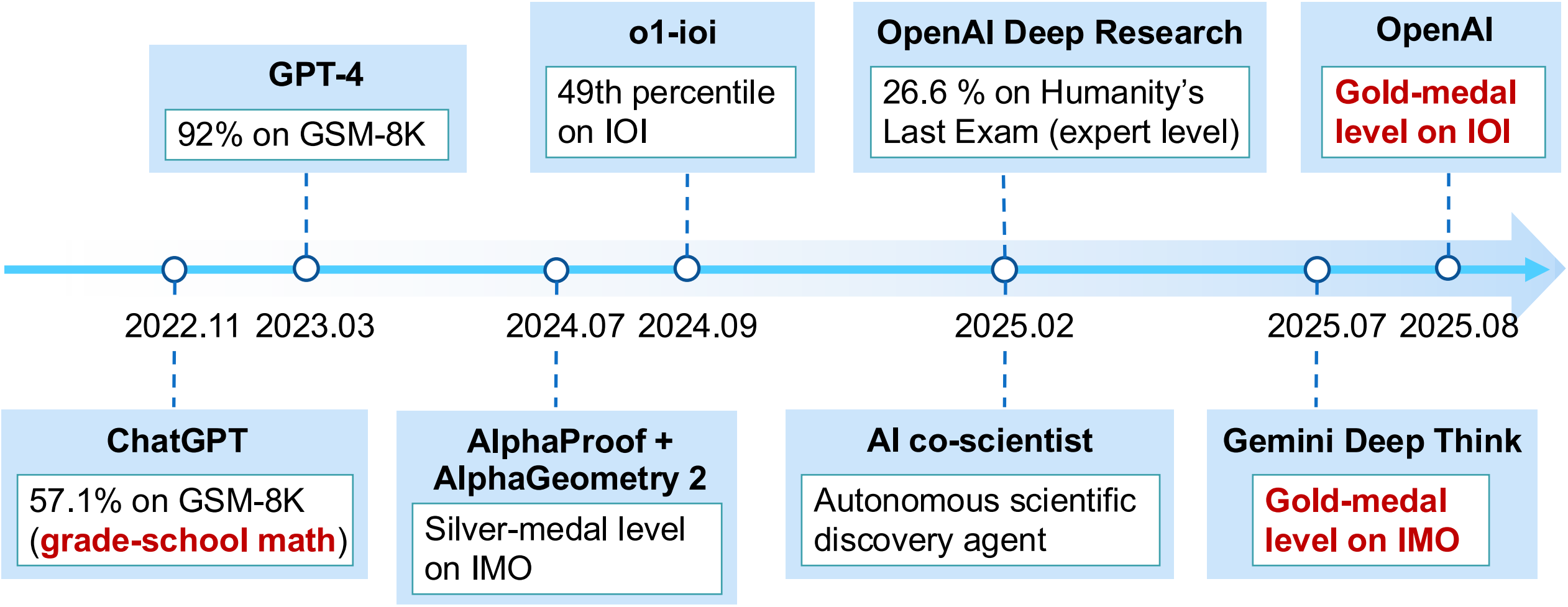
数学建模智能体理论与实战

刘浩 博士
香港科技大学（广州）
2025年12月26日

数学建模智能体

MM-Agent: LLM as Agents for Real-world Mathematical Modeling Problem. NeurIPS 2025

The Intellectual Development of LLMs



The Reasoning Barrier in Real-World Applications

Well-defined

Problem 1. A line in the plane is called sunny if it is not parallel to any of the x-axis, the y-axis, and the line $x + y = 0$.

Let $n \geq 3$ be a given integer. Determine all nonnegative integers k such that there exist n distinct lines in the plane satisfying both of the following:

- for all positive integers a and b with $a + b \leq n + 1$, the point (a, b) is on at least one of the lines; and
- exactly k of the n lines are sunny.

Well-defined

Gardens by the Bay is a large nature park in Singapore. In the park there are n towers, known as supertrees. These towers are labelled 0 to $n - 1$. We would like to construct a set of **zero or more** bridges. Each bridge connects a pair of distinct towers and may be traversed in **either** direction. No two bridges should connect the same pair of towers.

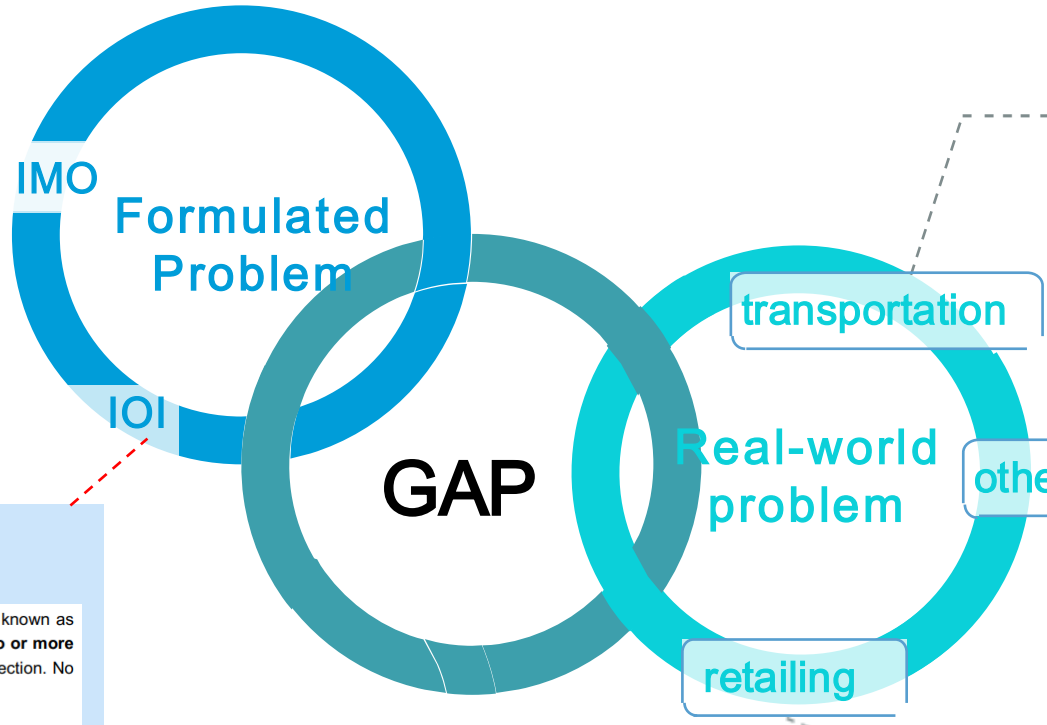
A path from tower x to tower y is a sequence of one or more towers such that:

- the first element of the sequence is x,
- the last element of the sequence is y,
- all elements of the sequence are **distinct**, and
- each two consecutive elements (towers) in the sequence are connected by a bridge.

Note that by definition there is exactly one path from a tower to itself and the number of different paths from tower i to tower j is the same as the number of different paths from tower j to tower i.

The lead architect in charge of the design wishes for the bridges to be built such that for all $0 \leq i, j \leq n - 1$ there are exactly $p[i][j]$ different paths from tower i to tower j, where $0 \leq p[i][j] \leq 3$.

Construct a set of bridges that satisfy the architect's requirements, or determine that it is impossible.



Open-ended

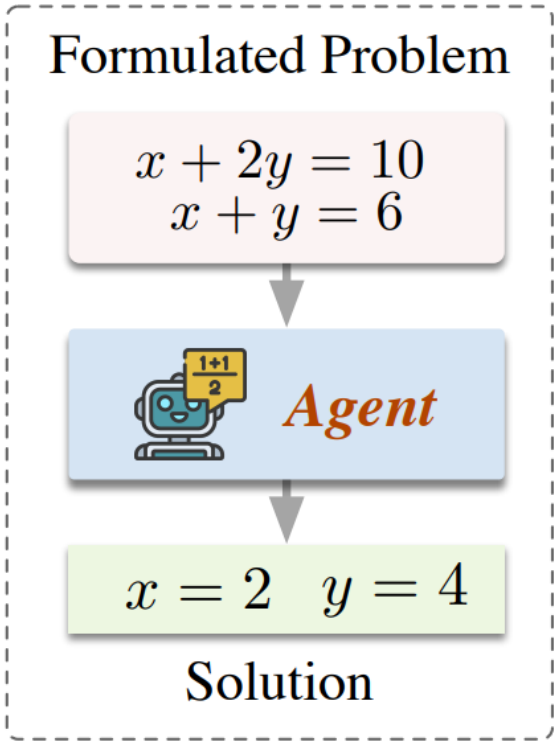
How should a city plan its public transportation system to maximize efficiency and minimize environmental impact?

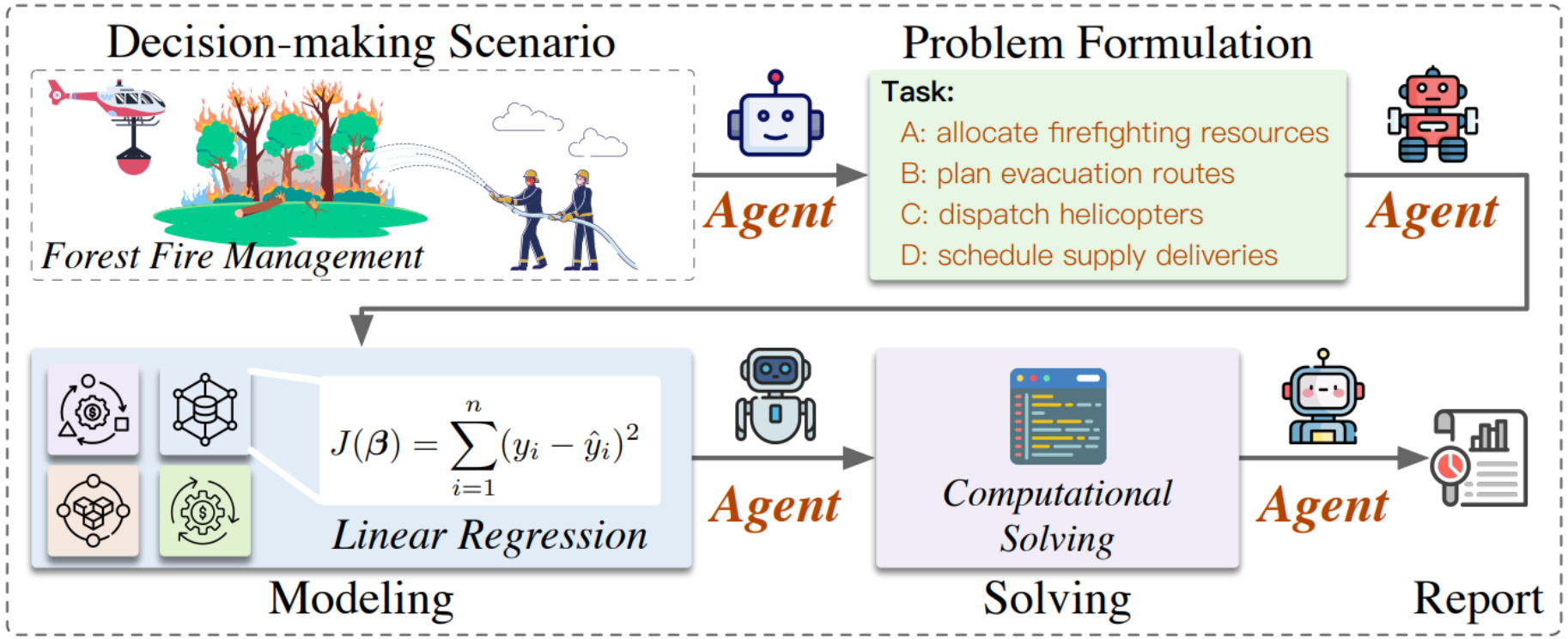
...

Fuzzy description

How should a retail store optimize its staff scheduling to minimize labor costs while ensuring customer satisfaction?

Our Work: MM-Agent





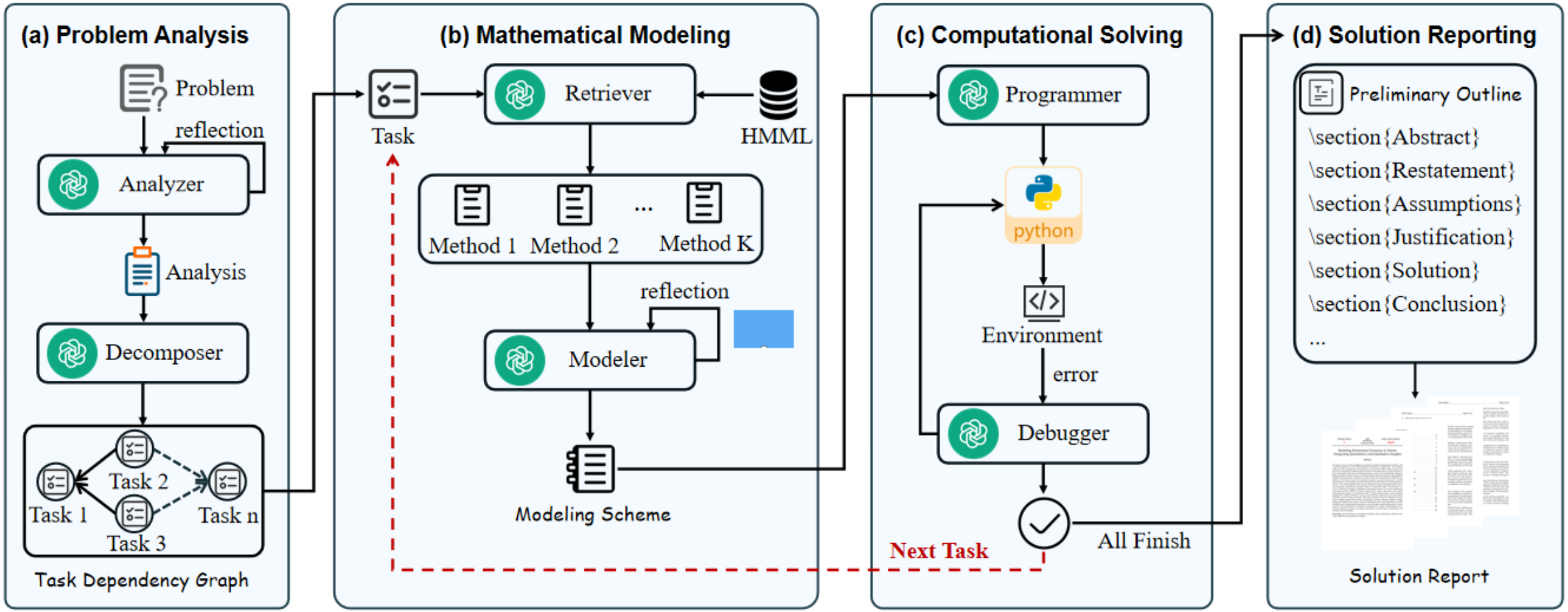
Solving well-defined problem in textbook

Solving open-ended Mathematical Modeling problem in the real-world

MM-Agent assisted two undergraduate teams in winning the Finalist Award (top 2.0% among 27,456 teams) in MCM/ICM 2025.

Architecture of MM-Agent

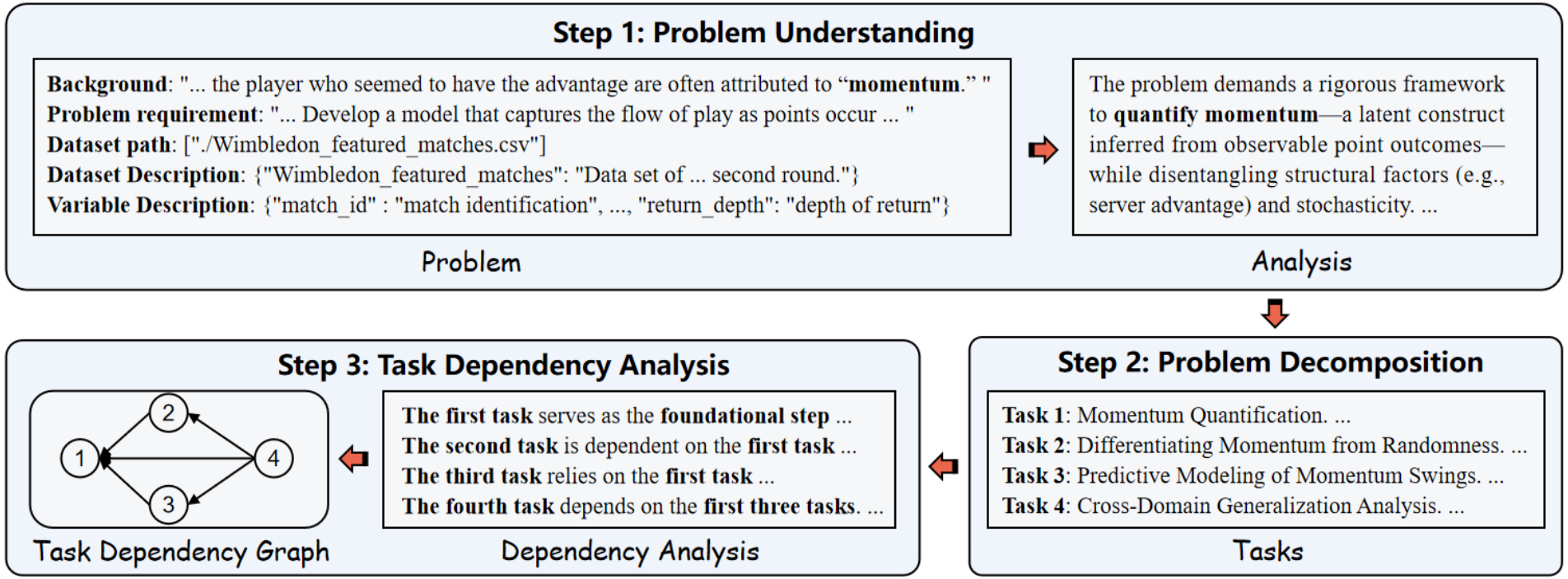
- MM-Agent automates the problem formulation and structured mathematical modelling process.



MM-Agent: LLM as Agents for Real-world Mathematical Modeling Problem. NeurIPS 2025.

Problem Analysis

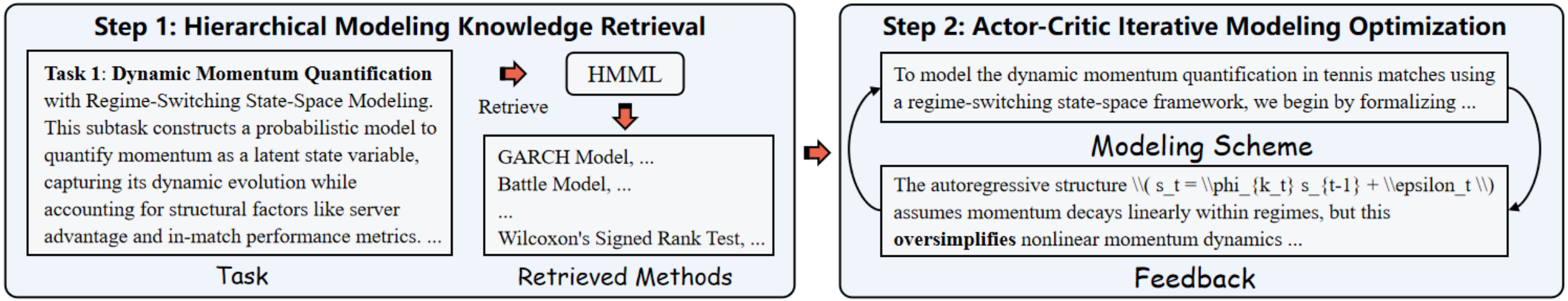
- **Automated problem understanding, problem decomposition, and task dependency analysis**
  - 1.Extract background, goals, and key information. 2. Identify subtasks based on components. 3. Map logical/variable dependencies between subtasks .



Mathematical Modeling

☐ Hierarchical Mathematical Modeling Optimization

- ➢ Mimic human cognition: retrieve modeling methods from a knowledge base and solve each subtask via actor-critic optimization.

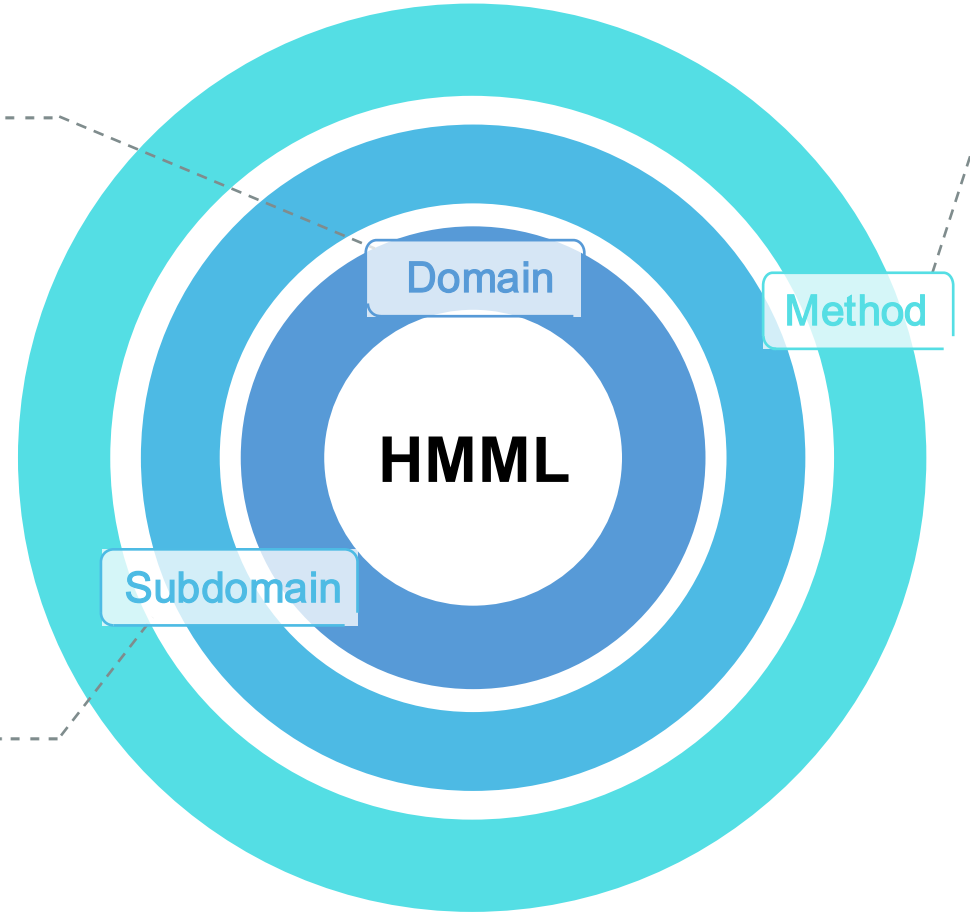


Hierarchical Mathematical Modeling Library

- To enhance the mathematical modeling capabilities of LLM agents, we constructed a structured three-level hierarchy designed to improve retrieval effectiveness.

Domains of mathematical modeling methods, including Optimization, Machine Learning, Operations Research, Prediction, and Evaluation.

Subdomains of mathematical modeling methods, such as Operations Research, which includes Graph Theory, Programming Theory, Stochastic Programming, and so on.



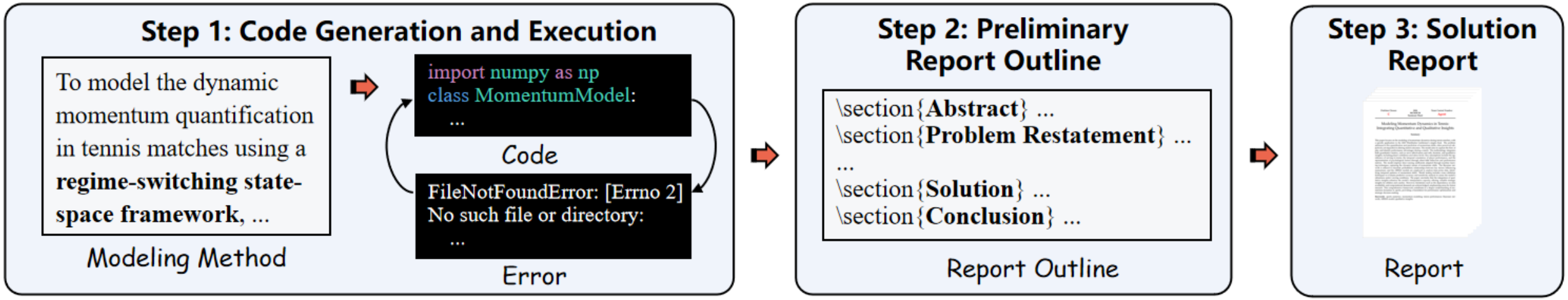
Specific mathematical modeling methods, such as Linear Programming, which include:

<Method>: Linear Programming
<core_idea>: ... linear objective functions and constraints
<application>: ... resource allocation, production planning, ...

Computational Solving and Solution Reporting

☐ Code execution based problem solving and report generation

- ➢ Use MLE-Solver to conduct experiments and auto-generate a structured report with results and findings



How to Evaluate Open-ended, Ill-Defined Tasks?

- **Benchmark Design:** Mathematical modeling competitions offer a natural, open-ended setting for evaluating LLM agents, where problems mirror real-world complexity and require abstraction, reasoning, and solution design.

- **Evaluation:**

Analysis Evaluation
- Is the problem clearly defined?
- Are key factors and logic between sub-tasks well-structured?

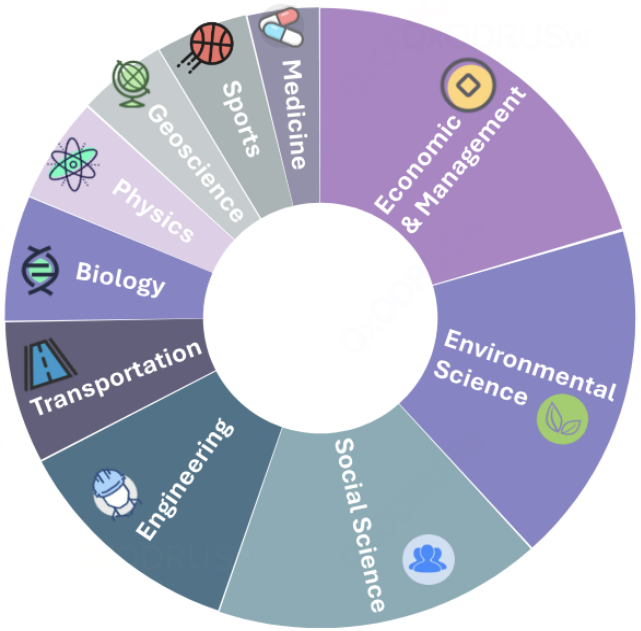
Modeling Rigorousness
- Are assumptions reasonable and stated explicitly?
- Are methods and models accurate, rational, and scientific?

Practicality & Scientificity
- Is the model applicable in real-world scenarios?
- Does it yield useful, scientifically valid insights?

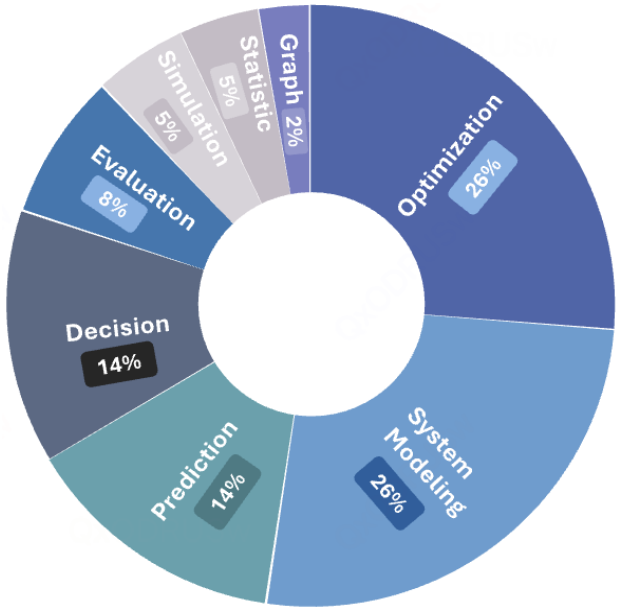
Result & Bias Analysis
- Are the results interpretable and in-depth?
- Is bias addressed and robustness ensured

MM-Bench

- **Data Summary:** MM-Bench consists of 10 domains, 8 task types, and a total of 111 problem samples, all sourced (MCM and ICM).



Domains



Problem Type

Problem example:

Background: "In the 2023 Vmbledon Gentlemen's final,.... The incredible swings, sometimes for many points or even games, that occurred in the player who seemed to have the advantage are often attributed to "momentum."

Dataset path: / Wimbledon_featured_ matches': 'Data set of ... second round.

Variable Description: "match_id': 'match identification',

Problem Example: 2024 C

Experiments Results

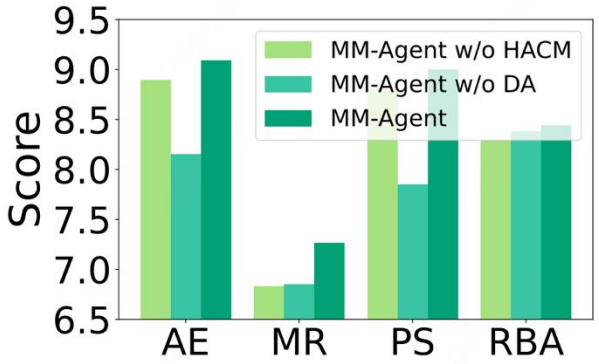
- **MM-Agent** achieves **SOTA performance** across all metrics and backbone models.
- Results are **robust across years**, showing no sign of data leakage.

Table 1: Experimental results on the 2021–2024 and 2025 mathematical modeling competitions. AE, MR, PS, and RBA denote *Analysis Evaluation*, *Modeling Rigorousness*, *Practicality and Scientificity*, and *Result and Bias Analysis*, respectively.

Methods	2021–2024					2025				
	AE ↑	MR ↑	PS ↑	RBA ↑	Overall ↑	AE ↑	MR ↑	PS ↑	RBA ↑	Overall ↑
Human										
Human Team	9.04	6.20	8.79	7.62	7.91	9.25	7.42	8.92	6.50	8.02
GPT-4o										
GPT-4o	7.62	3.86	8.48	5.17	6.28	7.67	3.67	8.90	5.75	6.50
DS-Agent	8.18	7.08	8.72	7.47	7.86	8.25	**7.33**	8.92	7.10	7.90
ResearchAgent	7.97	6.80	8.82	7.37	7.74	8.00	7.30	8.60	7.00	7.73
Agent Laboratory	8.56	6.35	8.63	5.56	7.28	8.75	5.58	8.58	5.33	7.13
MM-Agent	**9.15**	**7.28**	**9.00**	**8.44**	**8.85**	**8.86**	7.21	**9.00**	**8.43**	**8.38**
DeepSeek-R1-671B										
DeepSeek-R1	7.23	4.79	8.69	4.50	6.30	7.42	4.25	8.50	5.25	6.35
DS-Agent	8.25	6.88	8.74	7.19	7.77	7.92	6.33	9.00	7.60	7.71
ResearchAgent	8.13	7.04	8.77	6.92	7.72	8.00	6.75	8.83	7.58	7.79
Agent Laboratory	8.65	5.96	8.70	5.91	7.31	8.83	5.50	8.83	5.58	7.19
MM-Agent	**9.54**	**8.25**	**9.06**	**8.54**	**8.85**	**9.50**	**8.33**	**9.25**	**8.58**	**8.92**

Experiments Results

- **Ablation Study**: We evaluate MM-Agent by removing key components:
  - *w/o DA* (no task dependency analysis)
  - *w/o HACM* (no hierarchical actor-critic modeling)
- **Cost Efficiency:** MM-Agent achieves **competitive performance** with **low computational cost**.



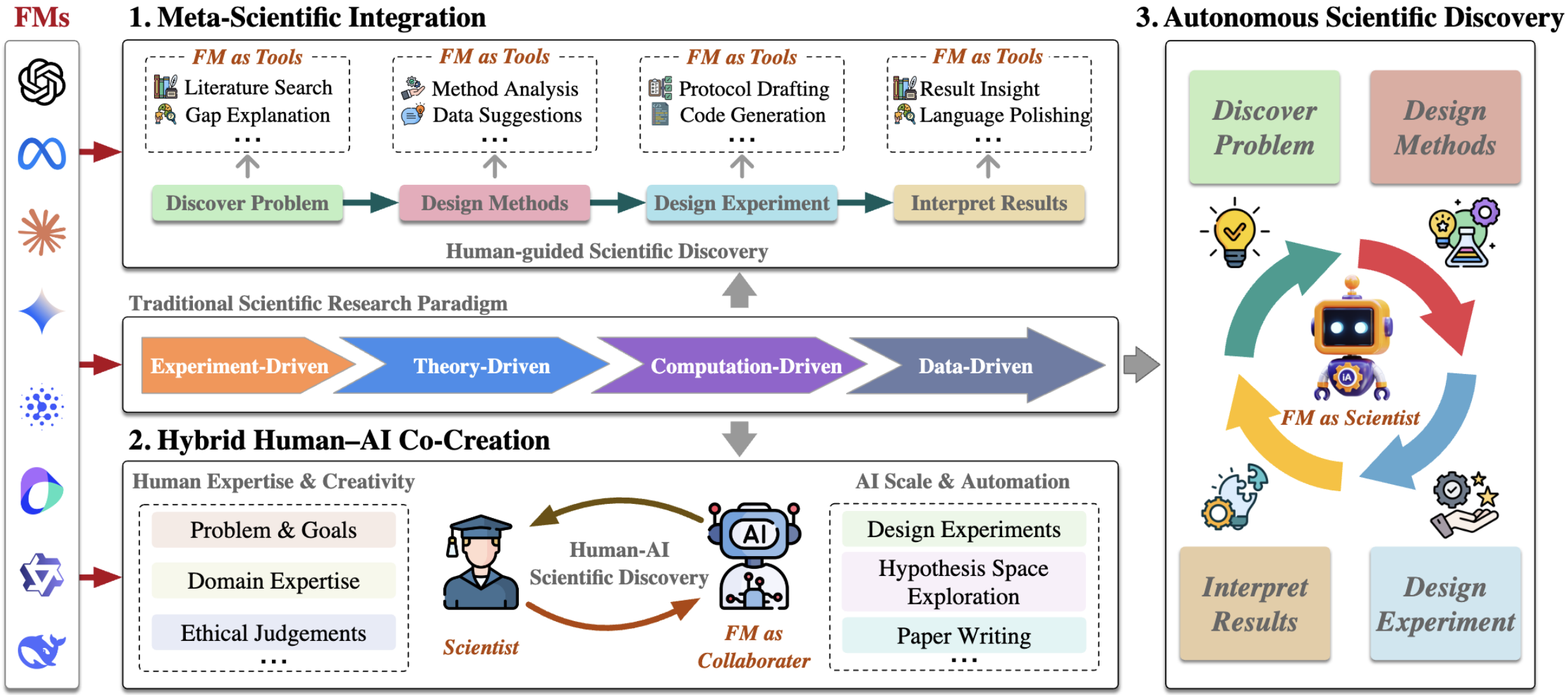
(a) GPT-4o

(b) DeepSeek-R1-671b

Table 2: Experimental results on average token consumption, cost, and runtime.

Methods	Token	Cost($)	Runtime(s)
GPT-4o			
DS-Agent	198,186	0.77	1,044
ResearchAgent	170,732	0.67	459
Agent Laboratory	746,159	2.14	1,015
MM-Agent	240,877	0.88	906
DeepSeek-R1-671b			
DS-Agent	315,171	0.42	6,444
ResearchAgent	222,030	0.28	4,816
Agent Laboratory	974,313	0.88	10,595
MM-Agent	530,363	0.56	7,529

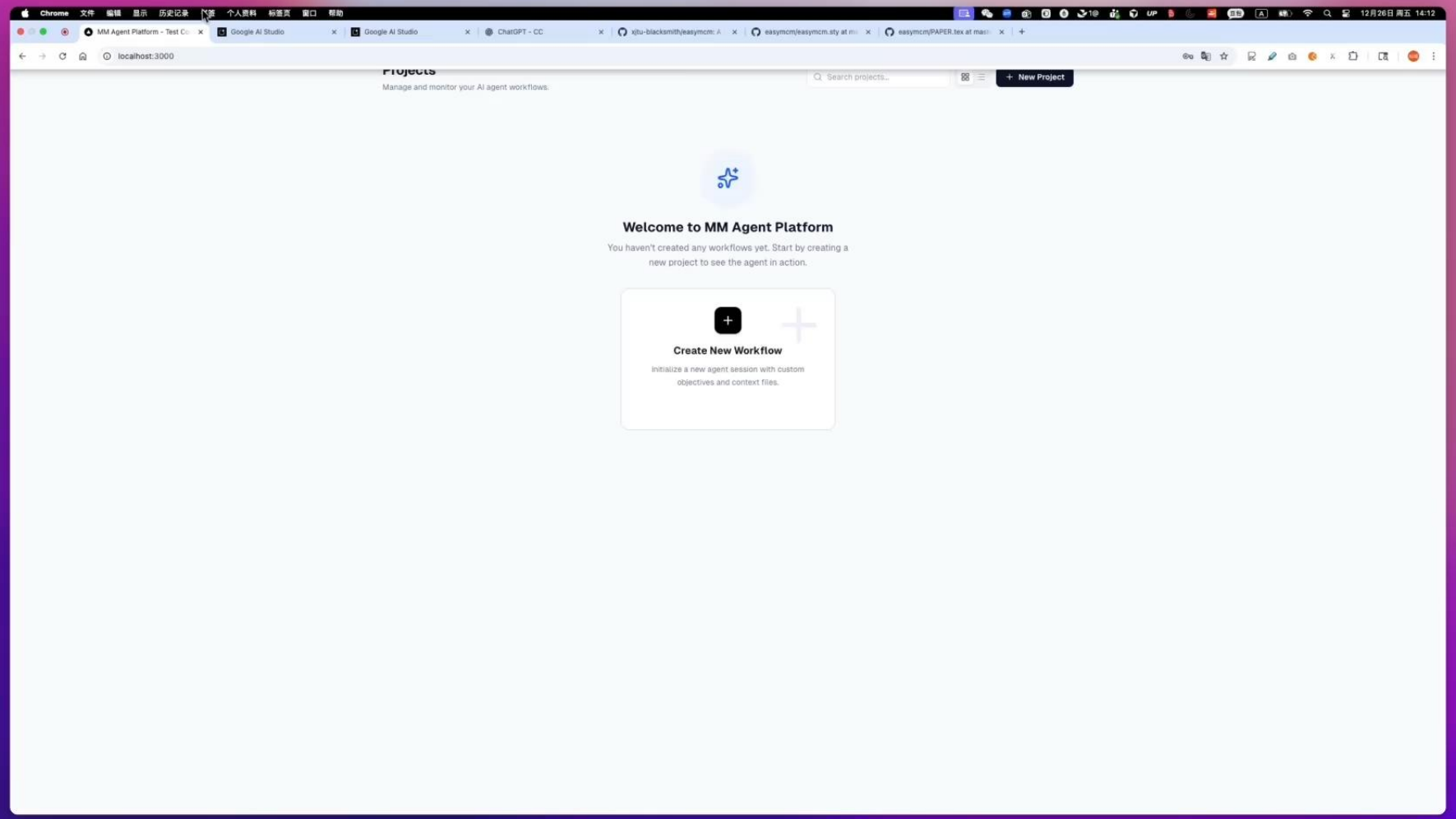
Future Directions



Foundation Models for Scientific Discovery: From Paradigm Enhancement to Paradigm Transition. NeurIPS 2025

数学建模智能体实战

Demo



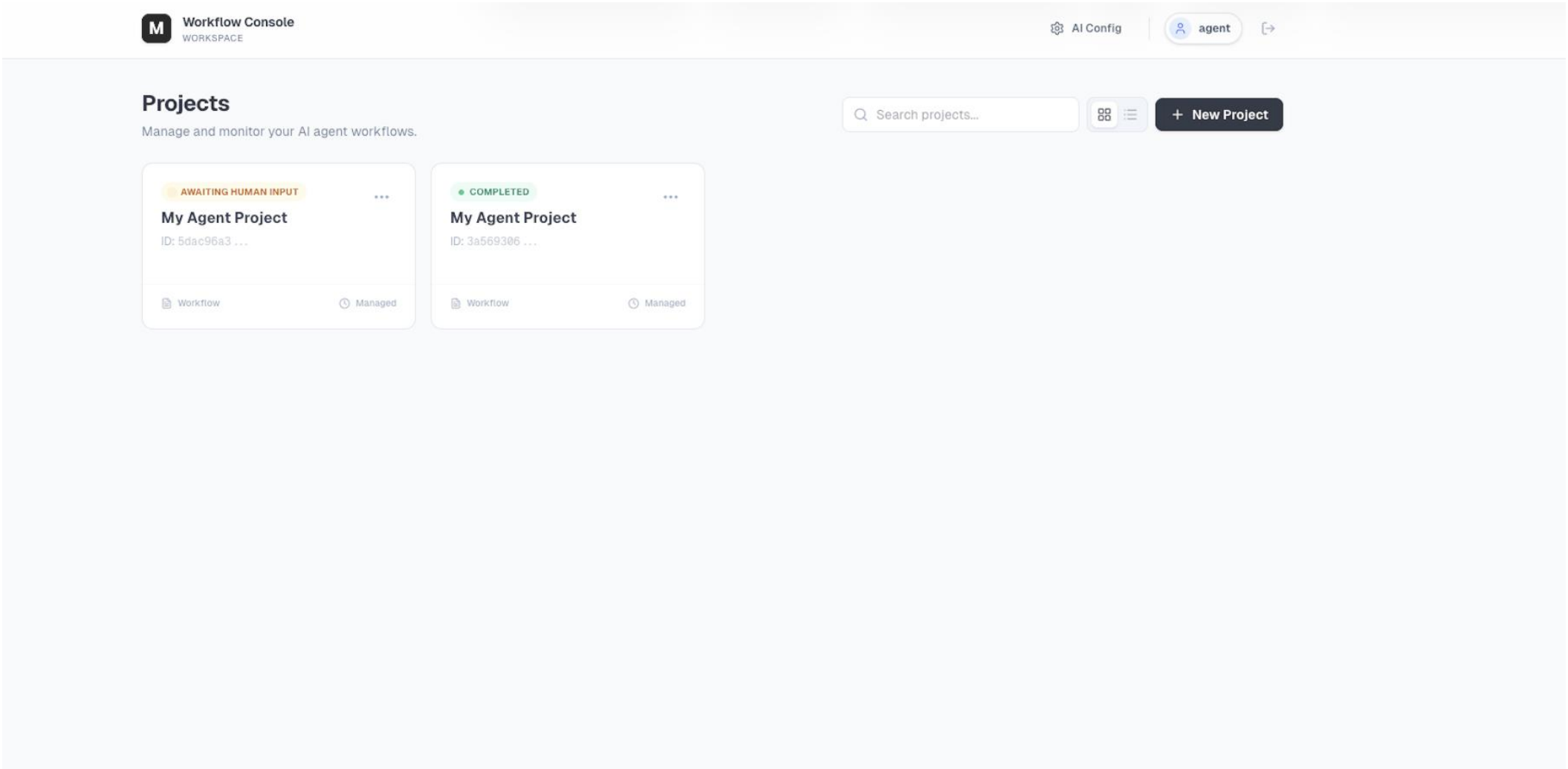
项目链接

群聊：美赛数据建模 agent

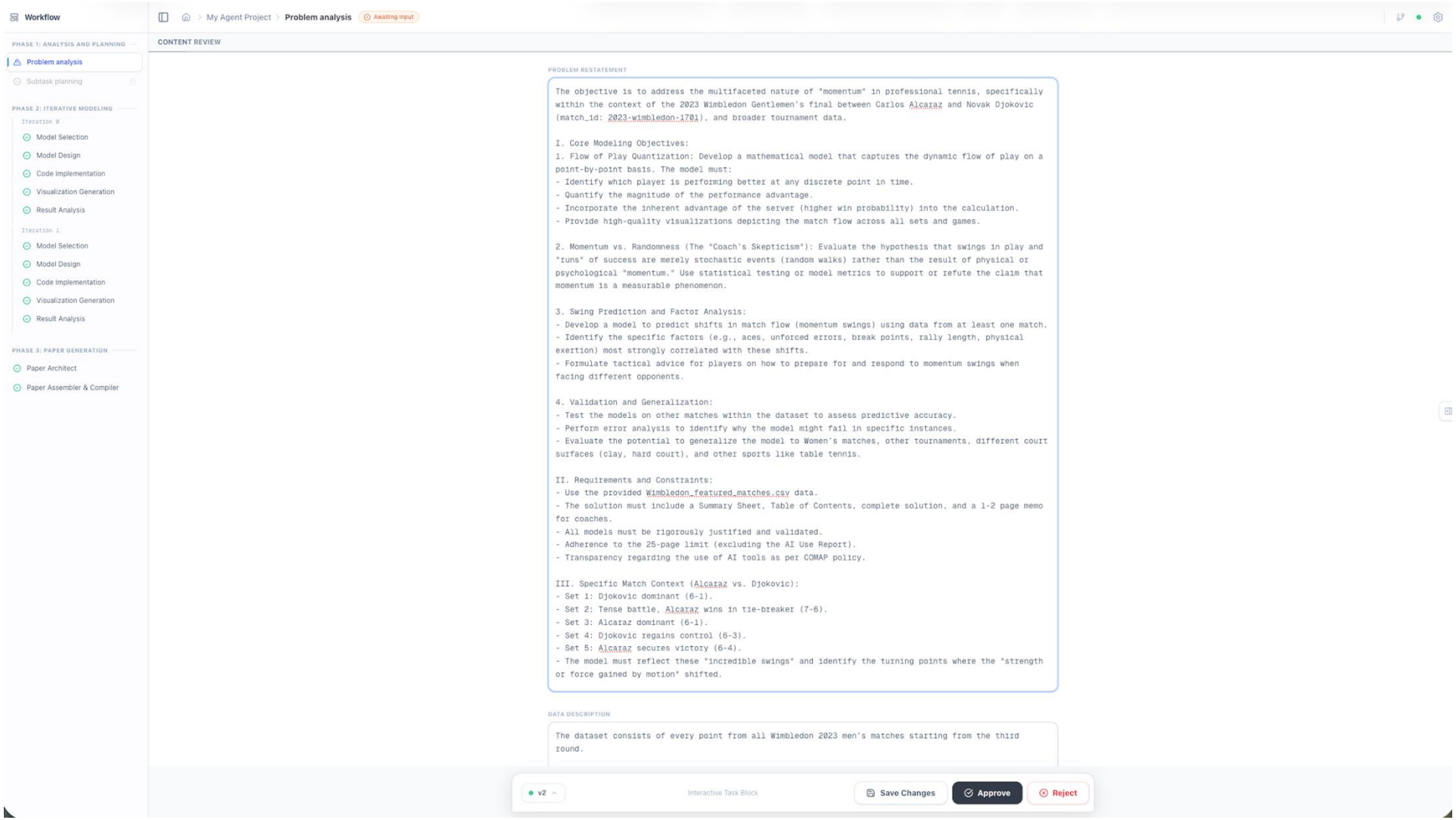
该二维码7天内(1月2日前)有效，重新进入将更新

项目agent服务群

管理界面：创建自己的数学建模项目。

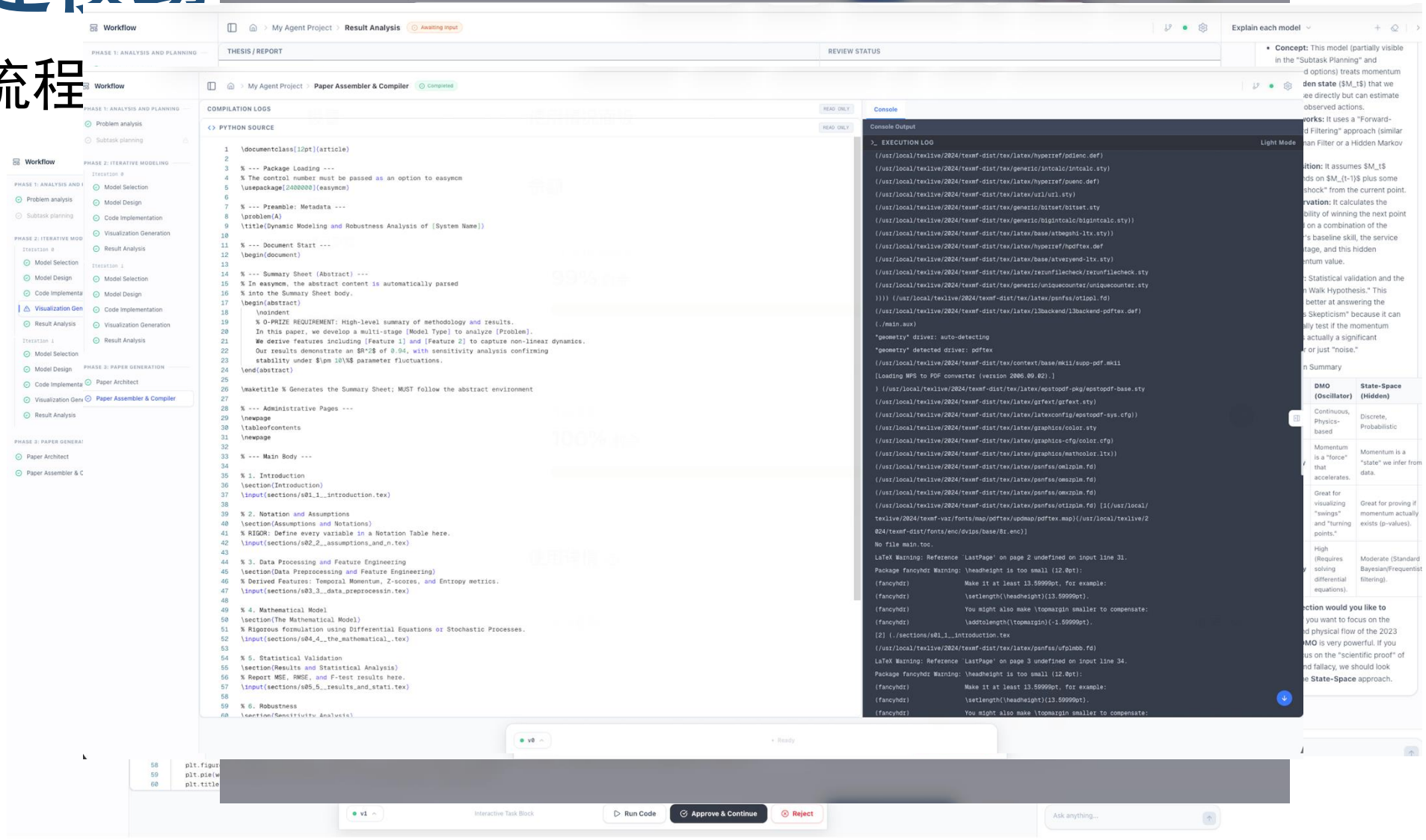


项目建立：上传问题与数据集。



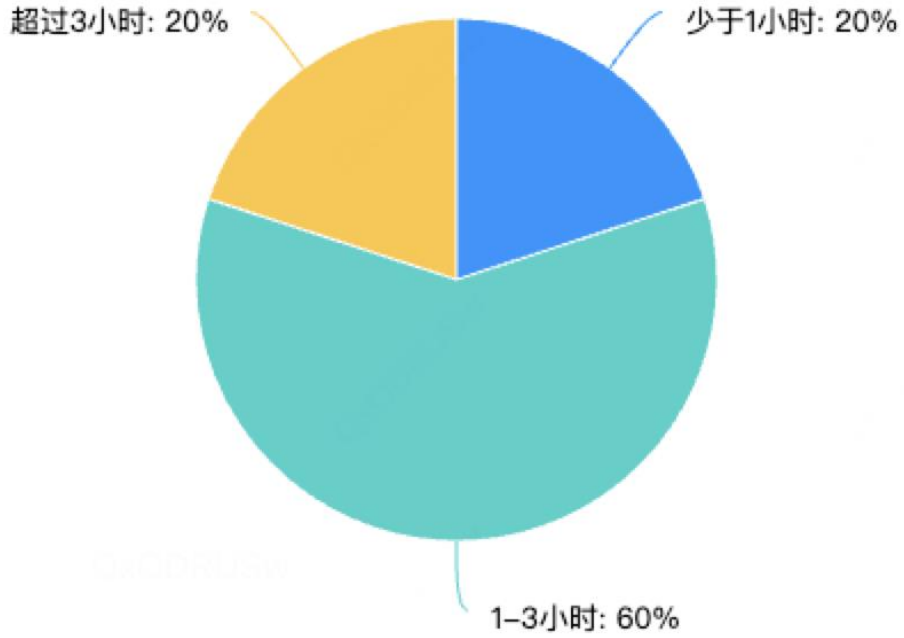
数学建模助

- 建模流程

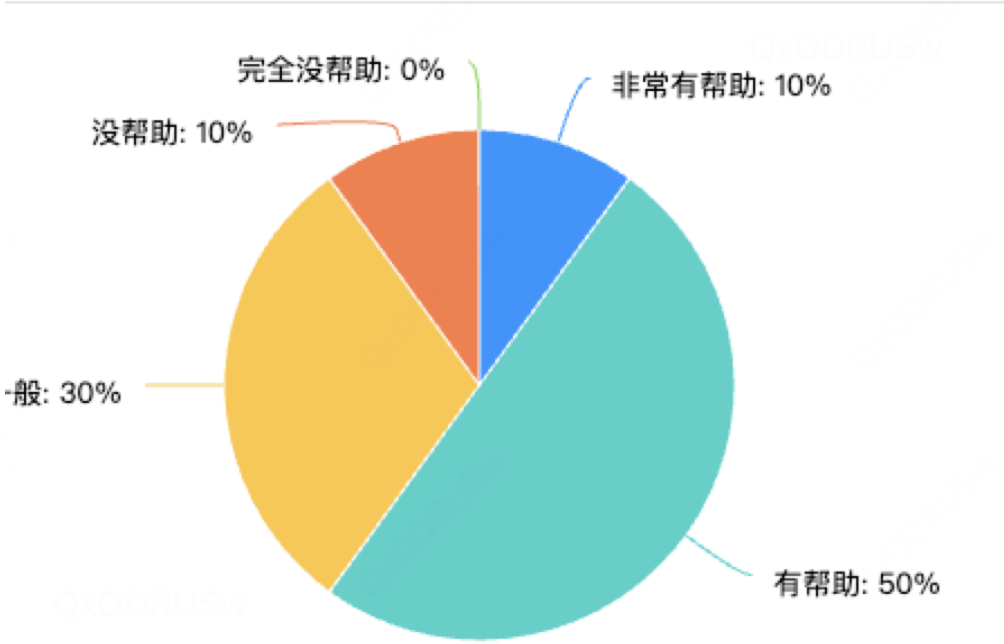

Workflow
My Agent Project
Result Analysis
Awaiting Input
PHASE 1: ANALYSIS AND PLANNING
Problem analysis
Subtask planning
PHASE 2: ITERATIVE MODELING
Model Selection
Model Design
Code Implementation
Visualization Generation
Result Analysis
PHASE 3: PAPER GENERATION
Paper Architect
Paper Assembler & Compiler
THESIS / REPORT
REVIEW STATUS
Explain each model
Paper Assembler & Compiler
Completed
COMPILATION LOGS
PYTHON SOURCE
READ ONLY
Console
Console Output
EXECUTION LOG
Light Mode
\documentclass[12pt]{article}
% --- Package Loading ---
% The control number must be passed as an option to easymcm
\usepackage[2400000]{easymcm}
% --- Preamble: Metadata ---
\problem{A}
\title{Dynamic Modeling and Robustness Analysis of {System Name}}
% --- Document Start ---
\begin{document}
% --- Summary Sheet (Abstract) ---
% In easymcm, the abstract content is automatically parsed
% into the Summary Sheet body.
\begin{abstract}
\noindent
% O-PRIZE REQUIREMENT: High-level summary of methodology and results.
In this paper, we develop a multi-stage [Model Type] to analyze [Problem].
We derive features including [Feature 1] and [Feature 2] to capture non-linear dynamics.
Our results demonstrate an R^2 of 0.94, with sensitivity analysis confirming
stability under $\pm 10\%$ parameter fluctuations.
\end{abstract}
\maketitle % Generates the Summary Sheet; MUST follow the abstract environment
% --- Administrative Pages ---
\newpage
\tableofcontents
\newpage
% --- Main Body ---
% 1. Introduction
\section{Introduction}
\input{sections/s01_1__introduction.tex}
% 2. Notation and Assumptions
\section{Assumptions and Notations}
% RIGOR: Define every variable in a Notation Table here.
\input{sections/s02_2__assumptions_and_n.tex}
% 3. Data Processing and Feature Engineering
\section{Data Preprocessing and Feature Engineering}
% Derived Features: Temporal Momentum, Z-scores, and Entropy metrics.
\input{sections/s03_3__data_preprocessin.tex}
% 4. Mathematical Model
\section{The Mathematical Model}
% Rigorous formulation using Differential Equations or Stochastic Processes.
\input{sections/s04_4__the_mathematical_.tex}
% 5. Statistical Validation
\section{Results and Statistical Analysis}
% Report MSE, RMSE, and F-test results here.
\input{sections/s05_5__results_and_stati.tex}
% 6. Robustness
\section{Sensitivity Analysis}
(/usr/local/texlive/2024/texmf-dist/tex/latex/hyperref/pdfenc.def)
(/usr/local/texlive/2024/texmf-dist/tex/generic/intcalc/intcalc.sty)
(/usr/local/texlive/2024/texmf-dist/tex/latex/hyperref/puenc.def)
(/usr/local/texlive/2024/texmf-dist/tex/latex/url/url.sty)
(/usr/local/texlive/2024/texmf-dist/tex/generic/bitset/bitset.sty
(/usr/local/texlive/2024/texmf-dist/tex/generic/bigintcalc/bigintcalc.sty))
(/usr/local/texlive/2024/texmf-dist/tex/latex/base/atbegshi-ltx.sty))
(/usr/local/texlive/2024/texmf-dist/tex/latex/hyperref/hpdftex.def
(/usr/local/texlive/2024/texmf-dist/tex/latex/base/atveryend-ltx.sty)
(/usr/local/texlive/2024/texmf-dist/tex/latex/rerunfilecheck/rerunfilecheck.sty
(/usr/local/texlive/2024/texmf-dist/tex/generic/uniquecounter/uniquecounter.sty
)))) (/usr/local/texlive/2024/texmf-dist/tex/latex/psnfss/ot1ppl.fd)
(/usr/local/texlive/2024/texmf-dist/tex/latex/l3backend/l3backend-pdftex.def)
(./main.aux)
"geometry" driver: auto-detecting
"geometry" detected driver: pdftex
(/usr/local/texlive/2024/texmf-dist/tex/context/base/mkii/supp-pdf.mkii
[Loading MPS to PDF converter (version 2006.09.02).]
) (/usr/local/texlive/2024/texmf-dist/tex/latex/epstopdf-pkg/epstopdf-base.sty
(/usr/local/texlive/2024/texmf-dist/tex/latex/grfext/grfext.sty)
(/usr/local/texlive/2024/texmf-dist/tex/latex/latexconfig/epstopdf-sys.cfg))
(/usr/local/texlive/2024/texmf-dist/tex/latex/graphics/color.sty
(/usr/local/texlive/2024/texmf-dist/tex/latex/graphics-cfg/color.cfg)
(/usr/local/texlive/2024/texmf-dist/tex/latex/graphics/mathcolor.ltx))
(/usr/local/texlive/2024/texmf-dist/tex/latex/psnfss/omlzplm.fd)
(/usr/local/texlive/2024/texmf-dist/tex/latex/psnfss/omszplm.fd)
(/usr/local/texlive/2024/texmf-dist/tex/latex/psnfss/omxzplm.fd)
(/usr/local/texlive/2024/texmf-dist/tex/latex/psnfss/ot1zplm.fd) [1{/usr/local/
texlive/2024/texmf-var/fonts/map/pdftex/updmap/pdftex.map}{/usr/local/texlive/2
024/texmf-dist/fonts/enc/dvips/base/8r.enc}]
No file main.toc.
LaTeX Warning: Reference `LastPage' on page 2 undefined on input line 31.
Package fancyhdr Warning: \headheight is too small (12.0pt):
(fancyhdr) Make it at least 13.59999pt, for example:
(fancyhdr) \setlength{\headheight}{13.59999pt}.
(fancyhdr) You might also make \topmargin smaller to compensate:
(fancyhdr) \addtolength{\topmargin}{-1.59999pt}.
[2] (./sections/s01_1__introduction.tex
(/usr/local/texlive/2024/texmf-dist/tex/latex/psnfss/ufplmbb.fd)
LaTeX Warning: Reference `LastPage' on page 3 undefined on input line 34.
Package fancyhdr Warning: \headheight is too small (12.0pt):
(fancyhdr) Make it at least 13.59999pt, for example:
(fancyhdr) \setlength{\headheight}{13.59999pt}.
(fancyhdr) You might also make \topmargin smaller to compensate:
DMO (Oscillator)
State-Space (Hidden)
Continuous, Physics-based
Discrete, Probabilistic
Momentum is a "force" that accelerates.
Momentum is a "state" we infer from data.
Great for visualizing "swings" and "turning points."
Great for proving if momentum actually exists (p-values).
High (Requires solving differential equations).
Moderate (Standard Bayesian/Frequentist filtering).
v0
v1
Interactive Task Block
Run Code
Approve & Continue
Reject
Ask anything...

用户调研

- 2025年助力两个美赛队伍取得F奖。
- 调研结果：对于建模思路有很大的帮助！



使用时长



帮助程度

谢谢聆听！

刘浩

北京 海淀

liuh@ust.hk

https://raymondhliu.github.io/

香港科技大学(广州)
THE HONG KONG
UNIVERSITY OF SCIENCE AND
TECHNOLOGY (GUANGZHOU)