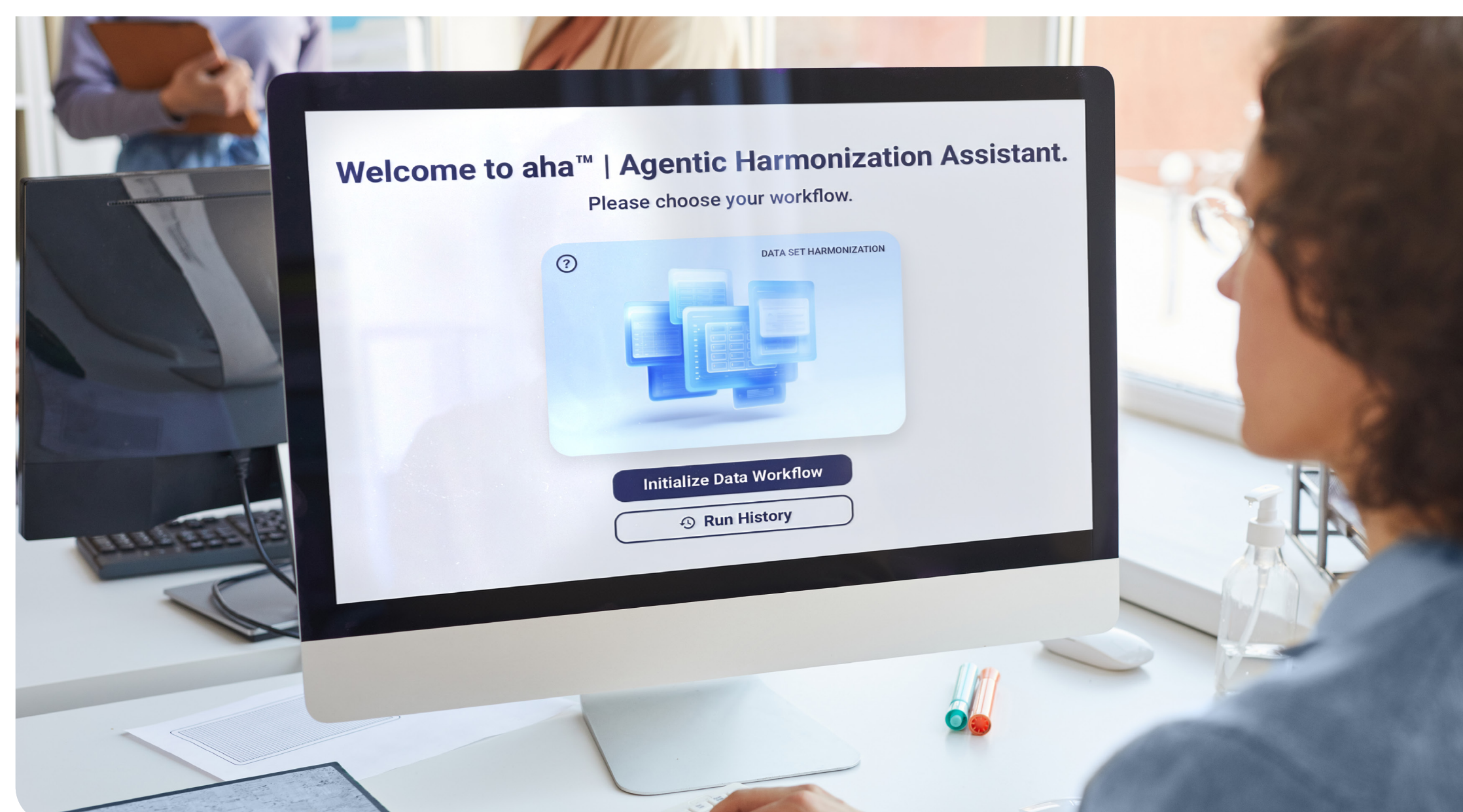




BACKGROUND

The Alzheimer's Disease (AD) research ecosystem contains a rapidly growing collection of clinical, imaging, sociodemographic, and molecular datasets. Yet inconsistencies in schemas, variable encodings, and batch effects require extensive preparation for cross-cohort analyses and constrain their utility for building large-scale trustworthy AI models¹. To address this bottleneck, we introduce the **Agentic Harmonization Assistant (aha™)**, a multi-agent system that reasons, plans, executes, and validates end-to-end dataset harmonization.



RESULTS

aha™ reduces harmonization time by 93%, decreasing the required effort from 28 hours of manual work to 2.5 hours for 100 variables. We evaluated **aha™** performance per module.



PROFILING:

We assessed the accuracy of the data extraction performed by the profiling.

We evaluate the correctness of the data extraction on three different domains:

- Data Type: Accuracy, Precision, Recall, and F1-Score measuring the correct selection of the data type between the possible options (float, int, binary, categorical_nominal, categorical_ordinal, text, or date).
- Variable description: Cosine Similarity between the generated description and the ground-truth on the data dictionary.
- Ranges or categories: IoU between the values extracted by the system and the ground-truth on the data dictionary

STRATEGY & PLANNING:

We evaluated the correctness of the selected harmonization plan on 4 items.

- Variable Alignment: F1 score measuring whether predicted variables were correctly matched to ground-truth variables.
- Decision Match: F1 score for selecting the correct harmonization decision (matched, mapped, partially mapped, calculated, or not mapped).
- Transformation Selection: F1 score evaluating whether the correct set of transformation operations (e.g., map, rename, case_when) was selected for each variable.
- Transformation Correctness: Mean accuracy of each operation's parameters (e.g., correct value mappings, rename targets, conditions), scored in [0, 1] per operation across aligned pairs.

EXECUTION:

We examined the proper execution of each transformation plan using RMSE for numerical variables and accuracy for categorical variables.

VALIDATION:

We analyzed the reliability of the reasoning used to approve or reject a harmonized variable. The module evaluates the correctness of the harmonization over five items:

- Concept alignment
- Data type compatibility
- Value/category coverage
- Missing and placeholders
- Stats sanity check.

Additionally, we created a subset of curated correct and incorrect harmonizations, along with manually generated rationales explaining why each variable in this subset was harmonized correctly or not. With this subset and the Prometheus 2 evaluator², we measured the performance of our model in producing accurate reasoning on the five-item rubric and alignment between the rubric and the final decision for each variable (harmonized/not-harmonized). The final score used a penalized factor for lower scores.



METHODS

aha™ is powered by multiple large language models operating through in-context learning. It is a four-stage system (**Figure 1**) that begins with a profiling module, which characterizes both the standard and unharmonized datasets using their data dictionaries to generate machine-readable variable profiles that combine descriptive statistics with semantic information. These profiles are then used by the strategy and planning module to create a harmonization plan for each variable, including confidence scores and required transformations, which are refined through an iterative process. The execution module implements the plan. Finally, the validation module evaluates the harmonized output against the standard dataset and generates a consistency analysis.

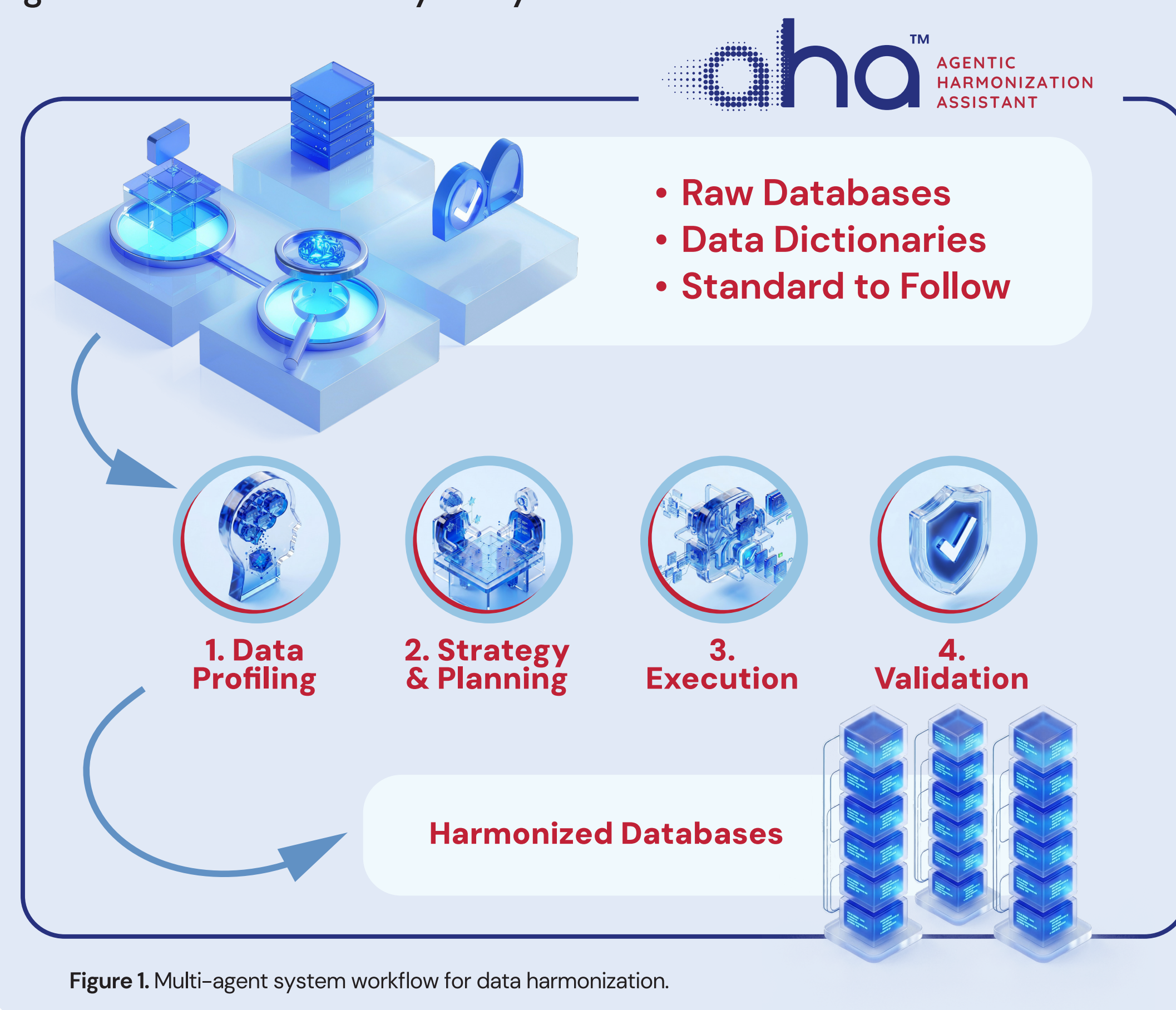


Figure 1. Multi-agent system workflow for data harmonization.

CONCLUSIONS

Our interim findings confirm that **aha™** is a powerful and scalable solution for harmonizing AD data across global cohorts. We continue to enhance model performance, enabling accurate outputs that support research and clinical applications. **aha™** will continue to improve through ongoing fine-tuning and integration of additional modalities such as neuroimaging and omics harmonization, positioning it as a foundational platform for AI-driven AD research and precision prevention strategies.

DISCLOSURE OF FUNDING

This research was funded by IGC Pharma, LLC.

REFERENCES

1. Hao, X., Li, X., Zhang, GQ. et al. An ontology-based approach for harmonization and cross-cohort query of Alzheimer's disease data resources. BMC Med Inform Decis Mak 23 (Suppl 1), 151 (2023). <https://doi.org/10.1186/s12911-023-02250-z>
2. S. Kim et al. Prometheus 2: An Open Source Language Model Specialized in Evaluating Other Language Models. Accessed: Jun. 22, 2026. [Online]. Available: <https://arxiv.org/abs/2405.01535>



SCAN ME