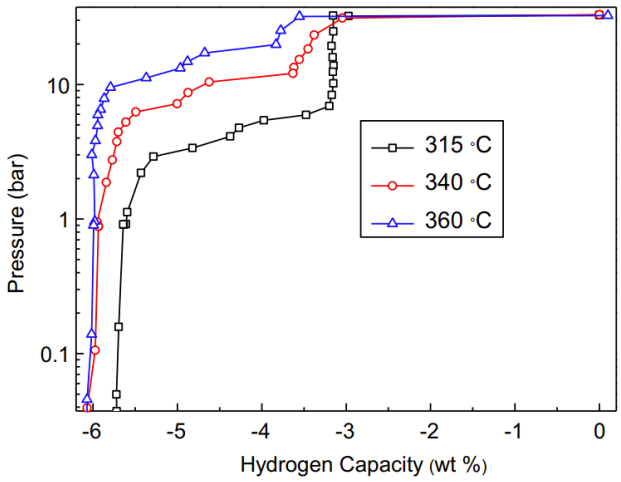

Figure S15 Original PCT figure from 10.1016/j.ijhydene.2011.02.001

Table S4. Comparison table of the AI-extracted results (left) and the ground-truth annotations (right)

AI Extracted Data	Ground Truth Data
[{ "ID": "1", "doi": "10.1016/j.ijhydene.2011.02.001", "Formula": "LiAlH4-MgH2-TiF3", "Hydrogenation temperature": ["300", "°C"], "Hydrogenation pressure": ["2", "MPa"], "Dehydrogenation temperature": ["315", "°C"], "Dehydrogenation pressure": ["0.3", "MPa"], "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"], "Gravimetric hydrogen density": ["6", "wt%", "3", "MPa", "315", "°C"], "H2 experiment release": ["2.48", "wt%"], "H2 experiment adsorption": ["2.68", "wt%"], "enthalpy change (ΔH, kJ/mol H2)": ["78", "kJ/mol H2", "82", "kJ/mol H2"], "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"], "Material type": "Multi-component Hydride",	[{ "ID": "1", "doi": "10.1016/j.ijhydene.2011.02.001", "Formula": "LiAlH4-MgH2-TiF3 composite", "Hydrogenation temperature": "NA", "Hydrogenation pressure": "NA", "Dehydrogenation temperature": ["315", "°C"], "Dehydrogenation pressure": ["3", "bar"], "Volumetric hydrogen capacity": "NA", "Gravimetric hydrogen density": ["5.7", "wt %", "0.05", "bar", "315", "°C"], "H2 experiment release": ["5.5", "wt %"], "H2 experiment adsorption": "NA", "enthalpy change (ΔH, kJ/mol H2)": ["78", "kJ/mol H2", "82", "kJ/mol H2"], "entropy change (ΔS, J/mol H2·K)": ["148", "J/mol K", "150", " J/mol K "], "Material type": "Multi-component Hydride", "Interstitial Hydride Category": "NA",

```
    "Interstitial Hydride Category": "NA",
    "Publication year": "2011"
  },
  {
    "ID": "2",
    "doi": "10.1016/j.ijhydene.2011.02.001",
    "Formula": "LiAlH4-MgH2-TiF3",
    "Hydrogenation temperature": ["300", "°C"],
    "Hydrogenation pressure": ["2", "MPa"],
    "Dehydrogenation temperature": ["340", "°C"],
    "Dehydrogenation pressure": ["0.5", "MPa"],
    "Volumetric hydrogen capacity": ["NA", "NA",
"NA", "NA", "NA", "NA"],
    "Gravimetric hydrogen density": ["6", "wt%", "3",
"MPa", "340", "°C"],
    "H2 experiment release": ["2.48", "wt%"],
    "H2 experiment adsorption": ["2.68", "wt%"],
    "enthalpy change (ΔH, kJ/mol H2)": ["78", "kJ/mol
H2", "82", "kJ/mol H2"],
    "entropy change (ΔS, J/mol H2·K)": ["NA", "NA",
"NA", "NA"],
    "Material type": "Multi-component Hydride",
    "Interstitial Hydride Category": "Others",
    "Publication year": "2011"
  },
  {
    "ID": "3",
    "doi": "10.1016/j.ijhydene.2011.02.001",
    "Formula": "LiAlH4-MgH2-TiF3",
    "Hydrogenation temperature": ["300", "°C"],
    "Hydrogenation pressure": ["2", "MPa"],
    "Dehydrogenation temperature": ["360", "°C"],
    "Dehydrogenation pressure": ["1", "MPa"],
    "Volumetric hydrogen capacity": ["NA", "NA",
"NA", "NA", "NA", "NA"],
    "Gravimetric hydrogen density": ["6", "wt%", "3",
"MPa", "360", "°C"],
    "H2 experiment release": ["2.48", "wt%"],
    "H2 experiment adsorption": ["2.68", "wt%"],
    "enthalpy change (ΔH, kJ/mol H2)": ["78", "kJ/mol
H2", "82", "kJ/mol H2"],
    "entropy change (ΔS, J/mol H2·K)": ["NA", "NA",
"NA", "NA"],
    "Material type": "Multi-component Hydride",
    "Interstitial Hydride Category": "Others",
    "Publication year": "2011"
  }
]
```

```
    "Publication year": "2011"
  },
  {
    "ID": "2",
    "doi": "10.1016/j.ijhydene.2011.02.001",
    "Formula": "LiAlH4-MgH2-TiF3 composite",
    "Hydrogenation temperature": "NA",
    "Hydrogenation pressure": "NA",
    "Dehydrogenation temperature": ["340", "°C"],
    "Dehydrogenation pressure": ["5", "bar"],
    "Volumetric hydrogen capacity": "NA",
    "Gravimetric hydrogen density": ["6.1", "wt %",
"0.08", "bar", "340", "°C"],
    "H2 experiment release": ["5.5", "wt %"],
    "H2 experiment adsorption": "NA",
    "enthalpy change (ΔH, kJ/mol H2)": ["78", "kJ/mol
H2", "82", "kJ/mol H2"],
    "entropy change (ΔS, J/mol H2·K)": ["148", "J/mol
K", "150", " J/mol K "],
    "Material type": "Multi-component Hydride",
    "Interstitial Hydride Category": "NA",
    "Publication year": "2011"
  },
  {
    "ID": "3",
    "doi": "10.1016/j.ijhydene.2011.02.001",
    "Formula": "LiAlH4-MgH2-TiF3 composite",
    "Hydrogenation temperature": "NA",
    "Hydrogenation pressure": "NA",
    "Dehydrogenation temperature": ["360", "°C"],
    "Dehydrogenation pressure": ["10", "bar"],
    "Volumetric hydrogen capacity": "NA",
    "Gravimetric hydrogen density": ["6.1", "wt %",
"0.08", "bar", "360", "°C"],
    "H2 experiment release": ["5.7", "wt %"],
    "H2 experiment adsorption": "NA",
    "enthalpy change (ΔH, kJ/mol H2)": ["78", "kJ/mol
H2", "82", "kJ/mol H2"],
    "entropy change (ΔS, J/mol H2·K)": ["148", "J/mol
K", "150", " J/mol K "],
    "Material type": "Multi-component Hydride",
    "Interstitial Hydride Category": "NA",
    "Publication year": "2011"
  }
]
```


Based on this case study, we summarize four common failure modes:

<u>Failure mode 1:Hallucination / spurious fields</u>

As shown in **Figure R3**, the figure contains only a dehydrogenation (release) process, for which the correct behavior is to record only the corresponding release-related temperature/pressure information. However, the model incorrectly hallucinated additional information (e.g., temperature/fields associated

with a non-existent hydrogenation process), which leads to direct penalties in the item-level accuracy scoring.

Failure mode 2: Numerical inaccuracy due to visual reading errors

According to the ground truth, the $LiAlH_4$–MgH_2–TiF_3 composite reaches maximum release capacities of 5.7, 6.1, and 6.1 wt.% at 315°C, 340°C, and 360°C, respectively. The model, however, extracted these values as approximately 6 wt.% across conditions, likely due to multimodal visual estimation errors when reading plotted curves/axis ticks.

Failure mode 3: Misinterpretation of multi-plateau PCT behavior

While many PCT curves exhibit a single plateau, Figure R3 contains two plateaus. The model extracted the release amount near the end of the first plateau (≈ 2.48 wt.%), whereas the ground truth corresponds to the release amount after the second plateau (≈ 5.5–5.7 wt.%). This indicates that the model can struggle to correctly interpret and select the appropriate plateau endpoint in multi-step PCT behavior.

Failure mode 4: Missing data when key quantities appear only in derived plots.

When thermodynamic parameters (e.g., enthalpy/entropy from a van't Hoff plot) are only provided in secondary plots or figure-derived analyses and not stated explicitly in the main text near the figure, the current workflow may fail to recover them, resulting in missing-field penalties.

10. Diversity of test set

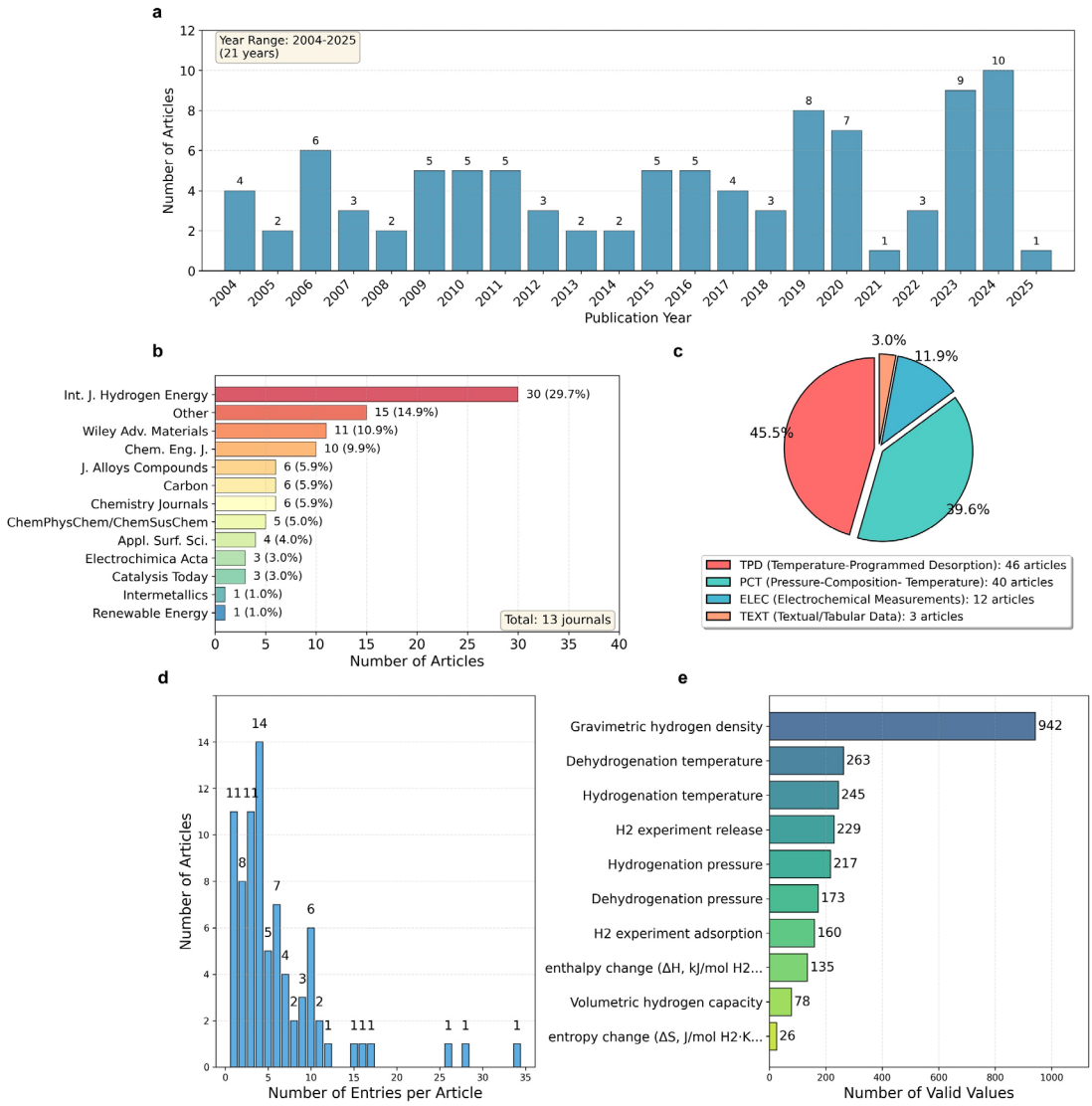

Figure S16. Diversity summary of the 100 evaluation articles, including publication years, journal distribution, figure/technique categories, entry complexity, and attribute coverage.

- **Publication years:** The 100 evaluation papers span 2004–2025 (21 years) and cover 13 journals**.** The largest single source is *International Journal of Hydrogen Energy* **(**30/100, 29.7%**),** with the remaining papers distributed across multiple venues (e.g., *Advanced Materials* 11, *Chemical*

Engineering Journal 10, *Journal of Alloys and Compounds* 6, *Carbon* 6, and several others), plus an "Other" category (15/100, 14.9%) representing additional journals.

- **Figure types and experimental techniques.** The test set includes multiple experimental modalities commonly used in hydrogen-storage studies. In particular, the evaluated papers contain TPD **(**Temperature-Programmed Desorption) figures **(**46 articles), PCT (Pressure–Composition–Temperature) figures (40 articles), electrochemical measurements (12 articles**)**, and textual/tabular-only data (3 articles). These categories reflect the diversity of experimental techniques and reporting styles encountered in the literature.
- **Data complexity.** The number of extracted entries per article varies substantially, with most papers containing only a few entries per paper (typical range ≈ 1–10), while a smaller fraction contains much larger numbers of entries**,** forming a long tail up to approximately 35 entries per article**.** In addition, the coverage of different quantitative attributes is uneven, reflecting real-world reporting diversity (*e.g.*, gravimetric capacity is most frequently available, whereas entropy is reported far less often).

10. Robustness of the DIVE workflow to multi-curve figures, overlapping curves, and low-resolution images

To provide concrete evidence of DIVE's robustness to multi-curve figures, overlapping curves, and low-resolution images, we include a representative case study here.

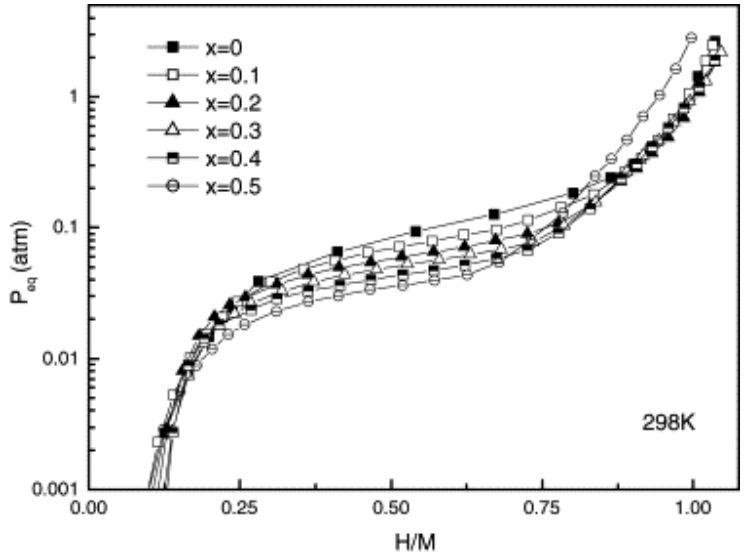

Figure S17. The desorption P–C isotherms for $La_2Mg(Ni_{1-x}Co_x)_9$ (x = 0–0.5) alloys at 25 °C (DOI: 10.1016/j.electacta.2004.08.004)

As shown in **Figure S17** (DOI: 10.1016/j.electacta.2004.08.004;), the figure contains six closely spaced PCT curves with substantial overlap, presented primarily in black-and-white. Curve identity is largely indicated by subtle scatter-marker differences, and the overall image resolution is relatively low. This figure therefore represents a challenging scenario for visual digitization.

Table S5. AI extracted data for multi-curve, overlapping curves, and low-resolution images

AI extracted data (Gemini 2.5 pro + DeepSeek-R1-0528-Qwen3-8B)
[{ "ID": "1", "doi": "10.1016/j.electacta.2004.08.004", "Formula": "La_2MgNi_9", "Hydrogenation temperature": ["25", "°C"], "Hydrogenation pressure": ["1", "atm"], "Dehydrogenation temperature": ["25", "°C"], "Dehydrogenation pressure": ["0.093", "atm"], "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"], "Gravimetric hydrogen density": ["1.03", "wt%", "0.093", "atm", "25", "°C"],

```
  "H2 experiment release": ["0.65", "H/M"],
  "H2 experiment adsorption": "NA",
  "enthalpy change (ΔH, kJ/mol H2)": ["NA", "NA", "NA", "NA"],
  "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"],
  "Material type": "Interstitial Hydride",
  "Interstitial Hydride Category": "AB3",
  "Publication year": "2004"
 },
 {
  "ID": "2",
  "doi": "10.1016/j.electacta.2004.08.004",
  "Formula": "La2Mg(Ni0.9Co0.1)9",
  "Hydrogenation temperature": ["25", "°C"],
  "Hydrogenation pressure": ["1", "atm"],
  "Dehydrogenation temperature": ["25", "°C"],
  "Dehydrogenation pressure": ["0.075", "atm"],
  "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"],
  "Gravimetric hydrogen density": ["1.04", "wt%", "0.075", "atm", "25", "°C"],
  "H2 experiment release": ["0.6", "H/M"],
  "H2 experiment adsorption": "NA",
  "enthalpy change (ΔH, kJ/mol H2)": ["NA", "NA", "NA", "NA"],
  "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"],
  "Material type": "Interstitial Hydride",
  "Interstitial Hydride Category": "AB3",
  "Publication year": "2004"
 },
 {
  "ID": "3",
  "doi": "10.1016/j.electacta.2004.08.004",
  "Formula": "La2Mg(Ni0.8Co0.2)9",
  "Hydrogenation temperature": ["25", "°C"],
  "Hydrogenation pressure": ["1", "atm"],
  "Dehydrogenation temperature": ["25", "°C"],
  "Dehydrogenation pressure": ["0.06", "atm"],
  "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"],
  "Gravimetric hydrogen density": ["1.05", "wt%", "0.06", "atm", "25", "°C"],
  "H2 experiment release": ["0.55", "H/M"],
  "H2 experiment adsorption": "NA",
  "enthalpy change (ΔH, kJ/mol H2)": ["NA", "NA", "NA", "NA"],
  "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"],
  "Material type": "Interstitial Hydride",
  "Interstitial Hydride Category": "AB3",
  "Publication year": "2004"
 },
 {
  "ID": "4",
  "doi": "10.1016/j.electacta.2004.08.004",
  "Formula": "La2Mg(Ni0.7Co0.3)9",
  "Hydrogenation temperature": ["25", "°C"],
  "Hydrogenation pressure": ["1", "atm"],
  "Dehydrogenation temperature": ["25", "°C"],
  "Dehydrogenation pressure": ["0.05", "atm"],
  "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"],
  "Gravimetric hydrogen density": ["1.06", "wt%", "0.05", "atm", "25", "°C"],
  "H2 experiment release": ["0.5", "H/M"],
  "H2 experiment adsorption": "NA",
  "enthalpy change (ΔH, kJ/mol H2)": ["NA", "NA", "NA", "NA"],
  "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"],
  "Material type": "Interstitial Hydride",
  "Interstitial Hydride Category": "AB3",
  "Publication year": "2004"
 },
```

```
  {
   "ID": "5",
   "doi": "10.1016/j.electacta.2004.08.004",
   "Formula": "La2Mg(Ni0.6Co0.4)9",
   "Hydrogenation temperature": ["25", "°C"],
   "Hydrogenation pressure": ["1", "atm"],
   "Dehydrogenation temperature": ["25", "°C"],
   "Dehydrogenation pressure": ["0.04", "atm"],
   "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"],
   "Gravimetric hydrogen density": ["1.0", "wt%", "0.04", "atm", "25", "°C"],
   "H2 experiment release": ["0.45", "H/M"],
   "H2 experiment adsorption": "NA",
   "enthalpy change (ΔH, kJ/mol H2)": ["NA", "NA", "NA", "NA"],
   "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"],
   "Material type": "Interstitial Hydride",
   "Interstitial Hydride Category": "AB3",
   "Publication year": "2004"
  },
  {
   "ID": "6",
   "doi": "10.1016/j.electacta.2004.08.004",
   "Formula": "La2Mg(Ni0.5Co0.5)9",
   "Hydrogenation temperature": ["25", "°C"],
   "Hydrogenation pressure": ["1", "atm"],
   "Dehydrogenation temperature": ["25", "°C"],
   "Dehydrogenation pressure": ["0.036", "atm"],
   "Volumetric hydrogen capacity": ["NA", "NA", "NA", "NA", "NA", "NA"],
   "Gravimetric hydrogen density": ["0.989", "wt%", "0.036", "atm", "25", "°C"],
   "H2 experiment release": ["0.4", "H/M"],
   "H2 experiment adsorption": "NA",
   "enthalpy change (ΔH, kJ/mol H2)": ["NA", "NA", "NA", "NA"],
   "entropy change (ΔS, J/mol H2·K)": ["NA", "NA", "NA", "NA"],
   "Material type": "Interstitial Hydride",
   "Interstitial Hydride Category": "AB3",
   "Publication year": "2004"
  }
]
```


For **Figure S17**, the AI-extracted structured output (**Table S6**) demonstrates that DIVE can still recover fine-grained quantitative trends. In particular, the extracted dehydrogenation equilibrium pressures capture a consistent monotonic shift across compositions, resolving values of 0.093, 0.075, 0.060, 0.050, 0.040, and 0.036 atm. Despite the strong overlap near curve endpoints, the extracted gravimetric hydrogen density also preserves the expected trend, yielding 1.03, 1.04, 1.05, 1.06, 1.00, and 0.989 wt.% across the six curves. These results indicate that DIVE remains reliable even when curves are highly similar, partially overlapping, and visually difficult to separate. Overall, this example suggests that DIVE can maintain high recognition fidelity under non-ideal image conditions commonly encountered in the literature.

11. Accuracy breakdown for Step 1 (caption-based figure identification)

Step 1 task definition. For Step 1, we classify each figure caption into four categories:

- **0:** None (not relevant)
- **1:** PCT (pressure–composition–temperature isotherms)
- **2:** ELEC (electrochemical discharge)
- **3:** TPD (temperature-programmed desorption / related curves)

Evaluation protocol. We collected **>800 captions** from the **100-paper evaluation set** and compared the LLM's predicted labels with human-verified ground truth. For computing PR/F1, we treat **{1,2,3} as "relevant"** and **0 as "not relevant."** The results are:

- **Precision: 88.89%** (of captions predicted as relevant, 88.89% are truly relevant)
- **Recall: 89.31%** (of truly relevant captions, 89.31% are retrieved)
- **F1: 89.10%**
- **Accuracy: 88.67%**

Table S6. Full per-caption results

doi	Caption	AI	Ground truth	result
10.1016/j.jallcom.2004.12.158	Fig. 1. Observed (points) and calculated (line) neutron diffraction patterns for an ErNi_{3} sample deuterided at 100 bar (a), calculated patterns of the contributing phases $\upbeta_{2}$ - (b) and $\upgamma$ -deuteride (c). Vertical bars indicate positions of Bragg peaks; difference pattern represented at the bottom; $\lambda=1.493814(19)\mathrm{\normalfont\AA}$.	3	0	FP
10.1016/j.ijhydene.2023.03.329	Fig. 3 e PeC-T curves of the annealed $\mathbf{LaY}_{1\cdot9}\mathbf{Ni}_{10}\mathbf{Mn}_{0\cdot5}\mathbf{Al}_{0\cdot2}$ alloys (a) and the XRD patterns of Alloy-1050 at different hydrogen absorption states (b).	1	1	TP
……				
Full per-caption results (**811 rows**) can be in caption_evaluation_results.csv in our repository (https://github.com/gtex-hydrogen-storage/DIVE).				