Agents of Intelligence: Lowering the Barrier to AI for Domain Scientists

Ravi Shende, Tainã Coleman, Hena Ahmed, Ishayu Ghosh, Ismael Perez, Tara Javidi and İlkay Altıntaş

A collaboration between University of California, San Diego and Loyola Marymount University

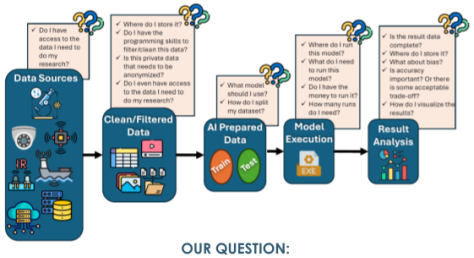
Abstract

Artificial intelligence is transforming how science is done, yet most domain scientists face a steep barrier to adopting it: a fragmented landscape of models, platforms, and data pipelines that rewards tacit expertise accumulated over years. We introduce **Agents of Intelligence (AoI)**, a modular, composable, human-in-the-loop framework of AI agents that sense workflow context, automate routine decisions, recommend models and resources, and learn from usage. Grounded in interviews with Schmidt AI postdocs across disciplines, AoI integrates with the National Data Platform to support the full research lifecycle from data acquisition to model deployment, making AI-assisted science accessible across domains.

The Expertise Gap

We interviewed 11 Schmidt AI postdoctoral researchers at UC San Diego spanning medical imaging, neuroscience, molecular modeling, ecology, genomics, earth systems, quantum chemistry, and more.
The same barriers appeared in every domain:
B1. No formal AI/ML training: scientists are domain experts, not ML engineers
B2. Compute resource confusion: which cluster? which GPU? how much memory?
B3. Data prep & metadata friction: cleaning, formatting, documenting data eats weeks
B4. Workflow fragmentation: tools don't talk to each other

Domain Scientist's View



OUR QUESTION:
Can AI agents encode the tacit computational expertise domain scientists lack, and put it at their fingertips?

The Framework

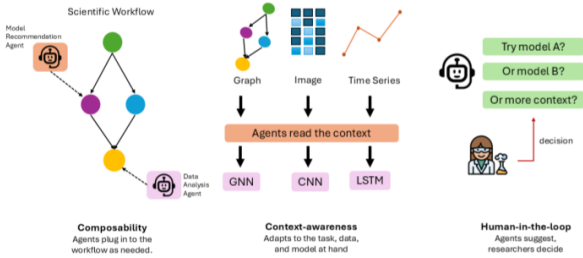


A framework of AI agents as context-aware research companions

4 capabilities

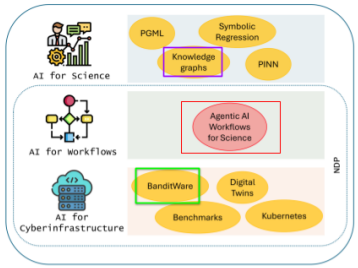
- **Sense:** track workflow state and user intent
- **Automate**: handle routine steps, reduce cognitive load
- **Recommend**: suggest models, resources, visualizations
- **Learn:** improve over time from feedback across domains

3 design principles

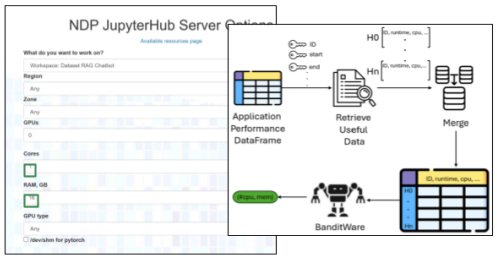


What AoI Has Unlocked

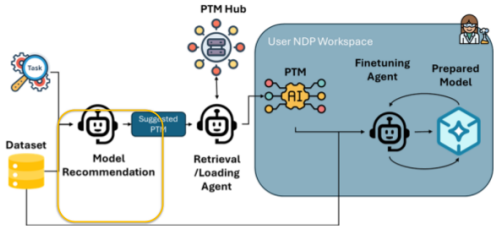
AoI is designed to accept the addition of actors in the various different layers. From the domain science to the application and cyberinfrastructure.

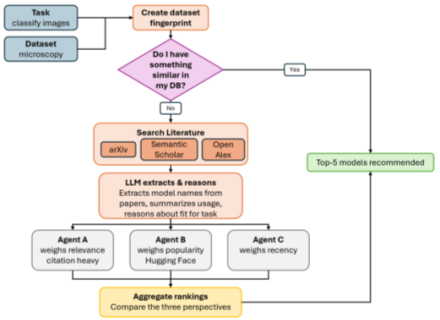


BanditWare: the right hardware without the guesswork (B2)
Scientists submit jobs without knowing which hardware fits. BanditWare uses a contextual bandit to learn workflow hardware fit without large historical datasets. **Result:** better runtimes, less wasted compute, no manual tuning.



Model Recommendation & Fine-Tuning (B1)
Given a dataset and task, a multi-agent system (1) ranks the top-5 pre-trained models, (2) retrieves the best fit from a model hub, and (3) fine-tunes it in the user's workspace. What used to take weeks of searching, installing, and debugging now takes minutes.

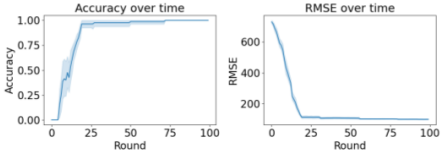




Evaluation: BanditWare

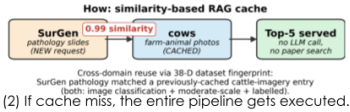
BanditWare (Use Case: Matrix Multiplication)
BanditWare's recommendations are most useful for hardware-sensitive applications, such as highly parallelizable programs. A common parallel bottleneck operation in scientific workflows is matrix multiplication.
Matrix Squaring dataset: Dataset of matrix squaring calculation runtimes across 5 different hardwares. Matrix sizes range from 5000 to 12,500. Runtimes range from 1 to 30 minutes. Accuracy is the percent of test runs where BanditWare selects the hardware with the lowest actual runtime. RMSE is the runtime prediction error, calculated per hardware then averaged across all 5 platforms.

BanditWare Performance Over Time



Evaluation: Model Recommendation

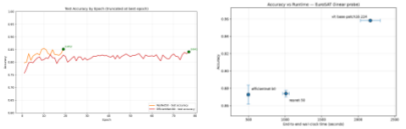
(1) The pipeline runs a cosine similarity to previous results, this way if similarity ≥ 85%, the results are served with no LLM call necessary



Note: cache hits do not suffer from the curse of dimensionality because most dimensions are one-hot encodings

(2) If cache miss, the entire pipeline gets executed.

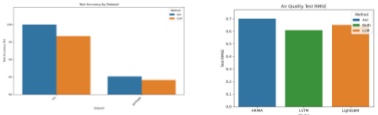
(3) If chosen by the user, the framework, as a last step, will call a fine tune agent that retrieves the smallest model from the top 5 suggestion from Hugging Face and prepared the last layer of the model to match the data. The idea is to offer a ready model as fast as possible.



To evaluate the Recommendation Agent, we tested the agent vs its internal LLM (Qwen3) directly on 3 datasets. The top recommended model from each was then trained and tested on the dataset using cross validation to select hyperparameters when applicable on an 80/20 train/test split. The datasets were:

1. **IRIS tabular classification**
2. **UCI Air Quality Time Series regression (1-step forecasts of CO)**
3. **UCI Garbage type Image classification**

In the above plot, the model proposed for the Garbage Dataset by AoI was both more accurate and quicker to learn; both use early stopping when the validation (not test) accuracy decreases.
Below, we compare the accuracy/RMSE of AOI vs its base LLM suggestions on the same datasets: Iris (tabular classification), Garbage Types - UCI (Image classification), Air Quality - UCI (Time Series Regression). In the graphs below, the accuracy on both classification datasets was better on the AOI recommended models (TabPFN vs SVM and ResNet vs EfficientNet). The RMSE is slightly better (lower) on the first suggestion for the LLM compared to AoI, but both recommenders' second choice were the same and performed best.



Coming Up…

- **Knowledge Graph Builder:** Ingests data from heterogeneous datasets, finds matching ontologies and builds the knowledge graph. Application: Digital twin representation.
- **Sandbox for agent creation:** SKILL.md based files for agent creation
- **Data Harmonization Agent:** agent that prepares dataset to be added to the federated NDP data catalog.

Acknowledgements

This work was supported by NSF award 2134904 for Burn-Pro3D and NSF award 2333609 for the National Data Platform and Schmidt Sciences. The Nautilus Kubernetes Cluster of the National Research Platform, funded by NSF 2112167 and 2120019, was also used for this research. The authors thank their team at the San Diego Supercomputer Center for their collaboration.

Schmidt Sciences
UC San Diego
SCIL Societal Computing and Innovation Lab