

PS 01.003 Developing and Benchmarking Agentic LLM Frameworks for Accelerated and Accessible Proteomics Data Analysis

Maximilien Burq, Aljaz Potocnik, Dejan Stepec, Yelena Bronevetsky, Charles C. Kim, Peter Cimermercic



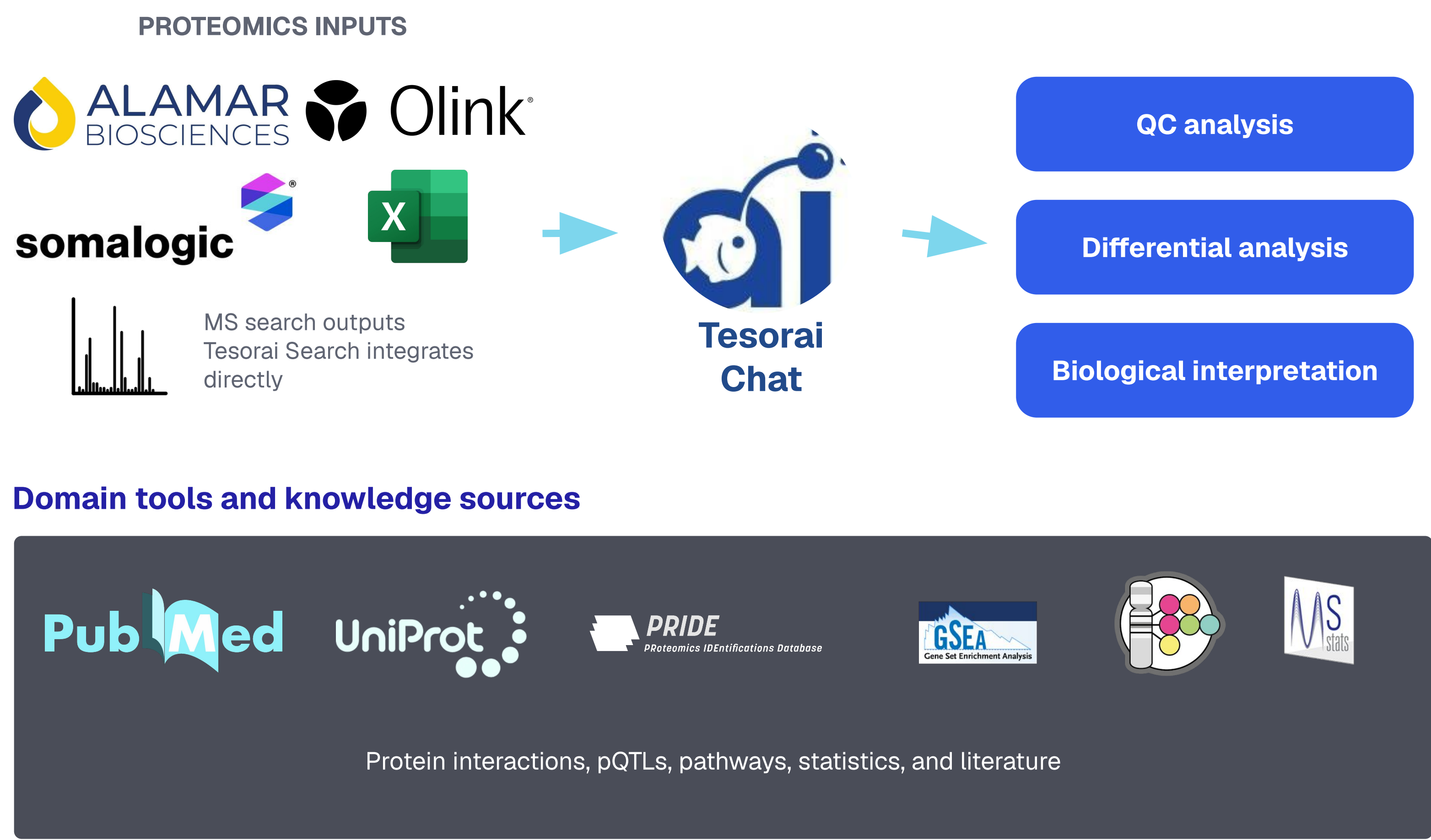
Background. Proteomics data analysis is notoriously complex, characterized by diverse experimental designs (abundance proteomics, cross-linking, affinity purification, spatial proteomics), analytical approaches (DDA/DIA, TMT, Olink), high data sparsity, and significant batch effects. Currently, extracting biological insights requires a specialized team with expertise in bioinformatics and statistics. While Large Language Models (LLMs) offer a potential interface for wet-lab scientists to perform rapid data exploration, their domain-specific accuracy and reliability remain unproven. We developed Tesorai Chat, an agentic framework designed to bridge this gap, and evaluated its performance against leading generalist and specialist models.

Methods. Tesorai Chat employs a multi-agent planning mode integrated with domain-specific libraries and external knowledge graphs (e.g., protein interactions, pQTLs, and literature). We developed a benchmark of 94 common proteomics use cases, categorized by complexity (easy, medium, hard). We compared Tesorai Chat against Claude 4.6 Sonnet, GPT-5.4 Codex, and Phylo. Each query was executed three times to assess both accuracy (percentage of correct outputs) and repeatability (consistency of the logic/answers across runs).

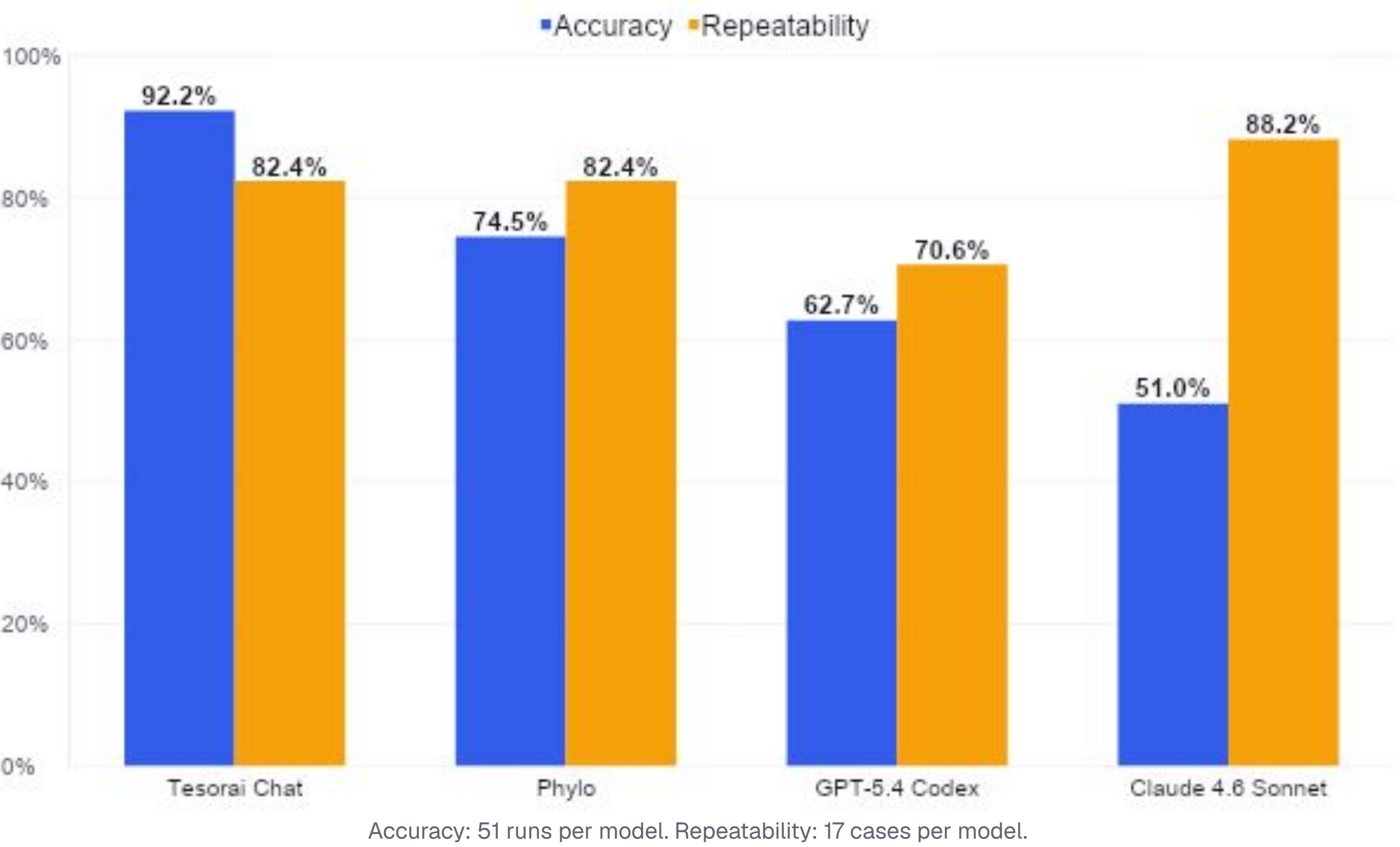
Results. Tesorai Chat significantly outperformed other models in accuracy, achieving 92.2%, followed by Phylo (74.5%), GPT-5.4 Codex (62.7%), and Claude 4.6 Sonnet (51.0%). Claude showed the highest repeatability (88.2%), while Tesorai Chat and Phylo both scored 82.4%. GPT-5.4 Codex scored 70.6% for repeatability.

Conclusions. General-purpose LLMs currently lack the precision required for reliable proteomics analysis. Agentic frameworks like Tesorai Chat augment LLMs with specialized tools and knowledge bases, providing the accuracy needed to support researchers without continuous bioinformatics intervention. Domain-specific agentic AI can make complex proteomics workflows faster and more reproducible.

Tesorai Chat connects diverse proteomics inputs to domain-specific analysis



92.2% accuracy across benchmarked proteomics analyses



Benchmark coverage

- Data extraction
- Data computation
- Statistical analysis
- Knowledge and database lookup
- API and knowledge-graph queries
- Web and literature lookup

Example workflow in Tesorai Chat

Start from a guided analysis or ask your own question

Chat executes the analysis and returns an interpretable result

If we're not here,
find us at Booth #2
in Innovation Row.

