



AutoAI Cluster Presentation

Research title: Agentic Tools for Scientific Research Automation

Presenter: K.Markov Members: I.Paik, Q.Zhao, Y.Liu, Y.Watanobe, J.Villegas, M.Mozgovoy, L.Jing, T.Truong

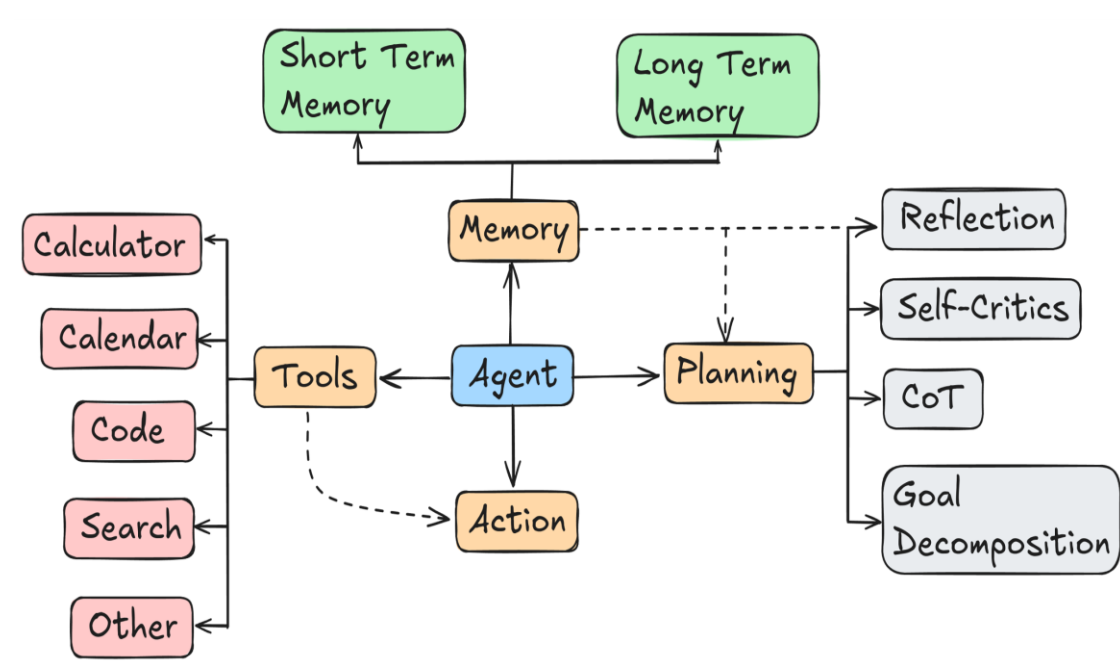


Summary

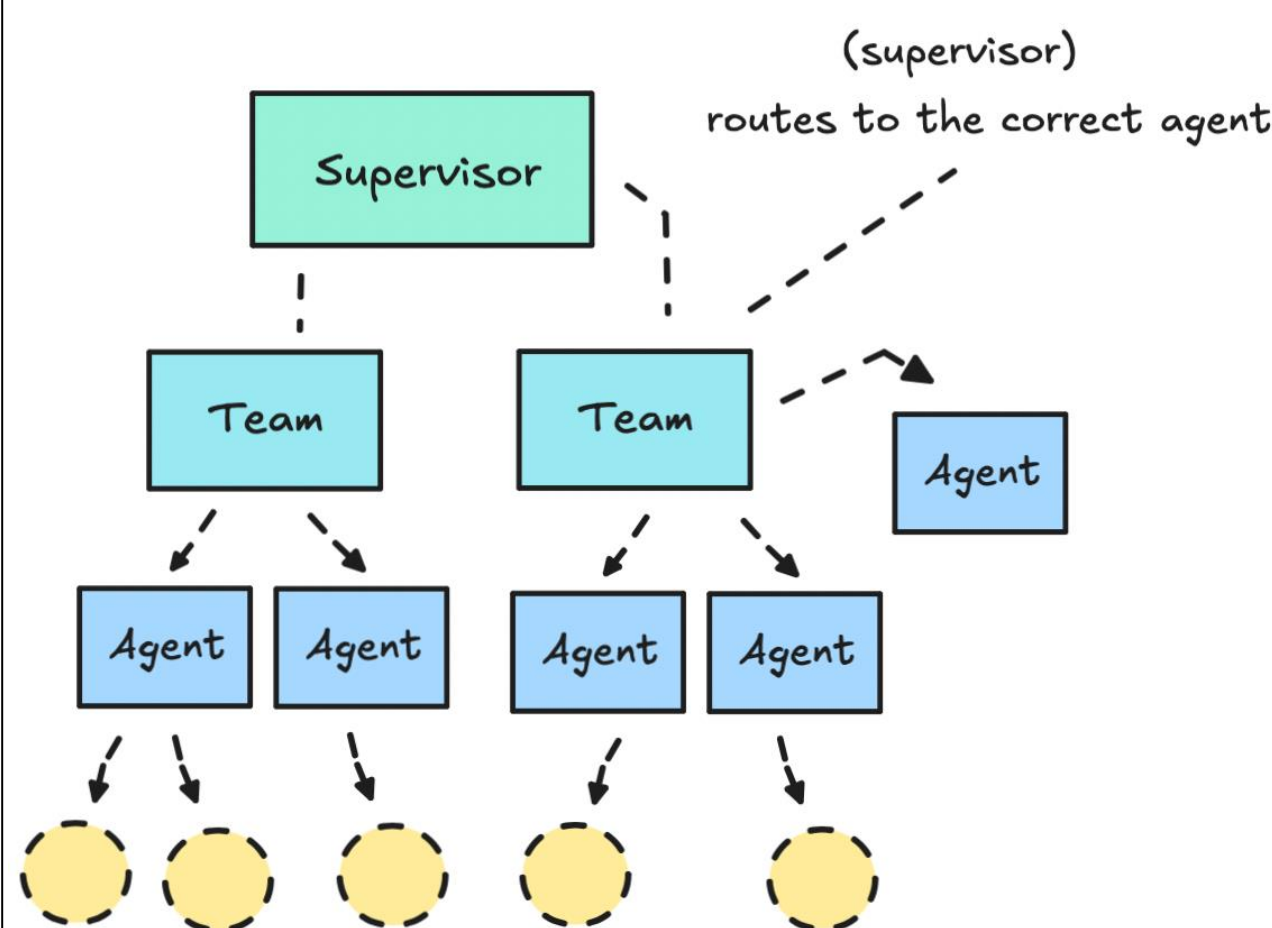
Artificial intelligence is evolving toward autonomy in science and data ecosystems. **Scientific Agents** progress from *tools* to *scientists*, automating hypothesis generation, experimentation, and interpretation. **Data Agents** follow a multi-level hierarchy advancing from human-assisted data management to self-orchestrating and generative systems. Both paradigms aim for autonomous reasoning, transparency, and responsible governance. Together, they represent converging paths toward unified AI ecosystems where Scientific Agents formulate hypotheses and Data Agents manage data, forming a closed, intelligent loop for scientific discovery and knowledge creation.

Background

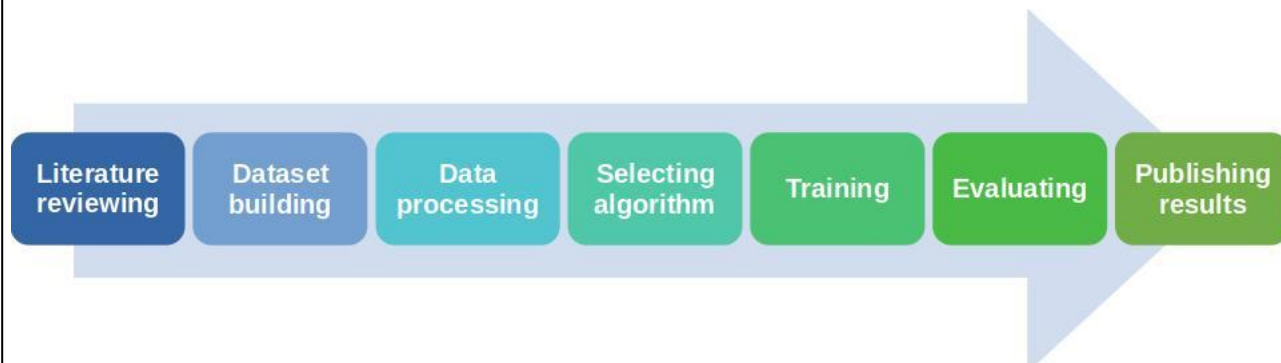
What is an **AI Agent**? A system that autonomously performs tasks by designing workflows with available tools.



Agentic system: collaborative network of multiple autonomous AI agents that work together to achieve common goal.



ML/DL research pipeline automation:



Some currently available AI tools:

- For literature search & review



- For code generation & execution



- For summarization & writing



Specialized Agents

Scientific Agents

LLM system automating scientific tasks — ranging from hypothesis generation and experiment design to simulation and report writing [1].

•Level1: AI as Tool

(Simple & static tasks)

Literature search, code generation, etc.

Direct human supervision

•Level2: AI as Analyst

(Goal-oriented reasoning)

Statistical analysis, model discovery, etc.

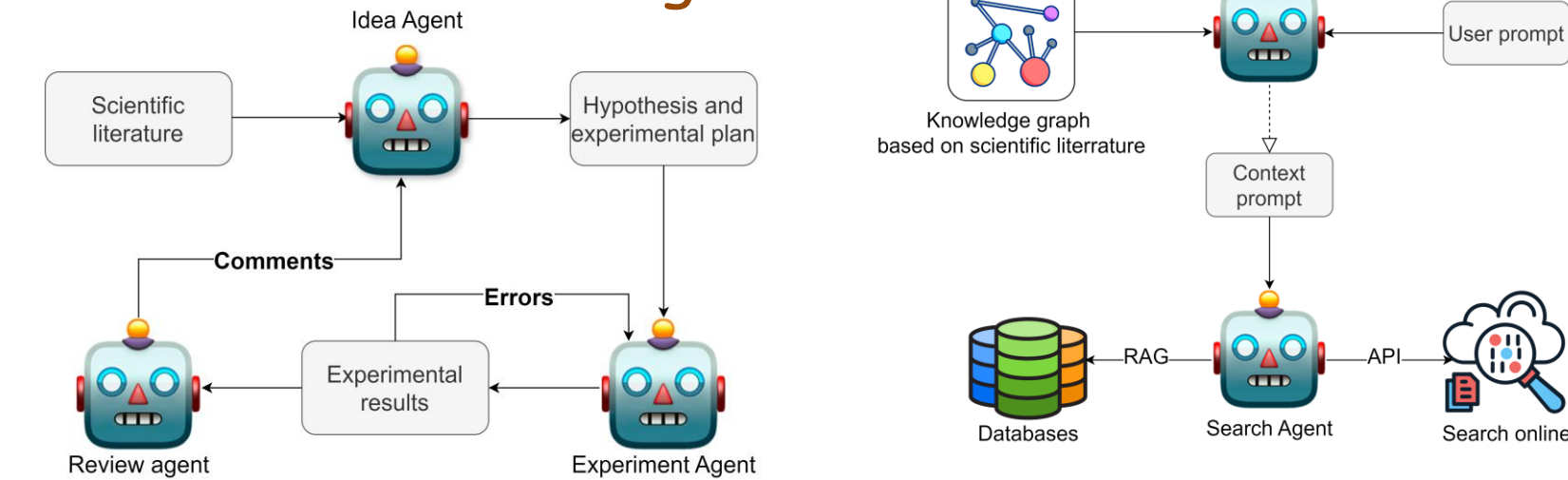
Reduced human intrusion

•Level3: AI as Scientist

(multi-stage navigation)

Hypothesis generation, end-to-end research cycles.

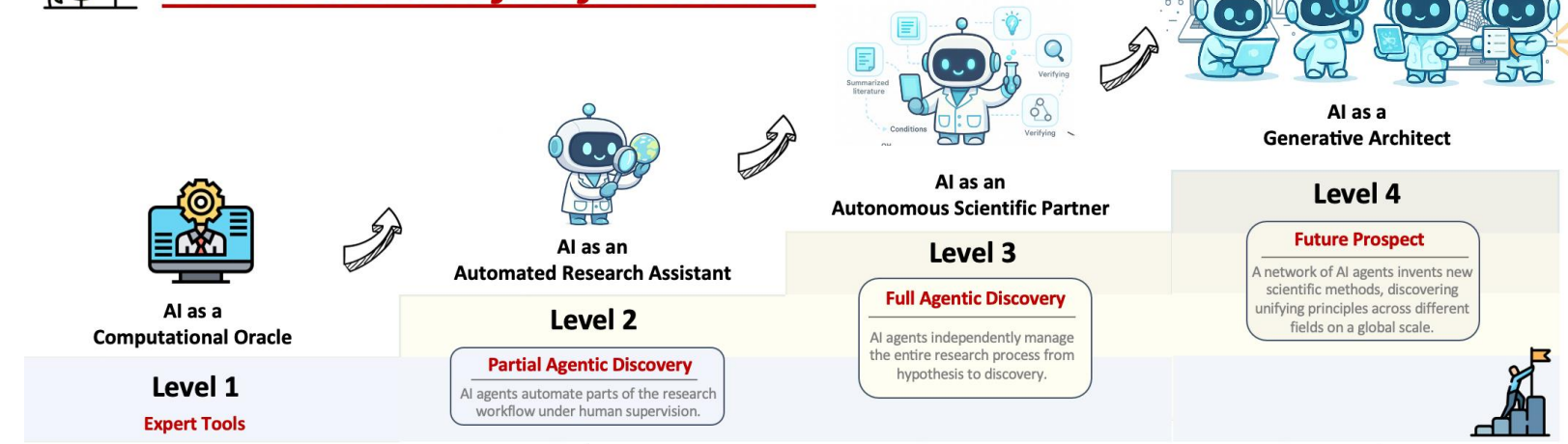
Minimal human oversight



Scientific agent with iterative hypothesis refinement

Scientific agent with external knowledge

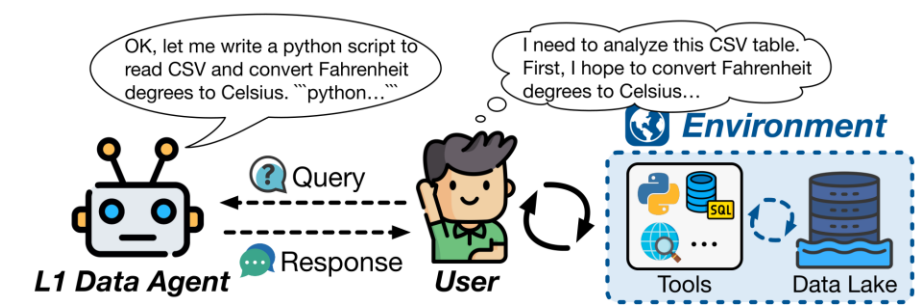
The Evolution of AI for Science



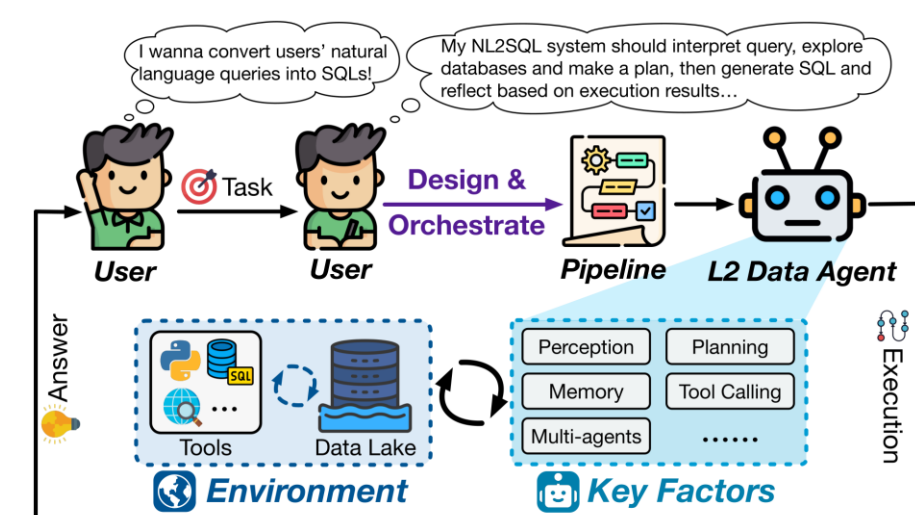
From computational tools to creative collaborators: the four-level AI evolution in science [3].

Data Agents

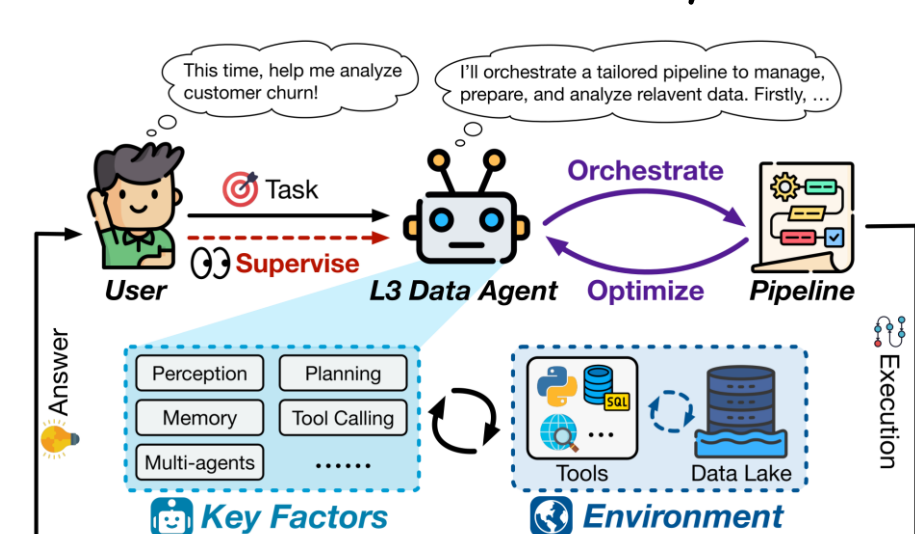
LLM system performing data management, preparation, and analysis with minimal human intervention [2].



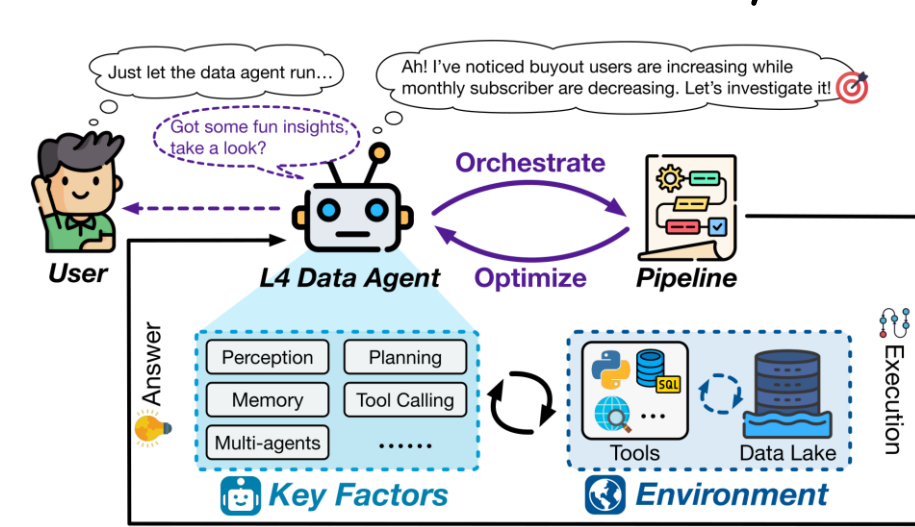
Assistance



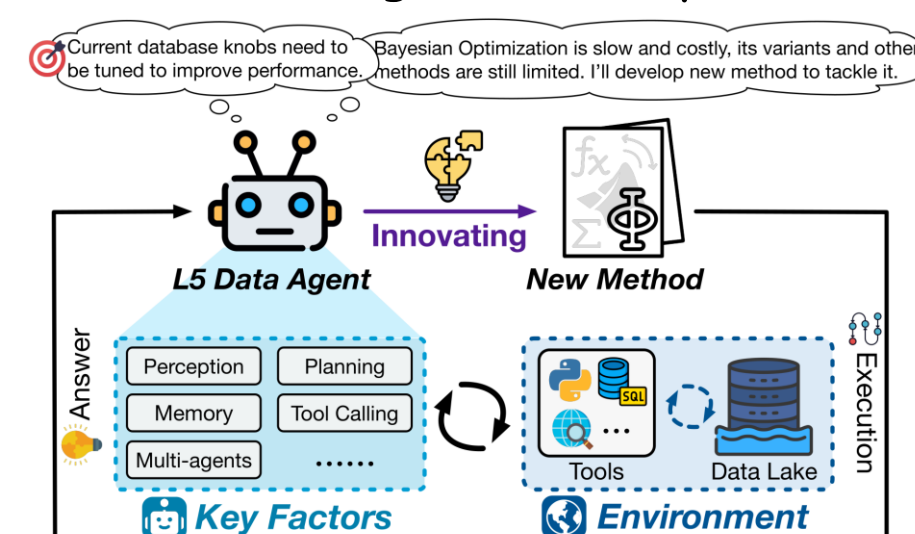
Partial Autonomy



Conditional Autonomy



High Autonomy



Full Autonomy

Comparative Insights

Scientific and Data Agents converge toward autonomous, reasoning-driven research ecosystems.

Comparison:

- Scientific Agents focus on hypothesis-driven discovery.
- Data Agents focus on end-to-end data orchestration.
- Shared goals: transparency, governance, and continuous learning.

Dimension	Scientific Agents	Data Agents
Focus	Scientific reasoning & discovery	Data lifecycle orchestration
Autonomy Scale	3 levels (Tool-Analyst-Scientist)	5 levels (L1-L5)
Core Process	Scientific method stages	Data management pipeline
Current Frontier	AI Scientist (self-refining)	Proto-L3 orchestration
Challenges	Ethics, transparency, validation	Accountability, taxonomy clarity

Feature Directions

Future Vision: Autonomous AI Research Ecosystems where Scientific and Data Agents collaborate:

- Scientific agents generate hypotheses.
- Data agents gather, clean, and analyze data.
- Both form a closed loop of scientific reasoning.

Common Challenges:

- Transparency & interpretability
- Continuous self-improvement
- Integration of reasoning + perception + action

References

- Ren, Shuo, et al. "Towards scientific intelligence: A survey of llm-based scientific agents." arXiv preprint arXiv:2503.24047 (2025).
- Zhu, Yizhang, et al. "A Survey of Data Agents: Emerging Paradigm or Overstated Hype?." arXiv preprint arXiv:2510.23587 (2025).
- Wei, Jiaqi, et al. "From ai for science to agentic science: A survey on autonomous scientific discovery." arXiv preprint arXiv:2508.14111 (2025).