Report on
Global Artificial Intelligence Championships (GAIC) Math 2024

AGI Odyssey

July 5, 2024

1 Summary

The Global Artificial Intelligence Championships (GAIC) Math 2024 organized by AGI Odyssey was a pioneering contest that aimed to advance the development of artificial intelligence tools for solving advanced mathematical problems across multiple levels of difficulty. The contest attracted wide participation, with 55 submissions from 32 teams, showcasing the significant interest and progress in this emerging field.

The problems spanned three levels – High School, university-level, and Olympiad – testing participants' abilities in areas ranging from foundational algebra and geometry to advanced calculus, linear algebra, and logic-intensive Olympiad problems. The open-ended nature of many questions, requiring precise numerical, algebraic, or vector answers, presented a considerable challenge that could not be easily overcome through guesswork.

The top three finishers achieved impressive results, with the first-place team scoring an overall 69.3% accuracy on 387 problems. The second-place team provided details on their multi-agent system based on large language models like GPT-4, incorporating techniques such as prompt engineering, chain-of-thought reasoning, and specialized methods like table lookup for statistical problems. However, details on the solutions from the first- and third-place teams were unavailable.

The contest highlighted the potential of AI to enhance mathematical problem-solving and understanding while fostering interdisciplinary collaboration. The insights gained will prove invaluable for refining future AI mathematics competitions, ensuring they continue pushing boundaries while inspiring participants and driving innovation in this rapidly evolving field.

The contest significantly contributed to the intersection of artificial intelligence and mathematical problem-solving. It highlighted the potential of AI to enhance understanding and innovation in solving complex mathematical problems. Feedback from this event will be instrumental in refining the structure and content of future contests, ensuring they continue to challenge and inspire participants while advancing the field of AI for mathematics.

The rest of this report summarizes the contest's structure, questions, participant engagement, and results, aiming to serve as a reference for the development of future mathematical AI contests.

2 Contest Structure and Test Question Sets

2.1 Contest structure

The contest was announced in January 2024 on a wide range of social media, including the X, huggingface.co, Reddit, Medium, and WeChat, opening to worldwide participants. The award prizes were set to be 75,000 USD for the first place, 20,000 USD for the second, and 5,000 USD for the third place. **The top-ranked teams will be invited to present their work at a GAIC Symposium in July 2024.**

Participants could be individuals, teams (up to 6 members), or organizations. They retained ownership of their AI models. Registration was free and open until March 1, 2024.

The 24-hour contest took place on March 16, 2024, starting at 12am Eastern Standard Time. The test questions were released in both PDF and Latex formats on the GAIC website. Participants submitted their responses on Codabech. All participants were required to share their responses on the competition leaderboard post-submission. Each participant was allowed a maximum of 5 submission attempts for the contest. GAIC recognized the submission with the highest accuracy from each team as its final score.

Post-contest requirements. Participants must visit the AGI Odyssey GAIC website post-contest to provide details about the AI models utilized. This step ensures transparency and aids in post-contest analysis. Only teams that provide specific details about the models they used will be recognized for participating in the competition and eligible for the final awards.

Executive committee. The executive committee oversaw the contest, playing a pivotal role in orchestrating the math competition from start to finish. Their responsibilities ranged from recruiting and selecting a diverse set of contest problems to finalizing them. They established the competition criteria and developed comprehensive guidelines to direct the contest operation team. This guidance ensured the smooth execution of the event and the fairness and efficiency of the evaluation process, ultimately contributing to the overall success and integrity of the contest.

2.2 Test question sets

The contest features various original mathematical problems categorized into three difficulty levels: High School, University, and mathematical Olympiad. Each category was carefully designed to test different aspects of mathematical theory and problem-solving ability:

- **High School Level** focused on foundational topics such as geometry, algebra, and (pre-)calculus.
- **University Level** focused on subjects typically studied by non-math majors, such as calculus, linear algebra, probability, statistics, and ordinary and partial differential equations.
- **Olympiad Level** comprised of highly challenging problems similar to those found in International Mathematical Olympiad, testing deep logical reasoning and mathematical skills.

Every test question has an objective answer that can be graded automatically. The final answers have three types: True-false (T-F), Multiple-choice questions (MCQ), and Open-answer (Open). In particular, the Open-answer questions have answers in various forms, including real numbers, polynomial expressions, vectors, and arrays. Thus, such questions reduce the possibility of guessing a correct answer to near zero.

Table 1 summarizes the distribution of question types.

Table 1: Distribution of Question Types by Difficulty Category

Category	**T-F**	**MCQ**	**Open**	Total
Olympiad	0	6	142	148
High School	1	120	17	138
University	15	0	86	101
Total	16	126	245	387

Mathematics professionals, ranging from experienced high school educators to university professors, carefully designed these original test questions. They were invited by AGI Odyssey to contribute their expertise to the contest. The executive committee was responsible for test question generation, review, formatting, testing, and revisions for GAIC Math 2024.

Table 2: Percentage of correctness of the Top 9 Teams

Team	Overall	High School	College	Olympic	T-F	MCQ	Open
qilixiang	69.3	**87.9**	72.0	**49.7**	**81.25**	**90.4**	**58.2**
Blacksheep	64.6	83.6	**81.0**	35.4	75.0	85.6	53.3
Good	63.8	85.7	71.0	38.1	75.0	87.2	51.6
ABC	62.8	85.0	69.0	37.4	68.75	88.0	50.0
sciencelover	61.0	85.0	68.0	33.3	75.0	87.2	47.1
robospace	43.9	81.4	42.0	9.5	62.5	83.2	23.0
overthink	39.0	32.9	72.0	22.4	43.75	26.4	45.5
lynn9675	27.9	48.6	34.0	4.1	50.0	47.2	16.8
mathx	27.4	44.3	35.0	6.1	50.0	40.0	19.3

3 Participating Teams and Results

3.1 Results overview

The competition attracted broad participation, with 48 registered teams representing an international mix from the United Kingdom, China, and the United States. Among these, 68.75% 33 teams competed individually, while 31.25% (15 teams) engaged in a group format. The competition received 55 submissions from 32 teams with diverse backgrounds in mathematics and computer science. A team from the United Kingdom achieved second place. Unfortunately, the teams that secured first and third places did not respond to our multiple email inquiries and follow-ups.

Table 2 and Figure 1 present the detailed results for the top 9 teams, highlighting their performance across various question levels and answer types.

All the top five teams achieved impressive overall accuracy above 61%, with the first-place team scoring 69.3%. These top five scores were significantly higher than the sixth score, which was 39%.

Results by question levels. Teams consistently score higher at the high school level problems than in university or Olympiad levels, indicating that these problems are relatively easier for the AI models. In particular, for the high-school level questions, the top five teams' scores were close, distributed between 85% and 87.9%, with the best score achieved by the first-place team. For the university level questions, the scores ranged from 68% to 81% among the top five teams. The best score was achieved by the team 'Blacksheep', and it was 9% better than the second-best. This suggested the strength or focus of the team 'Blacksheep' might be on more advanced undergraduate-level problems. The scores drastically dropped in the Olympiad category, illustrating the challenging nature of Olympiad level problems for all teams. Among the top five teams, scores ranged from 33.3% to 49.7%, for which the best score was 15.3% better than second-best.

Overall, the first-place team achieved the best-scores in the high school and Olympiad levels, while the second-placed team won the best-scores for the university level.

Results by answer types. The top five teams displayed high accuracy in True-false and Multiple-choice questions, with True-false scores between 71.4% and 78.6%, and Multiple-choice scores from 87.5% to 91.7%. This performance indicates that the top-ranked teams employed effective strategies in developing AI solutions to solve math problems in standardized test formats. Notably, the scores for True-false questions among the top five teams were quite close, with four teams achieving exactly 71.4%.

For open-answer questions, their scores ranged from 48.2% to 58.6%. Considering the challenging nature of these questions, these results are commendable.

The first-place team stood out significantly, achieving the highest scores across all three question types by substantial margins.

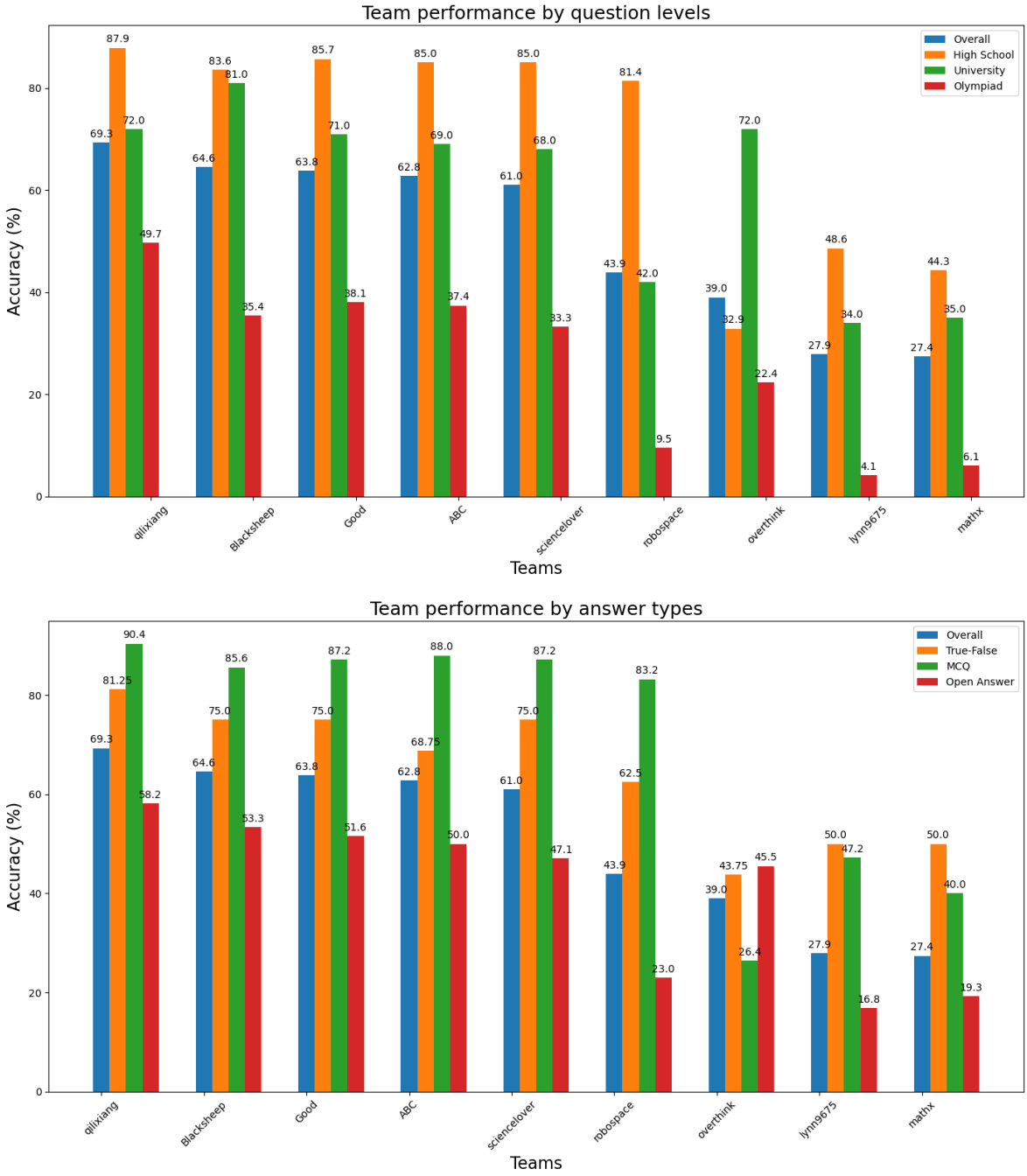

Figure 1: Chart displaying scores of the top 9 teams across various question levels and answer types.

3.2 Overview of the top three models

The second-place team provided detailed model information, while the other teams were out of touch after the contest.

- The first-place team's model: unavailable.

- The second-place team's model is available to the public at GitHub https://github.com/Jeff0741/GAIC-2024-Spring-EVS.git. The team designed a multi-agent solving system based on large language models, employing Prompts Engineering, CoT, Multi-Agents, Agents Interaction, and Tools. All the llm-based agents were based on `gpt-4-turbo` with `temprature=0`. They selected different solving methods for different problems, crafted more professional prompts for some challenging problems, and adopted a specialized table lookup method for statistical problems that require table lookup. The time required to complete a solution to a question was 5-25 minutes.
- The third-place team's model: unavailable.

Acknowledgement

The GAIC Math 2024 was made possible by the support from NetMind.AI and Protagolabs, the contributions from the supporting team, the mathematics professionals, the advisory board, and the executive committee.

Supporting team

Jianyang Chen	NetMind.AI
Siliang Deng	Protagolabs
Xinyao Hu	Protagolabs

Test question contributors

Dan Chen	Professor, Arizona State University, USA
Lingxia Chen	Educator, Potomac Falls High School, Virginia, USA
Chuangye Liu	Professor, Central China Normal University, China
Xiaoming Song	Professor, Drexel University, PA, USA
Junyan Zhang	Professor, University of Singapore
Hongjuan Zhou	Professor, Arizona State University, AZ, USA