MathNet: a Global Multimodal Benchmark for Mathematical Reasoning and Retrieval

Shaden Alshammari[*1] **Kevin Wen**[*1] **Abrar Zainal**[*3] **Mark Hamilton**[1]
Navid Safaei[3] **Sultan Albarakati**[2] **William T. Freeman**[1] **Antonio Torralba**[1]

[1] MIT [2] KAUST [3] Individual Researcher [*] Equal Contribution

Abstract

Mathematical problem solving remains a challenging test of reasoning for large language and multimodal models. However, existing benchmarks are limited in size, language coverage, and task diversity. We introduce *MathNet*, a large-scale, multilingual, and multimodal dataset of Olympiad-level problems. MathNet spans 40 countries, 10 languages, and two decades of competitions, comprising 13,026 expert-authored problems with solutions across diverse domains.

MathNet supports two tasks: (i) mathematical comprehension and (ii) mathematical retrieval, an underexplored but essential capability. For retrieval, we construct 39K pairs of mathematically equivalent problems to enable equivalence-based evaluation. Experimental results show that even state-of-the-art reasoning models (72% and 66% accuracy for GPT-5 and Gemini 2.5 Pro) are challenged, while embedding models exhibit substantial difficulty in retrieving equivalent problems.

MathNet provides the largest multilingual Olympiad dataset and the first retrieval benchmark for mathematical equivalence, which we will publicly release. [1]

1 Introduction

Recent LLMs and LMMs have made rapid strides on mathematical reasoning benchmarks, from grade-school arithmetic to competition mathematics [2, 10, 18]. In 2025, public reports claimed gold-medal–level performance at the International Mathematical Olympiad (IMO) by advanced AI systems [7, 19]. Moreover, there have been multiple incidents of AI systems reportedly solving open mathematical problems [17, 4].

Despite these advances, progress remains constrained by the lack of large, diverse, and systematically annotated benchmarks. Existing Olympiad-level datasets are typically drawn from community platforms such as AoPS and are predominantly English-only (see Table 1), restricting both linguistic and cultural coverage. To address this gap, we present *MathNet*—the first large-scale, multilingual, and multimodal dataset of Olympiad-level problems. Curated over two decades from 40 countries and spanning 10 languages, *MathNet* comprises 13,026 expert-authored problems across a wide range of mathematical domains. Its scale, diversity, and expert quality provide an unprecedented foundation for exploring mathematical generalization, cross-lingual transfer, and analogical reasoning.

Building on this foundation, we focus on a fundamental yet underexplored capability: *retrieving mathematically equivalent or related problems*. Unlike general semantic retrieval [14, 13, 15, 5], mathematical retrieval must be sensitive to symbolic structure, invariances, and transformations. For example, the problem of solving $x^2 + y^2 = 1$ is equivalent to one that poses $\sqrt{a^2 + b^2} = 1$,

[1] The dataset, benchmark, and demos are available at http://mathnet.netlify.app/.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI.

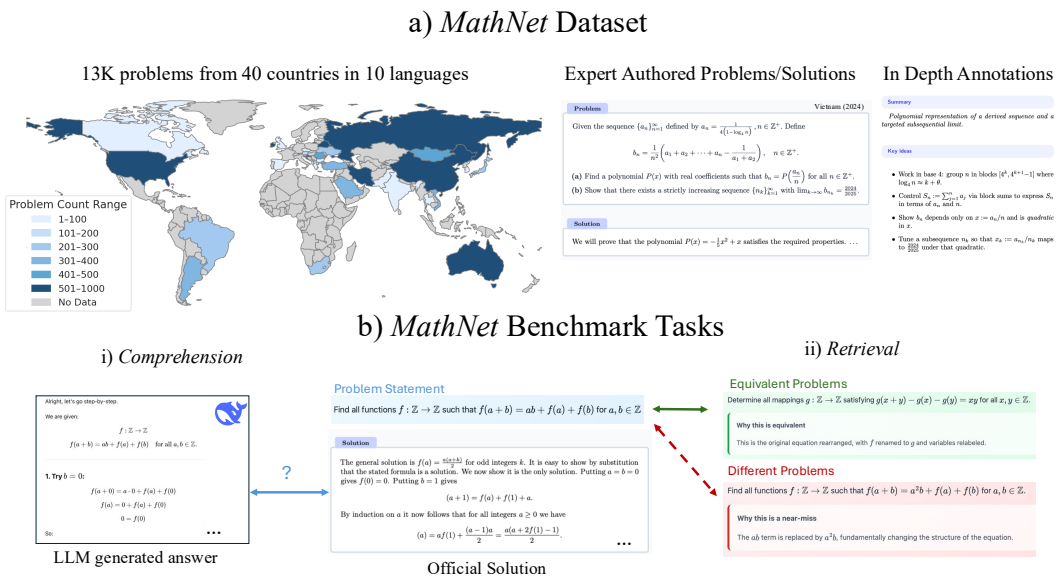


Figure 1: **Overview of MathNet.** (a) Dataset of 13K Olympiad-level problems across 40 countries and 10 languages with expert-authored solutions. (b) Benchmark tasks: comprehension (solution generation) and retrieval (equivalence-based problem matching).

or to a geometric formulation constraining a 2D vector to unit norm $|u|^2 = 1$. Crucially, however, these are not equivalent to solving $x + y = 1$. Current retrieval models fail to make this distinction: due to superficial lexical overlap [3], they often rank a problem containing $x + y = 1$ as closer to $x^2 + y^2 = 1$ than to the truly equivalent formulations.

A similar difficulty arises in research search. When mathematicians want to locate papers relevant to an expression such as $p_{i+1} - p_i \leq \Pi_i$, they often resort to paraphrased keyword queries like "bounds between consecutive primes," rather than retrieval based on symbolic structure.

This paper introduces *MathNet*, a benchmark designed to evaluate *math-aware retrieval* and its role in reasoning. Our contributions are:

1. **Dataset.** A 13K-problem corpus of Olympiad-style math with aligned LaTeX and natural-language statements, expert solutions, and metadata spanning 40+ countries and 10 languages.
2. **New Annotations and Similarity Axes.** 39,078 synthetic problem pairs with labeled equivalence classes, enriched with a principled taxonomy of similarity axes (invariance, resonance, and affinity) and 82 common math Olympiad concepts.
3. **Large-Scale Evaluation.** Benchmarking across 16 models on two primary tasks that measure mathematical comprehension and retrieval quality.

2 Dataset

We introduce *MathNet*, a large-scale benchmark designed to evaluate the reasoning and retrieval abilities of large language models (LLMs) and large multimodal models (LMMs). The benchmark contains both text-only and interleaved text–image problems, presented in multiple languages to broaden accessibility. In total, *MathNet* comprises 13,026 problems with expert-written solutions.

A key feature of *MathNet* is its fine-grained taxonomy of mathematical similarity, which enables systematic analysis of model performance across varying levels of structural and semantic overlap. We also define a novel retrieval task that tests models' ability to identify structurally related problems. Baseline models and evaluations highlight the benchmark's utility in assessing both problem-solving accuracy and mathematical understanding.

Benchmark	Size	Languages	Evaluation Type	Multimodal	Source	Difficulty
GSM8k [2]	8,500	EN	Numeric Answer	×	Crowdsourced grade-school problems	Grade School
MATH [10]	12,500	EN	Numeric Answer	×	Competition / textbook problems	High School
MATH-Vision [21]	3,040	EN	Expression / Proof	✓	Math Competitions	High School Olympiad
CMMLU [16]	1,594	ZH	MCQ	×	Chinese exam materials	High School / College
MMLU [9]	2,554	EN	MCQ	×	University / professional exams	College-Level Knowledge
C-Eval [11]	3,362	ZH	MCQ	×	Chinese college entrance / certification exams	College Entrance Exams
MMMU [22]	3,007	EN	MCQ / Expression	✓	Multimodal academic exams (college level)	College-Level Multimodal
AGIEval [23]	3,300	EN & ZH	MCQ / Expression	×	University entrance / qualification exams	University Entrance Exams
JEEBench [1]	515	EN	MCQ / Numeric Answer	×	Indian JEE Advanced past papers	JEE Advanced Exam
OlympiadBench [8]	6,142	EN & ZH	Proof / Expression	✓	AoPS Forum	Olympiad Level
OlympicArena [12]	3,233	EN & ZH	Proof / Process Evaluation	✓	AoPS Forum	Olympiad Level
Omni-Math [6]	4,428	EN	Proof / Process Evaluation	✓	AoPS Forum / Contest Pages	Olympiad Level
MathNet (ours)	13,026	EN, ZH, ES, RU AR, RO, DE, FA DE, UK	Proof / Process Evaluation	✓	Official Country Booklets and National Contests	Olympiad Level

Table 1: Comparison of mathematical reasoning benchmarks across different sizes, languages, evaluation types, and difficulty levels. We include both unimodal and multimodal datasets, spanning grade-school to Olympiad-level mathematics. Our proposed **MathNet** expands coverage to 10 languages and focuses on proof- and process-based evaluation with authentic national contest problems.

Data sources. Unlike prior benchmarks that rely on community platforms such as AoPS, *MathNet* is curated exclusively from officially published national contest materials, ensuring expert quality and consistency. The benchmark draws from 643 contest volumes across 40 countries (2006–2025). Full details are in Appendix A.1.

Pipeline overview. To construct the dataset, we standardized raw PDFs into structured Markdown, extracted aligned problem–solution pairs, and verified correctness through cross-model agreement. The pipeline uses multilingual OCR, LLM-based extraction, and dual-model validation with Humans and LLMs. For a complete description of each stage (document ingestion, solution retrieval, and semantic verification), see Appendix A.2.

Data Preparation and Release Our benchmark includes 13,026 problems, with 2,000 designated for model-based evaluation as *MathNet-ot*. We sample 1,000 problems across subjects to create *MathNet-val* for hyperparameter tuning or small-scale testing. *MathNet-val* problems have step-by-step solutions, supporting research like process-level evaluation. The remaining problems form *MathNet-test*, the official test set with unreleased answers for formal testing. The results in this paper are based on the entire benchmark dataset, including *MathNet-ot*, *MathNet-val*, and *MathNet-test*.

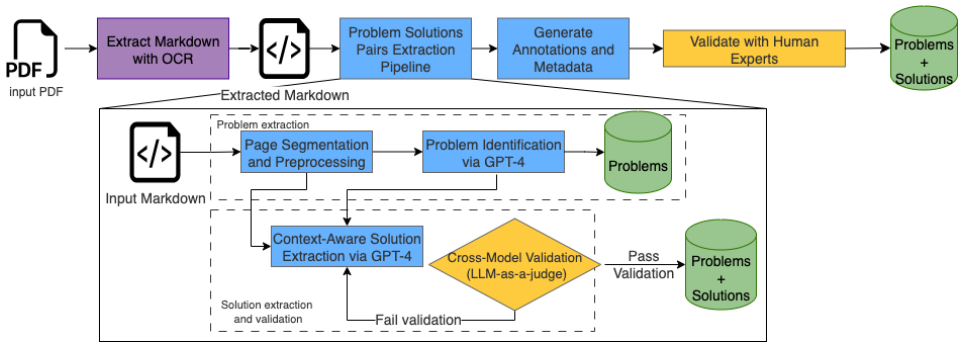


Figure 2: Problem–solution Extraction and Validation Pipeline.

3 Experiments

We evaluate on *MathNet* under two benchmarks: (a) **Math Comprehension** and (b) **Math Retrieval**. For comprehension, we follow the Omni-MATH protocol [6], using `GPT-4o` as the judge (98% agreement with human annotations). For retrieval, we test embeddings from state-of-the-art models. For more details refer to A.7.

Table 2: Experimental results on *MathNet*, expressed as percentages, with the highest score in each setting underlined and the highest scores across all settings bolded. When calculating the overall accuracy, for code generation problems, if any generated code for a problem passes all test cases, the problem is considered correct.

	Algebra	Number Theory	Geometry	Discrete Mathematics	**Macro-average**	**Micro-average**
	LLMs					
Llama-4-Maverick-17B	49%	50%	57%	31%	46.75%	46.29%
DeepSeek-V3	68%	38%	56%	36%	49.50%	45.59%
	LLMs + reasoning					
Gemini 2.5 Pro	74%	69%	61%	**61%**	66.25%	65.99%
Gemini 2.5 Flash	70%	70%	64%	57%	65.25%	64.88%
Claude 4 Sonnet	56%	33%	52%	**61%**	50.50%	52.00%
Claude 4 Opus	60%	38%	**87%**	41%	56.50%	56.57%
GPT-4o	49%	26%	68%	26%	42.25%	40.89%
GPT-5	**76%**	**78%**	75%	60%	**72.25%**	**72.12%**

Math Comprehension Results Table 2 summarizes accuracy across four mathematical domains. Baseline LLMs such as *Llama-4-Maverick-17B* and *DeepSeek-V3* achieve modest macro-averages in the mid-40s, indicating that direct pattern matching and shallow heuristics are insufficient for Olympiad-level problem solving.

Reasoning-augmented models (e.g., *Gemini 2.5 Pro* and *Gemini 2.5 Flash*) substantially improve performance, with macro-averages around 65%. However, their accuracy remains uneven across domains: while Algebra shows steady gains, **Number Theory and Discrete Mathematics remain the hardest categories**, reflecting difficulty with abstract reasoning, non-obvious solution paths, and combinatorial structures.

GPT-5 achieves the strongest and most balanced performance to date, with domain-level accuracies ranging from 60–78% and a macro-average of 72.25%. Yet even at this scale, **Discrete Mathematics and Number Theory continue to expose fundamental weaknesses**, underscoring the need for models that move beyond superficial correlations toward deeper structural and symbolic understanding.

Math Retrieval Results As shown in Table 3, retrieval on *MathNet* remains highly challenging at the top-1 level, with even the strongest models (`Qwen3-embedding-4B` and `Gemini-embedding-001`) achieving only $\sim$5% Recall@1. Performance improves markedly at higher cutoffs, with Recall@10 exceeding 80% in several domains. Among all models, Gemini-embedding-001 provides the most consistent gains, delivering the highest Recall@5 and Recall@10 across domains and the strongest aggregate performance (68.88% and 83.79%, respectively). In contrast, legacy embedding models such as `text-embedding-ada-002` and `text-embedding-3-small` perform substantially worse across all settings.

These results suggest that current general-purpose embedding models fail to capture the deep structural and symbolic relationships that define mathematical equivalence. A critical failure mode is that both LLMs and LMMs often rely on superficial textual overlap (e.g., matching on keywords such as "triangle" or "polynomial") rather than reasoning over the underlying mathematical concepts. The weak top-1 retrieval performance highlights that these models lack a robust internal representation of mathematical knowledge that would support analogical reasoning across problem variants. This gap underscores the need for embeddings explicitly trained to encode mathematical structure, rather than depending on incidental surface-level cues. For more results refer to Appendix A.9.

3.1 Discussion

Results on *MathNet* reveal a clear gap between the problem-solving ability of modern LLMs/LMMs and their understanding of mathematical structure. While models achieve impressive scores on answer-generation benchmarks, our retrieval task shows they lack a generalizable grasp of equivalence and analogy. The limited gains from visual augmentation further suggest that multimodal integration for symbolic tasks remains underdeveloped.

Table 3: Experimental results on *MathNet*, expressed as percentages for Recall@1, Recall@5, and Recall@10. The highest score in each setting is underlined, and the highest overall scores are bolded.

	Algebra			Number Theory			Geometry			Discrete Mathematics			All		
	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10	R@1	R@5	R@10
all-mpnet-base-v2	4.54%	73.06%	84.82%	4.67%	82.54%	91.82%	4.37%	74.76%	86.86%	4.25%	75.38%	86.92%	3.78%	57.7%	74.71%
multi-qa-mpnet-base-dot-v1	4.0%	69.4%	82.01%	3.73%	80.76%	89.88%	3.88%	71.73%	84.4%	3.98%	73.4%	85.05%	3.27%	55.08%	71.93%
cohere-embed-v4.0	2.73%	59.85%	71.97%	2.67%	68.85%	80.98%	2.35%	59.87%	73.09%	2.78%	63.4%	75.74%	2.24%	44.81%	60.33%
qwen3-embedding-4B	5.24%	78.74%	88.58%	4.62%	86.43%	92.71%	5.6%	79.05%	89.35%	5.96%	81.5%	89.96%	**4.96%**	64.95%	81.03%
gemini-embedding-001	5.5%	81.62%	89.31%	4.95%	87.43%	91.99%	5.49%	81.86%	89.73%	5.35%	82.8%	89.96%	4.83%	**68.88%**	**83.79%**
text-embedding-ada-002	2.05%	54.94%	66.49%	2.22%	63.35%	73.08%	2.16%	55.07%	67.74%	2.71%	57.51%	68.38%	1.94%	42.02%	55.58%
text-embedding-3-small	2.1%	47.47%	57.98%	1.89%	54.62%	64.46%	2.1%	47.61%	58.54%	2.84%	50.12%	59.75%	1.98%	35.49%	47.67%
text-embedding-3-large	3.19%	68.18%	78.41%	2.73%	75.25%	82.81%	3.2%	68.18%	78.34%	3.35%	69.52%	79.02%	2.74%	54.23%	68.88%

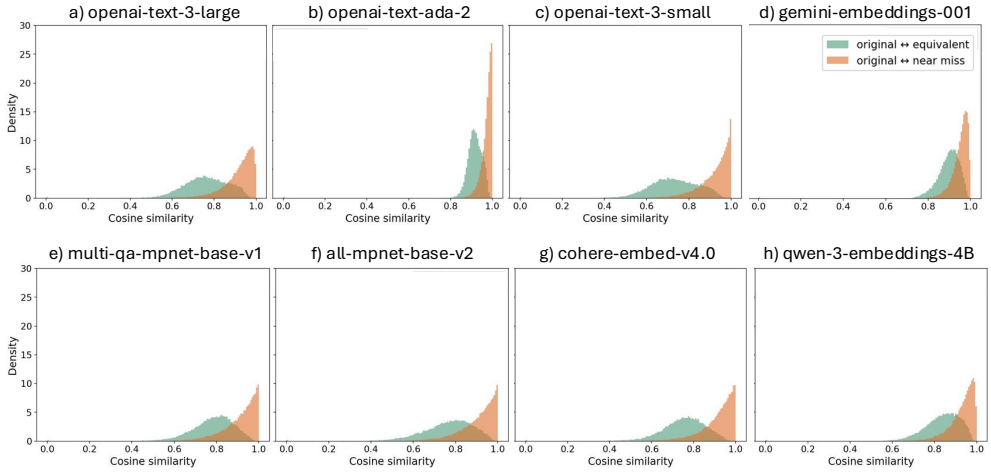


Figure 3: Cosine similarity distributions for equivalent (green) and near-miss/hard negatives (orange) problem pairs across different embedding models. Higher separation between the two distributions indicates a model's ability to distinguish structurally identical problems from those with small but critical alterations.

The strong performance of the formula-aware baseline indicates that structured, non-textual representations are crucial for retrieval. Progress in true mathematical reasoning may require moving beyond next-token prediction toward architectures that explicitly integrate symbolic reasoning.

4 Conclusion

We introduce *MathNet*, a large-scale multilingual, multimodal benchmark (13,026 expert-authored problems with 39K equivalence pairs) that evaluates both solution generation and *math-aware* retrieval. Experiments show frontier models achieve strong comprehension (e.g., GPT-5 at 72.25% macro-average) but retrieval is brittle—top-1 recall remains near 5% even for the best embeddings—indicating reliance on lexical cues over symbolic structure. These results motivate structure- and symbol-aware representations and tighter integration of retrieval with reasoning. We release *MathNet* to catalyze progress toward models that can organize, recall, and apply mathematical knowledge, not just produce correct final answers.

Bibliography

[1] Daman Arora, Gaurav Goyal, Harshit Arora, et al. Have LLMs advanced enough? a challenging problem solving benchmark for large language models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.

[2] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

[3] Debrup Das, Sam O' Nuallain, and Razieh Rahimi. Rader: Reasoning-aware dense retrieval models. *arXiv preprint arXiv:2505.18405*, 2025.

[4] Moran Feldman and Amin Karbasi. Gödel test: Can large language models solve easy conjectures?, 2025.

[5] Thibault Formal et al. Splade v2: Sparse lexical and expansion model for information retrieval. *arXiv:2109.10086*, 2021.

[6] Bofei Gao, Feifan Song, Zhe Yang, Zefan Cai, Yibo Miao, Qingxiu Dong, Lei Li, Chenghao Ma, Liang Chen, Runxin Xu, et al. Omni-MATH: A universal olympiad level mathematic benchmark for large language models. *arXiv preprint arXiv:2410.07985*, 2024.

[7] Google DeepMind. Advanced gemini deep think achieves gold-medal standard at the imo. Blog, 2025.

[8] Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Leng Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, et al. Olympiadbench: A challenging benchmark for promoting agi with olympiad-level bilingual multimodal scientific problems. *arXiv preprint arXiv:2402.14008*, 2024.

[9] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring massive multitask language understanding. In *International Conference on Learning Representations (ICLR)*, 2021.

[10] Dan Hendrycks et al. Measuring mathematical problem solving with the math dataset. *arXiv:2103.03874*, 2021.

[11] Yuzhuo Huang, Xunzhi Bai, Yifan Li, et al. C-eval: A multi-level multi-discipline chinese evaluation suite for foundation models. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks*, 2023.

[12] Zhen Huang, Zengzhi Wang, Shijie Xia, Xuefeng Li, Haoyang Zou, Ruijie Xu, Run-Ze Fan, Lyumanshan Ye, Ethan Chern, Yixin Ye, et al. Olympicarena: Benchmarking multi-discipline cognitive reasoning for superintelligent AI. *arXiv preprint arXiv:2406.12753*, 2024.

[13] Gautier Izacard et al. Unsupervised dense information retrieval with contrastive learning. *arXiv:2112.09118*, 2021.

[14] Vladimir Karpukhin et al. Dense passage retrieval for open-domain question answering. In *EMNLP*, 2020.

[15] Omar Khattab and Matei Zaharia. Colbert: Efficient and effective passage search via contextualized late interaction over bert. In *SIGIR*, 2020.

[16] Haonan Li, Yixuan Zhang, Fajri Koto, Yifei Yang, Hai Zhao, Yeyun Gong, Nan Duan, and Timothy Baldwin. CMMLU: Measuring massive multitask language understanding in chinese. In *Findings of the Association for Computational Linguistics: ACL 2024*, 2024.

[17] Fan Nie, Ken Ziyu Liu, Zihao Wang, Rui Sun, Wei Liu, Weijia Shi, Huaxiu Yao, Linjun Zhang, Andrew Y Ng, James Zou, et al. Uq: Assessing language models on unsolved questions. *arXiv preprint arXiv:2508.17580*, 2025.

[18] OpenAI. Gpt-4 technical report. arXiv:2303.08774v4, 2024.

[19] OpenAI. Openai system claims gold-medal score at the 2025 imo. Blog/press; use cautiously, 2025.

[20] rednote hilab. dots.ocr: Multilingual document layout parsing in a single vision-language model. https://github.com/rednote-hilab/dots.ocr, 2025. Accessed: YYYY-MM-DD.

[21] Ke Wang, Junting Pan, Weikang Shi, Zimu Lu, Houxing Ren, Aojun Zhou, Mingjie Zhan, and Hongsheng Li. Measuring multimodal mathematical reasoning with MATH-Vision dataset. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks*, 2024.

[22] Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, et al. MMMU: A massive multi-discipline multimodal understanding and reasoning benchmark for expert AGI. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[23] Wanjun Zhong, Ruixiang Cui, Sai Liang, et al. AGIEval: A human-centric benchmark for evaluating foundation models. In *Findings of the Association for Computational Linguistics: NAACL 2024*, 2024.

A Dataset

A.1 Competition Sources

Each year, IMO-participating countries publish official problem booklets for national contests and team selection exams. We collected 643 PDF volumes spanning 2006–2025, totaling more than 22,000 pages across 40 countries. All problems and solutions were authored by national teams, ensuring high-quality and consistent material.

A.2 Text Extraction

We standardized all contest booklets into structured Markdown using `dots-ocr` [20], a multilingual OCR and document parsing framework (Appendix Figure 8). This ensured robustness across digitally typeset, scanned, and bilingual documents.

A.3 Problem–Solution Extraction

Aligned problem–solution pairs were identified via a three-stage LLM-based pipeline (Figure 2): 1. **Document ingestion and problem extraction.** Contest booklets were segmented into pages, then processed by **GPT-4.1** to extract problem statements in LaTeX format with provenance metadata. 2. **Solution retrieval.** Solutions were located using a sliding window of four pages after the problem page, with GPT-4.1 extracting matches (prompt shown in Appendix 1). 3. **Semantic verification.** Extracted pairs were checked independently by **GPT-4.1** and **Claude 4 Opus**, which judged alignment and completeness (Appendix 2). Only pairs validated by both were accepted.

This multi-stage approach achieved high precision and recall across diverse contest formats.

Human Validation of Problem–Solution Extraction. To obtain a reliable estimate of extraction quality, we randomly sampled 100 problem–solution pairs from the dataset and conducted a controlled human evaluation. We recruited 20 annotators with academic backgrounds in mathematics, computer science, and engineering, and instructed them to independently assess each pair along two dimensions: (i) the correctness of alignment, i.e., whether the solution corresponds to the intended problem, and (ii) the completeness of coverage, i.e., whether the solution is fully captured. To facilitate annotation, we developed and publicly released a lightweight web-based interface that supports multimodal display of problems, solutions, and provenance metadata (e.g., source document and page number).

LLM-Based Stress Testing with Distractors. To assess dataset robustness and potential leakage, we employed a large language model (GPT-4.1) to generate a set of "distractor" problems. For each problem, we prompted the model to produce five plausible but incorrect statements and then instructed it to identify the correct problem from a mixture of its own distractors and the true related problems from our dataset. The model's low success rate indicates that the annotated problem connections in *MathNet* are non-trivial and cannot be inferred through simple surface-level patterns, thereby reinforcing the quality of our annotations.

Expert Review of Similarity Annotations. As an additional validation step, we asked experts to review a subset of 100 sampled problems with their associated distractors. At least two annotators independently assessed each problem–distractor set, and a senior expert resolved any disagreements through consensus. This procedure confirmed that the similarity annotations capture genuine mathematical structure rather than superficial lexical overlap, providing a complementary layer of assurance beyond the LLM-based evaluation.

A.4 Data Analysis

Figure 5b illustrates the diversity of our dataset across mathematical domains. Notably, Number Theory and Combinatorics account for a large share of the most difficult problems, reflecting their inherent complexity. In addition, the dataset is multilingual, with problems provided in ten different languages (see Appendix Table 5), which makes it particularly well-suited for evaluating cross-lingual reasoning.

A.5 Overview of competitions covered by MathNet

This section lists the national and regional competitions represented in *MathNet*, along with years covered and document sources, to clarify the dataset's institutional breadth.

[redacted]	Years	Competitions
Argentina	2003–2023	Cono Sur MO; Argentine National Olympiad; National Olympiad – First Day; National Olympiad – Second Day; National Olympiad – Level 2 First Day; National Olympiad – Level 2 Second Day; National Olympiad – Level 3 First Day; National Olympiad – Level 3 Second Day; Rioplatense Olympiad – Level A First Day; Iberoamerican MO; Olimpiada de Mayo; Olympiad Solutions – First Day; Olympiad Solutions – Second Day; National Olympiad; Rioplatense MO
Australia	2010–2024	AMOC Senior Contest; APMO; AIMO; Australian MO; EGMO (TST); IMO (TST); MCYA
Austria	2010–2024	Austrian MO – Regional; Austrian MO – Junior Regional; Austrian MO – National (Part 1); Austrian MO – National (Part 2); National Olympiad – Preliminary; National Olympiad – Final; Beginners' Competition; EGMO (TST); IMO (TST)
Balkans	2010–2025	BMO
Baltics	2009–2023	Baltic Way; Baltic Way Shortlist
Belarus	2010–2024	Belarusian MO; IMO (TST)
Brazil	2006–2012	OBM
Bulgaria	2007–2024	Bulgarian MO – Regional; Bulgarian MO – Final; Bulgarian Autumn Competition; Bulgarian Spring Competition; Bulgarian Winter Competition (Rousse, Varna, National); IMO (TST); BMO (TST); Other Bulgarian Competitions; JBMO (TST)
Canada	2010–2017	CMO
China	2007–2025	AMC 10/12; AIME; CMO (China); Chinese MO; China Southeastern MO; CWMO; CGMO; Hua Luogeng Cup; IMO (TST); Soviet Mathematical Competition; Russian Mathematical Competition; Putnam (China ed.)
Croatia	2010–2019	Croatian MO; National Olympiad – City; National Olympiad – County; National Olympiad – Final; MEMO; IMO/MEMO (TST)
Czech Republic	2000–2025	Czech MO – School; Czech MO – District; Czech MO – Regional; Czech MO – Final; Czech–Polish–Slovak Match; Czech–Slovak–Polish Match; Czech–Austrian–Polish–Slovak Match; CAPS Match; Olympiad Corner; IMO/EGMO/MEMO (TST)
Slovakia	2000–2025	Slovak MO – School; Slovak MO – District; Slovak MO – Regional; Slovak MO – Final; Czech–Slovak Match; Czech–Polish–Slovak Match; Czech–Slovak–Polish Match; Czech–Austrian–Polish–Slovak Match; CAPS Match; Olympiad Corner; IMO/EGMO/MEMO (TST)
Poland	2004–2025	Polish MO; Czech–Polish–Slovak Match; Czech–Slovak–Polish Match; Czech–Austrian–Polish–Slovak Match; CAPS Match; Olympiad Corner; IMO/EGMO/MEMO (TST)
Estonia	2010–2025	Estonian MO; Kangaroo; IMO (TST); Other Estonian Open Contests; EGMO (TST)
Greece	2007–2024	Hellenic MO – Archimedes; National Competition – Thales; National Competition – Euclides; BMO; JBMO; Mediterranean Competition; EGMO (TST); IMO (TST); JBMO (TST)
Hong Kong	2014–2017	Hong Kong MO; Hong Kong Team Selection Test; Preliminary Selection – IMO; IMO (TST); APMO; CHKMO
India	2006–2023	INMO; RMO; TSTs (IMO/EGMO/RMM); EGMO (TST); RMM (TST); IMO (TST); USA TST Exchange; ISL/ELMO (training/mock)
Iran	2010–2024	Iranian MO; IMO (TST)
Ireland	2007–2025	Irish MO; IMO (TST)
Japan	2006–2025	JMO; JJMO; IMO/EGMO (TST)
Mongolia	2009–2025	Mongolian MO; Mongolian National MO; IMO (TST); EGMO (TST)
Netherlands	2019–2025	Dutch MO; Junior MO; Kangaroo; Pythagoras Olympiad; BxMO; BxMO/EGMO (TST); IMO (TST)

Continued on next page

Country	Years	Competitions
North Macedonia	2008–2023	Macedonian MO; Macedonian Junior MO; National Olympiad – Regional; National Olympiad – Final; BMO; JBMO; Mediterranean Competition; EGMO (TST); IMO (TST); BMO (TST)
Romania	2010–2025	Romanian MO – District; Romanian MO – Final; RMM; BMO; JBMO; EGMO; IMAR Competition; Stars of Mathematics; Danube Competition; Clock-Tower School Competitions; IMO/BMO/JBMO/EGMO/RMM (TST)
Russia	2009–2025	Russian MO – Regional; Russian MO – Final; Euler Olympiad; All-Russian Olympiad (district, regional, national); IMO/EGMO (TST)
Saudi Arabia	2010–2025	Saudi MO; APMO (TST); EGMO (TST); IMO (TST); BMO (TST); JBMO (TST)
Singapore	2010–2025	SMO (Junior, Senior, Open); SIMOC Camp Quizzes; National Olympiad – Round 2 (all); IMO/EGMO (TST)
Slovenia	2008–2016	Slovenian National MO; International Kangaroo; IMO (TST)
South Africa	2010–2024	SAMO; National Olympiad – Senior; University Training Camps; Talent Search; Monthly Problem Sets; IMO (TST)
South Korea	2004–2024	KMO; National Olympiad; IMO (TST)
Spain	2012–2023	Spanish MO; National Olympiad – First Phase; National Olympiad – Final Phase; Iberoamerican MO; Mediterranean MO; Barcelona Contest; BarcelonaTech Math Contest; Arhimede Contest; IMO (TST)
Taiwan	2012–2024	Taiwan MO; National Olympiad Training Camps (Independent Study, Mock Exams, International Practice); IMO (TST)
Thailand	2007–2017	Thailand MO; TMO; IMO (TST)
Turkey	2008–2024	Turkish MO; Junior Turkish MO; National Olympiad; IMO (TST); JBMO (TST); EGMO (TST); Silk Road Mathematical Competition
UK	2006–2022	BMO (Rounds 1 & 2); BMO; EGMO (TST); IMO (TST); RMM (TST); CGMO (TST); Mathematics Ashes
USA	2001–2025	AMC 10/12; AIME; USAMO; USAJMO; IMO (TST); EGMO (TST); RMM (TST)
Ukraine	2005–2023	Ukrainian National MO; Regional Olympiads; Kyiv City Olympiad; Ukrainian Tournament of Mathematical Battles; Ukrainian Mathematical Competitions; Online Olympiads (Algebra, Combinatorics, Number Theory); Ukrainian Summer School Competitions; EGMO (TST); IMO (TST); RMM (TST); EMC
Vietnam	2001–2024	VMO; Vietnamese National Olympiad; IMO (TST)

A.6 Dataset Statistics

We report summary statistics including per-language and per-domain distributions, subtopic frequencies, and problem/solution length profiles, with additional visualizations. For access to full dataset, refer to http://mathnet.netlify.app/.

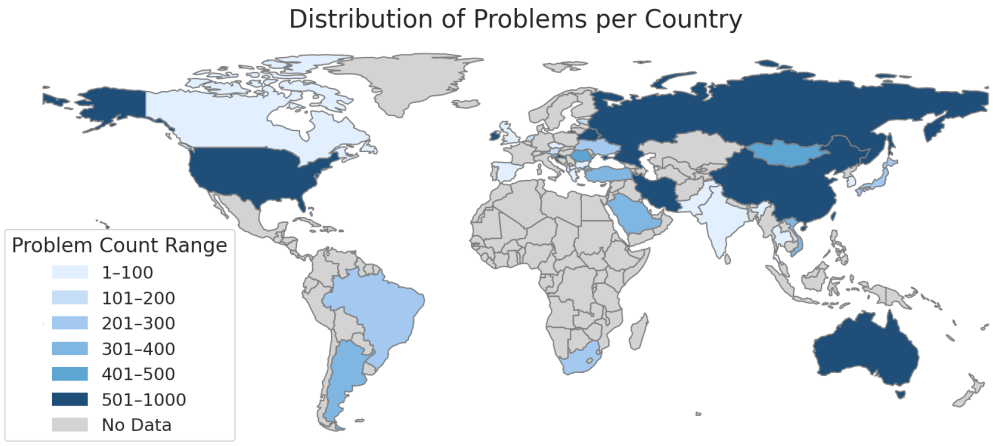
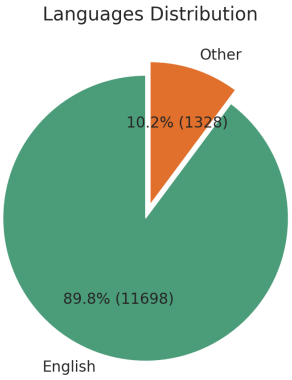
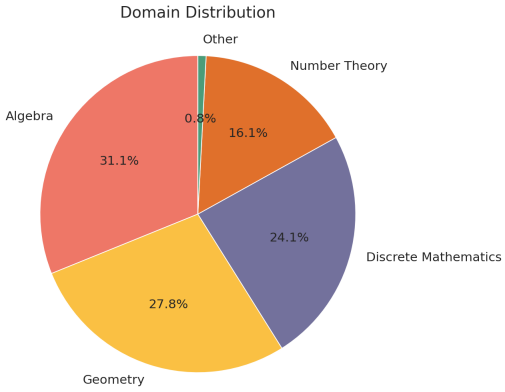


Figure 4: Problems Distribution per Country

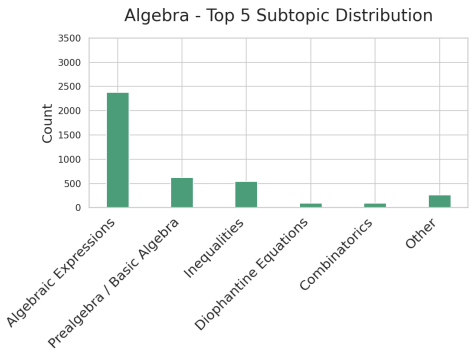


(a) Problems Distribution per Language

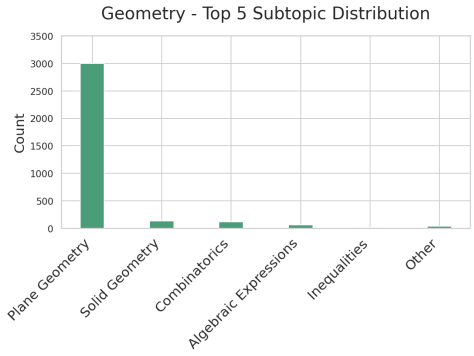


(b) Problems Distribution per Domain

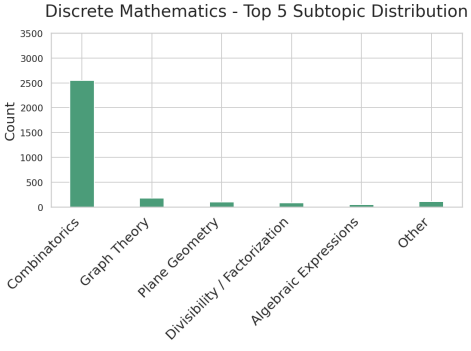
Figure 5: Distribution of problems across languages and domains.



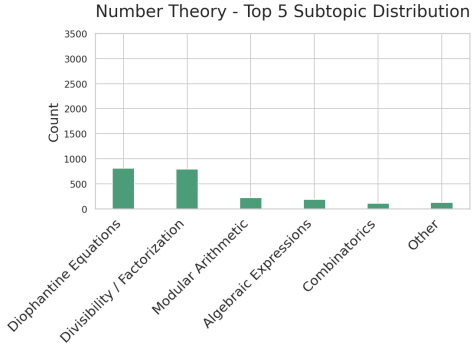
(a) Algebra



(b) Geometry

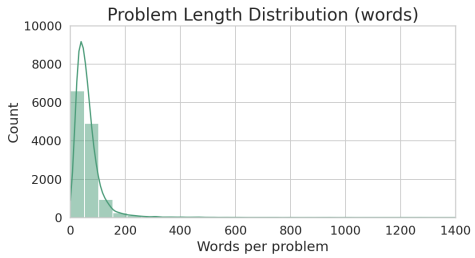


(c) Discrete Mathematics

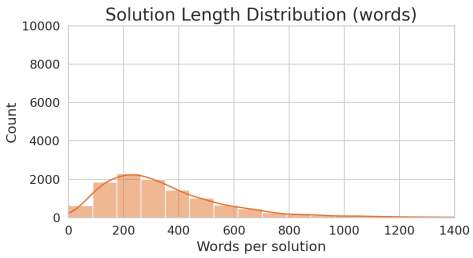


(d) Number Theory

Figure 6: Domain subtopic distribution



(a) Problem Length Distribution (Words)

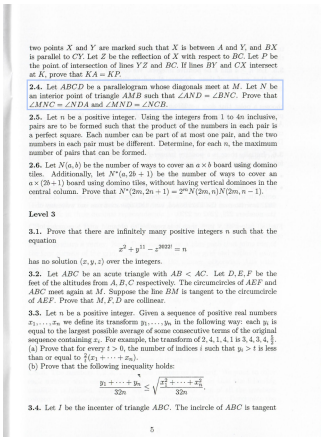


(b) Solution Length Distribution (Words)

Figure 7: Problems vs Solutions (Length Distribution) (words)

Language	English	Spanish	Arabic	Russian	Roman	Bulgarian	Persian	German	Chinese	Ukrainian
Count	11698	242	200	180	60	52	70	23	418	83

Table 5: Problems Distribution per Language



two points X and Y are marked such that X is between A and Y, and BX is parallel to CY. Let Z be the reflection of X with respect to BC. Let P be the point of intersection of lines YZ and BC. If lines BY and CX intersect at K, prove that $KA = KP$.

2.4. Let $ABCD$ be a parallelogram whose diagonals meet at M. Let N be an interior point of triangle AMB such that $\angle AND = \angle BNC$. Prove that $\angle MNC = \angle NDA$ and $\angle MND = \angle NCB$.

2.5. Let n be a positive integer. Using the integers from 1 to $4n$ inclusive, pairs are to be formed such that the product of the numbers in each pair is a perfect square. Each number can be part of at most one pair, and the two numbers in each pair must be different. Determine, for each n, the maximum number of pairs that can be formed.

2.6. Let $N(a, b)$ be the number of ways to cover an $a \times b$ board using domino tiles. Additionally, let $N^*(a, 2b+1)$ be the number of ways to cover an $a \times (2b+1)$ board using domino tiles, without having vertical dominoes in the central column. Prove that $N^*(2m, 2n+1) = 2^m N(2m, n) N(2m, n-1)$.

Level 3

3.1. Prove that there are infinitely many positive integers n such that the equation

$$x^2 + y^{11} - z^{2022} = n$$

has no solution (x, y, z) over the integers.

3.2. Let ABC be an acute triangle with $AB < AC$. Let D, E, F be the feet of the altitudes from A, B, C respectively. The circumcircles of AEF and ABC meet again at M. Suppose the line BM is tangent to the circumcircle of AEF. Prove that M, F, D are collinear.

3.3. Let n be a positive integer. Given a sequence of positive real numbers $x_1, \ldots, x_n$ we define its transform $y_1, \ldots, y_n$ in the following way: each y_i is equal to the largest possible average of some consecutive terms of the original sequence containing x_i. For example, the transform of $3, 4, 1, 4, 1$ is $3, 4, 3, 4, \frac{5}{2}$.
(a) Prove that for every $t > 0$, the number of indices i such that $y_i > t$ is less than or equal to $\frac{n}{t}(x_1 + \cdots + x_n)$.
(b) Prove that the following inequality holds:

$$\frac{y_1 + \cdots + y_n}{32n} \leq \sqrt{\frac{x_1^2 + \cdots + x_n^2}{32n}}.$$

3.4. Let I be the incenter of triangle ABC. The incircle of ABC is tangent

5

(a) PDF Sample

```
## Argentina2022 — Page 3

## Level 2

2.1. Determine the number of permutations $a_1, a_2, ..., a_{2021}$ of the numbers
2, 3, ..., 2022 such that $a_k$ is divisible by $k$, for all $k = 1, 2, ..., 2021$.

2.2. Let $m, n \ge 2$. You want to completely cover an $m \times n$ board without
any gaps or overlaps, using only pieces of the following two types:

![](images/Argentina2022_p3_data_e3b73661a3.png)

Type A

![](images/Argentina2022_p3_data_3fbe98ccdf.png)

Type B

Each type A piece must cover exactly 4 squares on the board, and each type B piece
must cover exactly 5 squares on the board. Rotating the pieces is allowed.
Determine all pairs $(m, n)$ for which this can be done.

2.3. Let ABC be a triangle with $AB < AC$. On the angle bisector of $\angle BAC$,

4

two points $X$ and $Y$ are marked such that $X$ is between $A$ and $Y$, and $BX$
is parallel to $CY$. Let $Z$ be the reflection of $X$ with respect to $BC$. Let
$P$ be the point of intersection of lines $YZ$ and $BC$. If lines $BY$ and $CX$
intersect at $K$, prove that $KA = KP$.
💡
2.4. Let $ABCD$ be a parallelogram whose diagonals meet at $M$. Let $N$ be an
interior point of triangle $AMB$ such that $\angle AND = \angle BNC$. Prove that
$\angle MNC = \angle NDA$ and $\angle MND = \angle NCB$.
```

(b) Markdown Problem Sample

```
## Argentina2022 — Page 10
💡
2.4. Let $N'$ be the symmetric of $N$ relative to $M$. Since the diagonals cut in
half, $AN'CN$ and $DN'BN$ are parallelograms. Hence, $\angle DN'A = \angle BNC$
and $\angle BN'C = \angle DNA$. As $\angle DNA = \angle BNC$, then the
quadrilaterals $DN'NA$ and $BNN'C$ are cyclic. Thus, $\angle MNC = \angle N'NC =
\angle AN'N$ ($AN' \parallel NC$) and $\angle AN'N = \angle NDA$ (cyclic $DN'NA$),
hence $\angle MNC = \angle NDA$. Similarly, $\angle MND = \angle NCB$.

![](images/Argentina2022_p10_data_aac9d871b5.png)

2.5. For each $m \in \mathbb{N}$, let $f(m)$ be the product of the primes that
appear with odd exponent in the prime factorization of $m$. It is easy to see that
given two positive integers $a$ and $b$, the product $ab$ is a perfect square if
and only if $f(a) = f(b)$.

For each $k \in \mathbb{N}$, let $S$ be the set of all $m$ in $\{1, 2, \dots, 4n\}
$ such that $f(m) = k$. Note that if $k$ is not squarefree, then $S$ is empty.
Also, as $f(m) \le m$ for all $m$, we have that each element of $\{1, 2, \dots,
4n\}$ belongs to exactly one of the sets $S_1, S_2, \dots, S_{4n}$.

By the initial observation each of the pairs that we are going to form must be
integrated by two elements of the same $S_j$; furthermore, any way of assembling
the pairs respecting that rule meets the conditions of the statement. Then, if
each set $S_j$ has $a_j$ elements, the maximum number of pairs that can be formed
is $\lfloor \frac{a_1}{2} \rfloor + \lfloor \frac{a_2}{2} \rfloor + \dots +
\lfloor \frac{a_{4n}}{2} \rfloor$.

We claim that $\lfloor \frac{a_j}{2} \rfloor$ is equal to the number of multiples
of 4 in $S_j$. Indeed, this is obvious if $S_j$ is empty; if instead $j$ is
squarefree, the elements of $S_j$ are the numbers of the form $j \cdot s^2$ with
$s$ covering the values from 1 to $a_j$. Since $j$ is squarefree, it has at most
one factor of 2, so $j \cdot s^2$ is a multiple of 4 if and only if $s$ is even.
```

(c) Markdown Solution Sample

```
{'id': 'p_1bef4f22098f',
 'document_name': 'ARG_2022',
 'is_problem': True,
 'cleaned_latex': 'Let $ABCD$ be a parallelogram whose diagonals meet at $M$.
                   Let $N$ be an interior point of triangle $AMB$ such that
                   $\\angle AND = \\angle BNC$. Prove that ...'
 'language': 'en',
 'one_sentence_description': "Given a parallelogram and a point inside one ..."
 'topics': ['Plane Geometry', 'Triangles', 'Quadrilaterals'],
 'parts': [],
 'statement_pages': ['P0'],
 'statement_images': [],
 'source': '/datadrive/dots_ocr/processed_output/ARG_2022',
 'origin_file': 'dots_ocr/out/ARG_2022/json/window_0004.json',
 '_cache_key': 'ARG_2022|p_1bef4f22098f|1bef4f22098ffbe8353e06a6',
 'solution_latex': 'Let $N'$ be the symmetric of $N$ relative to $M$.
                    Since the diagonals cut in half, $AN'CN$ and $DN'BN$ are
                    parallelograms ...'
 'solution_parts': [],
 'source_solution_page_number': [8, 12],
 'has_solution': True
}
```

(d) LLM Output Sample

Figure 8: Sample Input Data

A.7 Evaluation Protocol

Math Comprehension. For the problem-solving task, we adopt the evaluation criteria proposed by Omni-MATH, which compare model-generated solutions against expert-authored ground truth [6]. Following their

protocol, correctness is determined by exact match with the final boxed answer or by equivalence after symbolic simplification. In addition, we compute accuracy at the step level, checking whether intermediate reasoning aligns with expert annotations when available. This allows us to distinguish between models that arrive at the correct final answer by coincidence versus those that demonstrate consistent reasoning ability. We also report performance by subject domain (algebra, geometry, combinatorics, number theory), enabling a fine-grained analysis of model strengths and weaknesses.

Math Retrieval. The primary evaluation metric for our retrieval task is **Recall@k**, which measures whether any of the top-k retrieved problems correspond to a "correct" match from our equivalent versions of each problem. We report Recall@1, Recall@5, and Recall@10. To better understand embedding behavior, we further analyze cosine similarity distributions between equivalent problem pairs, unrelated pairs, and near misses (hard negatives), highlighting cases where models struggle to separate fine-grained distinctions.

A.8 Prompts

We include the core prompts used for extraction, evaluation, and metadata classification. These are the exact versions used in our experiments.

Listing 1: System prompt for solution extraction

````
sys_prompt = """
    You are an expert in extracting mathematical problems and solutions.
    I will provide you with:
        - One math problem
        - Multiple pages
    Extract the solution that matches the problem
    Important instructions:
    - If the problem statement is split into multiple numbered points, extract the
    solution in multiple points
    - Never leave ‘solution_text‘ empty. If no solution can be found, write ‘"Not
    found"‘ as the value.
    - If solution contains imgs make sure to extractt image path such as: ![](images
    /Argentina2022_p5_data_8b4126bbff.png)
    - If solution coontains tables make sure to extract the tables such as: <table><
    thead><tr><th>Team</th><th>T1</th><th>T2</th><th>T3</th><th>T4</th><th>T5</th><
    th>T6</th><th>T7</th><th>T8</th><th>Total</th></tr></thead><tbody><tr><td>T1</
    td><td>-</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</
    td><td>14</td></tr><tr><td>T2</td><td>0</td><td>-</td><td>2</td><td>2</td><td
    >2</td><td>2</td><td>2</td><td>2</td><td>12</td></tr><tr><td>T3</td><td>0</td><
    td>0</td><td>-</td><td>2</td><td>2</td><td>2</td><td>2</td><td>2</td><td>10</td
    ></tr><tr><td>T4</td><td>0</td><td>0</td><td>0</td><td>-</td><td>2</td><td>2</
    td><td>2</td><td>2</td><td>8</td></tr><tr><td>T5</td><td>0</td><td>0</td><td>0</
    td><td>0</td><td>-</td><td>2</td><td>2</td><td>2</td><td>6</td></tr><tr><td>T6</
    td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-</td><td>2</td><td>2</
    td><td>4</td></tr><tr><td>T7</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</
    td><td>0</td><td>-</td><td>2</td><td>2</td></tr><tr><td>T8</td><td>0</td><td>0</
    td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>-</td><td>0</td></tr></
    tbody></table>
    - Follow the JSON schema below precisely:
    ‘‘‘json
    {
        "has_solution": "bool, if solution was found and extracted set to true, else
     false"
        "solution_page_number": "the page number where the solution is found"
        "solution_latex": "extracted solution in latex format"
        "solution_parts": [
            "part_label": "label of the part"
            "part_latex": "extracted part solution in latex format"
        ]
    }
    ‘‘‘
"""
````

Listing 2: System prompt for evaluation

```
sys_prompt_eval = """
    You are an expert in evaluating mathematical problems and solutions.

    I will supply you with a problem and its solution(s), including alternative
    solutions if available.
    Your task is to evaluate based on the following criteria:

    1. **Extraction completeness:** All main parts of the solution must have been
    correctly extracted. Missing or truncated content should be noted.
    2. **Problem-solution match:** Ensure that the solution corresponds correctly to
     the provided problem. If they are mismatched or unrelated, it should be noted.
    3. **Solution completeness:** Check if the reasoning is fully present in the
    extracted solution. Minor implicit steps are acceptable, but missing entire
    parts should be flagged.

    **Important:**
    - Since these problems and solutions are authored by experts, do NOT reject for
    correctness. Focus only on extraction issues or mismatches.
    - Do NOT reject for minor omissions, terseness, formatting, or style.
    - Only reject if there is a **clear, significant extraction issue** or the
    solution does not match the problem.
    - Always provide a clear reason if rejecting, mentioning which criteria are
    affected

    Provide your evaluation strictly in JSON format:

    ```json
    {
      "final_verdict": "accept" or "reject",
      "reason": "A concise explanation for your decision, mentioning which criteria
    failed if rejected"
    }
        ```
    Do not add any extra commentary outside the JSON.
    """
```

Listing 3: System prompt for domain extraction

```
sys_prompt = f"""You are given a math problem and its solution. Your task is to
    analyze it and return structured metadata in JSON format.

# Step 1. Classification
Classify the problem using ONLY the following taxonomy [Domains  Subjects  Topics
    Subtopics]:
    {
        taxonomy
    }

IMPORTANT
* Each level(domain, subject, topic, subtopic) can have multiple entries if the
    problem involves more than one.
* Always use the exact names from the taxonomy.
* If you cannot classify to a certain Domain, classify to Other

# Step 2. Short Description
Write a concise description of the problem in the form:
"Problem Title: Short description"

* The title should be short(e.g., "Pythagorean Triple Problem", "Graph Coloring
    Puzzle").
* The description should briefly summarize the main mathematical task.

# Step 3. Key Ideas of the Solution
Extract the main strategies, theorems, and insights used in the solution as a bullet
    -point list.
```

```
* Focus on mathematical techniques, not surface-level steps.

- Follow the JSON schema below precisely:

# Output Format (JSON)
```json
    {{
        "classification": {{
            "domain": ["..."],
            "subject": ["..."],
            "topic": ["..."],
            "subtopic": ["..."]
        }},
        "description": "Problem Title: Short description",
        "key_ideas": [
            "First main idea",
            "Second main idea",
            "Third main idea"
        ]
    }}
```
"""
```

A.9 Additional Results

A.10 Error Analysis

Our analysis of model failures reveals three primary error types: Symbolic Misinterpretation, Logical Gaps, and Contextual Over-reliance.

Symbolic Misinterpretation: Embedding models frequently misunderstand the meaning of mathematical notation (e.g., confusing a^n with a_n), leading to incorrect transformations or calculations. This is particularly prevalent in Number Theory and Combinatorics.

Logical Gaps: Models often fail to follow a coherent logical chain, skipping crucial steps or introducing invalid assumptions. This manifests in both problem-solving and retrieval, where the model cannot connect a problem to its related counterpart because it misses the key linking insight.

Contextual Over-reliance Embedding models may over-rely on a few keywords, leading them to misclassify a problem based on a superficial match. For example, a problem about a "congruent triangle" might be retrieved for a query about "similar triangles" because of the keyword "triangle," despite the profound mathematical difference.

For retrieval, further illustrate the issue of lexical similarity over math equivalence, Figure 3 shows the distribution of cosine similarities between equivalent and non-equivalent problems. Surprisingly, non-equivalent pairs often exhibit higher similarity scores than equivalent ones. This counterintuitive trend highlights that embeddings frequently capture superficial lexical or symbolic overlap rather than true structural relationships, leading models to mis-rank distinct problems as closer than genuinely equivalent ones. This explains the weak Recall@1 performance observed in Table 3.

A.11 LLMs Usage in the Paper

The authors made use of large language models (LLMs) primarily to support the writing process, including polishing the text for clarity and readability. In addition, LLMs were employed to assist in refining the design of the project website as well as the interface used by annotators.

A.12 Taxonomy of Topics Commonly used in Math Olympiad

We provide the curated taxonomy used for labeling domains, subjects, topics, and subtopics. These labels ground our analyses and enable consistent cross-competition comparisons.

Sub-subtopic	Key Concepts
Geometry	
Plane Geometry	
Triangles	Centroid, incenter, circumcenter, orthocenter, ex-centers, Euler line, nine-point circle; geometric inequalities; trigonometry (metric relations)
Quadrilaterals	Cyclic, inscribed/circumscribed, Complete quadrangle, perpendicular diagonals
Circles	Angels, coaxal, tangents, radical axis, metric relations, Apollonius circle
Concurrency / Collinearity	Theorems of Ceva, Menelaus, Pappus, Desargues
Transformations	Translation, rotation, homothety, spiral similarity, inversion, the method of moving points
Advanced Configurations	Simson line, Miquel, Napoleon / Fermat / Brocard points, symmedians, polar triangles, harmonic/isogonal/isotomic conjugates, barycentric coordinates
Geometric Inequalities	Classical and advanced
Combinatorial Geometry	Helly, Sylvester, convex hulls, Pick theorem, Minkowski theorem, convex figures
Analytic / Coordinate Methods	Complex numbers, Cartesian coordinates, vectors, trigonometric relations
Miscellaneous	Angle/distance chasing, constructions, loci
Solid Geometry	
3D Shapes	Polyhedra, prisms, pyramids, spheres, cylinders, cones
Volume	Cavalieri's principle, Formulae and problem-solving
Surface Area	Formulae and applications
Other 3D problems	Mixed problems, reducing the problem into a plane geometry problem
Differential Geometry	
Curvature	Gaussian, mean
Manifolds	Surfaces, parametric
Geodesics	Shortest paths, great circles
Non-Euclidean Geometry	
Spherical Geometry	Spherical triangles, angles, area
Hyperbolic Geometry	Lines, models, inequalities
Algebra	
Prealgebra / Basic Algebra	
Integers	Sets of integers, Divisibility, primes, the Greatest Common Divisor (GCD), the Least Common Multiplier (LCM)
Fractions	Operations, simplification, comparison
Decimals	Conversion, operations, rounding
Simple Equations	Linear equations, word problems
Other	Number properties, prime factorization, divisors
Algebraic Expressions	
Polynomials	Operations, factorization, Algebraic identities, symmetric functions, Vieta's formula, interpolation formulae, complex numbers, roots of unity, Chebyshev polynomials and other trigonometric polynomials, irreducibility of polynomials, Descartes rule of signs, rootso of polynomials, Intermediate Value Theorem (IVT)

Continued on next page

Sub-subtopic	Key Concepts
Sequences / Series	Recurrences, Charachteristic equations, monotonocity, boundedness, periodicity, convergence and divergence, floors/ceilings, sums/products, telescoping sums, Abel summation
Functional Equations	Substitution, defining a new function, Cauchy's equations, Injectivity/surjectivity, Periodicity, application of Calculus and Mathematical Analysis, iterations
Inequalities	
Functional considerations	Linear/Quadratic solving techniques
Classical inequalities	Cauchy-Schwarz, QM-AM-GM-HM, Power Mean, Jensen's Inequality, smoothing, Muirhead, Chebyshev's inequality, majorization, combinatorial optimization
Discrete Mathematics	
Graph Theory	
Basic concepts	Vertices, edges, path, connected graphs, cycles, Hamiltonian cycle and path, trees
Matchings	Marriage Lemma, Tutte's theorem
Connectivity	Menger, max-flow min-cut
Extremal	Turán
Euler characteristic	$V - E + F$
Combinatorics	
Enumeration	Symmetry, basic counting techniques, recursion, bijection, inclusion-exclusion, double counting
Probability	Expected values, probabilistice methods, partitions, generating functions
Binomial coefficients	Algebraic properties
Pigeonhole principle	Applications
Invariants / Monovariants	Problem-solving
Coloring / Extremal	Graph problems
Induction	Standard and smoothing
Games / Greedy	Strategies, combinatorial games
Logic / Algorithms / Other	
Logic	Propositional/predicate logic, truth tables
Algorithms	Sorting, searching, Dynamic Programming (DP), greedy
Other	Miscellaneous problems, strategy development problems, interdeciplinary problems
Number Theory	
Divisibility / Factorization	
Primes	Properties, sieves, prime numbers tests
GCD	Euclidean algorithm; linear combinations; Bezout's identity
LCM	Computation; relation with GCD
Factorization	Trial, Fermat, Pollard
Modular Arithmetic	
Basic operations $\pmod n$, inverses $\pmod n$	Existence (when $\gcd(a, n) = 1$); computation (extended Euclidean algorithm)
Chinese Remainder Theorem (CRT)	Solving systems of congruences; applications in number theory and cryptography
Fermat / Euler / Wilson	Theorems; proofs; problem-solving applications
Polynomials mod p	Roots, factorization; applications to number theory problems
Residues / Primitive Roots	
Primitive roots	Existence modulo primes; modulo p^n; computation
Quadratic residues	Properties; Legendre symbol; Euler's criterion
Quadratic reciprocity	Law of quadratic reciprocity; applications
Multiplicative order $\pmod n$	Definition; computation; relation with primitive roots and cyclic groups
Diophantine Equations	

Continued on next page

Sub-subtopic	Key Concepts
Factorization Methods	Difference of squares, Sophie Germain identity, special factorizations; Unique Factorization Domains (Gaussian, Eisenstein integers); Norms in algebraic number fields; Vieta jumping
Modular Arithmetic & Congruences	Reductions modulo primes or powers; Quadratic residues, Legendre symbol; Multiplicative order & primitive roots; Hensel lifting; Local–global principles (solvability mod p)
Parametrization of Solutions	Pythagorean triples; Rational parametrization of conics (general quadratics); Higher-degree parametrizations (elliptic curves, quartics)
Inequalities & Size Arguments	Bounding arguments; Infinite descent; Minimal solutions (no smaller solution possible)
Special Equations	Pell's equation: continued fractions, fundamental solution, recurrence; Fermat-type: $x^4 + y^4 = z^2$,
Descent & Structural Methods	Infinite descent; Descent on elliptic curves; Geometry of numbers
Arithmetic Functions	
Euler's totient's function	Properties, applications
Number / Sum of divisors	Computation, properties
Sum of digits	Basic properties
Möbius inversion	Definition, applications
Algebraic Number Theory	
Algebraic numbers	Minimal polynomials, field extensions, solving Diophantine equations