Response to Discussion 2: Analysis-Planning and Result Agents with Reasoning

Team LAMBDA
The Hong Kong Polytechnic University
Hong Kong SAR, China

We are deeply grateful to Professor Xihong Lin for her thoughtful and constructive discussion of our work on LAMBDA [Sun et al., 2025a]. Her expertise in biostatistics and statistical genetics provides valuable perspective on how LAMBDA can be enhanced to better support complex, domain-specific scientific workflows. Professor Lin raises two particularly important suggestions: the addition of an analysis-planning agent and a result agent, both with enhanced reasoning capabilities. We address these suggestions in detail below.

As discussed in our comprehensive survey of large language model-based agents for statistics and data science [Sun et al., 2025b], the integration of reasoning capabilities into AI agents represents a critical frontier for advancing the field.

1 The Analysis-Planning Agent with Reasoning

Professor Lin correctly observes that LAMBDA's current two-agent setup (Programmer and Inspector) assumes users already know which analyses to run, what statistical methods to use, and what domain-specific community standards to apply. This expectation can be challenging for non-expert researchers. Extending LAMBDA with an analysis-planning agent could address this limitation by interacting with users to construct analysis plans that answer scientific questions of interest, adhere to field norms, and incorporate appropriate analytic reasoning.

We wholeheartedly agree with this vision. Many disciplines rely on well-established analytic pipelines and conventions developed through years of methodological and empirical research. For example, Professor Lin mentions Genome-Wide Association Studies (GWAS), which commonly require genotype quality control (Hardy-Weinberg equilibrium tests, missing-call filters), population-structure correction using ancestry principal components, association testing with mixed models to handle relatedness, and multiple-testing correction. These steps represent community consensus critical for reproducible science.

An analysis-planning agent would collaborate with users to:

1. Identify the scientific domain of the data and refine the user's research goals and questions.

2. Map the user's research questions to appropriate analytic procedures by applying statistical and domain-specific reasoning rules and tailoring them to the data at hand.

3. Propose a comprehensive analysis pipeline that adheres to community standards and lays the groundwork for a comprehensive analysis report.

Such a pipeline would include quality control procedures, data cleaning and preprocessing, exploratory analysis, modeling choices, result validation strategies, plans for generating intermediate results to solicit user feedback, and components of the final analysis report that addresses the user's scientific questions while following field conventions.

In our ongoing research, we are exploring approaches to incorporate reasoning modules into a fully end-to-end operational mode. Recent advanced techniques such as Chain-of-Thought (CoT), Tree of Thoughts (ToT), and Graph of Thoughts (GoT) can support complex tasks by enabling structured thinking, reflection, and exploration of multiple solution paths—particularly well-suited to open-ended data science workflows. Following Professor Lin's suggestion, we plan to extend the current two-agent system to a multi-agent framework that includes a planning agent and an analyst agent.

The planning agent will parse the user's question, assess available data and environmental resources, perform high-level reasoning, and allocate subtasks to specialized sub-agents. These sub-agents will use the environment's tools to execute their subtasks step by step and return feedback to the planning agent, which will iteratively update the plan until the overall task is completed.

By embedding domain knowledge into developing data preprocessing and modeling steps, the agent can work with the user to reason about the appropriateness of the analytic choices and make specific recommendations so the analysis plan mirrors best practices to address the scientific questions of interest.

2 The Result Agent with Reasoning

Professor Lin's suggestion to develop a result agent that collaborates with users to map analytic outputs onto field-specific interpretive frameworks and reporting norms is equally compelling. Trustworthy findings and contextualized result interpretation are essential for an intelligent data analysis agent to be practically useful.

Domain knowledge and community standards guide not only how data are analyzed but also how results should be validated, interpreted, and reported. For example, GWAS findings are typically validated in independent cohorts and followed by functional follow-up of implicated variants. A result agent should therefore be able to:

- Map results to domain-field-specific established interpretive frameworks.
- Compare findings with the literature to assess novelty and significance.
- Evaluate result validity.
- Provide interpretations that incorporate existing domain-specific knowledge.

- Caution against overinterpretation.

Accomplishing this requires scientific reasoning that integrates curated literature, ontologies, domain-specific reporting templates, and field guidelines. The development of such capabilities represents a significant but important challenge that would substantially enhance LAMBDA's practical utility.

Human-in-the-loop interaction is a strength in LAMBDA, although its current implementation focuses mainly on execution code error correction rather than deeper scientific dialogue. To increase real-world utility, human-agent dialogue should be adapted in multiple agents, including both the analysis-planning and result agents, by engaging users in deeper scientific conversation.

Specifically, the analysis-planning agent would collaborate with the user on designing conversational mechanisms that develop a comprehensive analysis plan meeting scientific goals. The result agent would enhance result intelligence by collaborating with the user through dialogue to examine results, make suggestions to improve analysis, and provide result interpretation.

Incorporating interactive visual interfaces in the result agent would allow users to examine whether the analysis plan and results make sense by casting them in the literature and providing recommendations for additional analysis. This feedback loop would allow the system to adapt dynamically and ensure that analytical results remain aligned with the user's scientific goals while providing scientific insights that guide further analysis and support meaningful, trustworthy result interpretation.

3 Enhancing Analytical and Scientific Reasoning

Professor Lin correctly notes that while the Programmer and Inspector agents execute code following user instructions and fix execution errors, they do not evaluate whether the analytic procedure and results are scientifically meaningful or contextualized within the literature, nor do they provide scientific insights of the results.

To improve analytical and scientific reasoning, future development could explore multi-layer reasoning architectures, assisted by a reasoning agent, that evaluate the conceptual validity of an analysis plan and the strengths and weaknesses of results. Such an architecture would combine statistical and data-science reasoning with domain-expert reasoning to ensure scientific justification and findings are sound and interpretable, while making suggestions for additional analyses based on current results.

A tree-of-thought-style component could further deepen reasoning by evaluating alternative analytic routes and interpretations before committing to a specific approach. This would allow the system to consider multiple hypotheses and present trade-offs to the user, enabling more informed decision-making.

We are also exploring reinforcement learning approaches for statistical reasoning. Professor Fan Zhou and Professor Bang Liu propose framing data analysis as a reinforcement learning environment to enhance reasoning capabilities. LAMBDA's modular architecture provides a natural foundation for this, where agent actions and environment states can be clearly defined.

The proposed framework includes a process-oriented action space that explicitly utilizes control tokens to delineate reasoning steps, and a hybrid reward model that combines rule-based feedback with an "LLM-as-a-judge" mechanism. Optimizing against these metrics would incentivize the agent to prioritize deep analytical insights over superficial execution.

4 Conclusion

We thank Professor Xihong Lin for her insightful suggestions regarding the development of analysis-planning and result agents with enhanced reasoning capabilities. These enhancements would transform LAMBDA into a domain-aware analytic assistant that ensures analyses both answer users' scientific questions and align with community standards, while providing greater practical utility and scientific contributions.

The addition of reasoning capabilities to LAMBDA represents an exciting frontier that aligns with our ongoing research agenda. We are committed to exploring multi-agent architectures that incorporate planning, analysis, and result interpretation in a unified framework that maintains human oversight and customization at every step.

References

Maojun Sun, Ruijian Han, Binyan Jiang, Houduo Qi, Defeng Sun, Yancheng Yuan, and Jian Huang. Lambda: A large model based data agent. *Journal of the American Statistical Association*, pages 1–13, 2025a.

Maojun Sun, Ruijian Han, Binyan Jiang, Houduo Qi, Defeng Sun, Yancheng Yuan, and Jian Huang. A survey on large language model-based agents for statistics and data science. *The American Statistician*, pages 1–14, 2025b.