Novelty Assessment Report

Paper: Achieving Expert-Level Agent from Foundation Model via Complexity Curriculum Reinforcement Learning with Synthetic Data
PDF URL: https://openreview.net/pdf?id=1sffPGGQyT
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Large language model (LLM) agents exhibit strong mathematical problem-solving abilities and can even solve International Mathematical Olympiad (IMO) level problems with the assistance of formal proof systems. However, due to weak heuristics for auxiliary constructions, AI for geometry problem solving remains dominated by expert models such as AlphaGeometry 2, which rely heavily on large-scale data synthesis and search for both training and evaluation. In this work, we make the first attempt to build a medalist-level LLM agent for geometry and present InternGeometry. InternGeometry overcomes the heuristic limitations in geometry by iteratively proposing propositions and auxiliary constructions, verifying them with a symbolic engine, and reflecting on the engine's feedback to guide subsequent proposals. A dynamic memory mechanism enables InternGeometry to conduct more than two hundred interactions with the symbolic engine per problem. To further accelerate learning, we introduce Complexity-Boosting Reinforcement Learning (CBRL), which gradually increases the complexity of synthesized problems across training stages. Built on InternThinker-32B, InternGeometry solves 44 of 50 IMO geometry problems (2000–2024), exceeding the average gold medalist score (40.9), using only 13K training examples, just 0.004% of the data used by AlphaGeometry 2, demonstrating the potential of LLM agents on expert-level geometry tasks. InternGeometry can also propose novel auxiliary constructions for IMO problems that do not appear in human solutions. We will release the model, data, and symbolic engine to support future research.

Core Task Landscape

This paper addresses: **Automated Geometry Theorem Proving**

A total of **50 papers** were analyzed and organized into a taxonomy with **22 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Algebraic and Symbolic Methods**
- **Formal Verification and Interactive Proof Systems**
- **Neural and Learning-Based Approaches**
- **Proof Generation and Readability**
- **Theorem Discovery and Generation**
- **Interactive and Dynamic Geometry Systems**
- **Specialized Geometric Domains and Techniques**
- **Surveys, Reviews, and Foundational Studies**

Complete Taxonomy Tree

- Automated Geometry Theorem Proving Survey Taxonomy
- Algebraic and Symbolic Methods
  - Wu's Method and Characteristic Sets (5 papers)
  - [3] Automated theorem proving (Monty Newborn, 2001) View paper
  - [5] Automated geometric theorem proving: Wu's method (Joran Elias, 2004) View paper
  - [29] Wu's method for automated geometry theorem proving and discovering (Shang-Ching Chou, 2000) View paper
  - [33] Mechanical theorem proving in geometries: Basic principles (Wu, 2012) View paper
  - [49] Ritt-Wu's decomposition algorithm and geometry theorem proving (S. Chou, 1990) View paper
  - Gröbner Basis Methods (2 papers)
  - [18] Automated geometry theorem proving using Buchberger's algorithm (B. Kutzler, 1986) View paper
  - [38] On the Application of Buchberger's Algorithm to Automated Geometry Theorem Proving (Bernhard Kutzler, 1986) View paper
  - Vector and Complex Number Methods (3 papers)
  - [9] Self-evident Automated Geometric Theorem Proving Based on Complex Number Identity (Xicheng Peng, 2023) View paper
  - [15] Automated geometry theorem proving by vector calculation (Shang-Ching Chou, 1993) View paper
  - [41] Self-evident automated proving based on point geometry from the perspective of Wu's method identity (Jingzhong Zhang, 2019) View paper
  - General Algebraic Frameworks (4 papers)
  - [1] Automatic geometry theorem proving (Tomas RECIO, 1999) View paper
  - [21] Geometry Theorem Proving (Jacques Fleuriot, 2001) View paper
  - [32] Automatic discovery of theorems in elementary geometry (T. Recio, 1999) View paper
  - [46] A new approach for automatic theorem proving in real geometry (Andreas Dolzmann, 1998) View paper
- Formal Verification and Interactive Proof Systems
  - Lean-Based Formalization (2 papers)
  - [4] Seed-Prover: Deep and Broad Reasoning for Automated Theorem Proving (Chen, 2025) View paper

- [8] LeanGeo: Formalizing Competitional Geometry problems in Lean (Wang Zihan, 2025) View paper
  - Coq-Based Formalization (2 papers)
  - [6] A matroid-based automatic prover and coq proof generator for projective incidence geometry (David Braun, 2024) View paper
  - [13] A machine proof system of point geometry based on Coq (Siran Lei, 2023) View paper
  - Constructive and Direct Proofs (2 papers)
  - [26] Euclid Machines (Wagner Sanz, 2025) View paper
  - [27] A machine-checked direct proof of the Steiner-lehmus theorem (Ariel Kellison, 2022) View paper
  - Autoformalization and Translation (2 papers)
  - [7] Autoformalizing euclidean geometry (Murphy, 2024) View paper
  - [25] Language Models for Verifiable Mathematical Automation Interaction, Integration, and Autoformalization (Qiaochu, 2024) View paper
- Neural and Learning-Based Approaches
  - Reinforcement Learning for Proof Search ★ (2 papers)
  - [0] Achieving Expert-Level Agent from Foundation Model via Complexity Curriculum Reinforcement Learning with Synthetic Data (Anon et al., 2026) View paper
  - [2] Aristotle: IMO-level Automated Theorem Proving (Achim, 2025) View paper
  - Olympiad-Level Problem Solving (1 papers)
  - [47] Solving olympiad geometry without human demonstrations. (Trieu H. Trinh, 2024) View paper
  - Partial Label and Supervised Learning (2 papers)
  - [31] Partial Label Learning for Automated Theorem Proving (Zombori, 2025) View paper
  - [48] Verification as learning geometric concepts (Rahul Sharma, 2013) View paper
- Proof Generation and Readability
  - Traditional Proof Production (3 papers)
  - [17] Automated Geometry Theorem Proving for Human-Readable Proofs (Ke Wang, 2015) View paper
  - [22] Machine proofs in geometry: Automated production of readable proofs for geometry theorems (Shang-Ching Chou, 1994) View paper
  - [28] Automated production of traditional proofs for constructive geometry theorems (S. Chou, 1993) View paper
  - Illustrated and Visual Proofs (1 papers)
  - [11] Automated generation of illustrated proofs in geometry and beyond (Predrag JaniÄiÄ, 2023) View paper
- Theorem Discovery and Generation
  - Automated Theorem Generation (3 papers)
  - [12] Automated Generation of Geometry Proof Problems Based on Point Geometry Identity (Lei Li, 2024) View paper
  - [35] Automatic theorem generation in plane geometry (Rajiv Bagai, 1993) View paper
  - [43] Automated Generation of Triangle Geometry Theorems (Alexander Skutin, 2021) View paper
  - Conjecture Completion and Discovery (2 papers)
  - [19] Automated Completion of Statements and Proofs in Synthetic Geometry: an Approach based on Constraint Solving (Janicic, 2024) View paper
  - [20] Considerations on approaches and metrics in automated theorem generation/finding in geometry (Pedro Quaresma, 2024) View paper
- Interactive and Dynamic Geometry Systems
  - GeoGebra Integration (3 papers)
  - [23] Automated theorem proving in GeoGebra: Current achievements (Francisco Botana, 2015) View paper
  - [30] Dealing with Degeneracies in Automated Theorem Proving in Geometry (Z. KovÃ¡cs, 2021) View paper
  - [34] A mechanical geometer (F. Botana, 2021) View paper
  - Proof Exploration Environments (1 papers)
  - [24] Proof exploration using dynamic geometry systems with integrated automated deduction capabilities (P. Quaresma, 2024) View paper
- Specialized Geometric Domains and Techniques
  - Projective and Non-Euclidean Geometry (1 papers)
  - [10] Automatic proving of geometric theorems (Henry Crapo, 1995) View paper
  - Solid and Higher-Dimensional Geometry (1 papers)
  - [39] Towards automated proving in solid geometry (Danijela SimiÄ, 2025) View paper
  - Geometric Construction Problems (1 papers)
  - [40] Solving geometric construction problems supported by theorem proving (V Marinkovic, 2014) View paper
  - Alternative Algebraic Representations (2 papers)
  - [36] A Deductive Database Approach to Automated Geometry Theorem Proving and Discovering (Shang-Ching Chou, 2000) View paper
  - [44] Automated Theorem Proving Practice with Null Geometric Algebra (Hongbo Li, 2019) View paper
- Surveys, Reviews, and Foundational Studies (6 papers)
  - [14] A Review on Mechanical Proving and Formalization of Mathematical Theorems (Si Chen, 2025) View paper
  - [16] Automated reasoning in geometry theorem proving with Prolog (Helder Coelho, 1986) View paper
  - [37] Automated theorem proving for elementary geometry (Marek Janasz, 2016) View paper
  - [42] Mechanical geometry theorem proving (Shang-Ching Chou, 1988) View paper
  - [45] What can formal systems do for mathematics? A discussion through the lens of proof assistants (Koutsoukou-Argyraki, 2022) View paper
  - [50] Standardizing the specification, verification, and exchange of product geometry: Research, status and trends (Vijay Srinivasan, 2008) View paper

Narrative

Core task: Automated geometry theorem proving. The field encompasses a diverse set of approaches for mechanically establishing the validity of geometric statements. At the highest level, the taxonomy distinguishes between classical algebraic and symbolic methods—such as the Wu method[5] and Gröbner basis techniques[18]—that reduce geometric problems to polynomial algebra, and formal verification frameworks that integrate geometry into interactive proof assistants like Coq[6,13]. Alongside these traditional pillars, neural and learning-based approaches have emerged, leveraging reinforcement learning and neural proof search to navigate large search spaces. Additional branches address proof generation and readability[11,17], theorem discovery and generation[20,32], interactive and dynamic geometry systems[23,24], and specialized domains including solid geometry[39] and olympiad-level problems[47]. Surveys and foundational studies[14] provide historical context and methodological overviews, tying together decades of research from symbolic engines to modern machine learning.

Recent work has intensified the use of reinforcement learning for proof search, exploring how agents can learn effective strategies in complex geometric environments. Expert Agent Curriculum[0] sits squarely within this neural and learning-based branch, focusing on curriculum design and expert guidance to improve RL-driven proof discovery. It contrasts with Aristotle IMO[2], which also targets challenging competition-level geometry but may emphasize different training regimes or neural architectures. Both efforts reflect a broader trend of applying deep learning to domains once dominated by symbolic reasoning, raising questions about the trade-offs between interpretability—where classical methods like Wu's algorithm[5] produce verifiable algebraic certificates—and the flexibility of learned heuristics. As the field matures, a key open question is how to blend symbolic guarantees with neural scalability, ensuring that automated provers remain both powerful and trustworthy across diverse geometric settings.

Related Works in Same Category

The following **1 sibling papers** share the same taxonomy leaf node with the original paper:

1. Aristotle: IMO-level Automated Theorem Proving

Authors: Achim, Tudor, Best, Alex, Tudor Achim, et al. (43 authors total) | **Year/Venue**: 2025 • arXiv.org | **URL**: View paper

Abstract

We introduce Aristotle, an AI system that combines formal verification with informal reasoning, achieving gold-medal-equivalent performance on the 2025 International Mathematical Olympiad problems. Aristotle integrates three main components: a Lean proof search system, an informal reasoning system that generates and formalizes lemmas, and a dedicated geometry solver. Our system demonstrates state-of-the-art performance with favorable scaling properties for automated theorem proving.

Relationship Analysis

Both papers belong to the Reinforcement Learning for Proof Search category, using RL to guide automated geometry theorem proving. They overlap in applying RL-based proof search strategies to IMO-level geometry problems and achieving medalist-level performance. However, InternGeometry focuses on complexity curriculum RL with synthetic data and long-horizon LLM-tool interactions (200+ steps) using only 13K training examples, while Aristotle integrates formal verification with informal reasoning and includes a dedicated geometry solver component as part of a broader multi-component system.

Contributions Analysis

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: InternGeometry: a medalist-level LLM agent for geometry problem solving

Description: The authors introduce InternGeometry, an LLM-based agent that solves IMO-level geometry problems by iteratively proposing propositions and auxiliary constructions, verifying them with a symbolic engine, and reflecting on feedback. A dynamic memory mechanism enables the agent to conduct over 200 interactions per problem.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Towards Geometry Problem Solving in the Large Model Era: A Survey

URL: View paper

Brief Assessment

Geometry Problem Survey[56] is a survey paper that reviews existing work in geometry problem solving but does not present a novel system. It discusses various approaches including LLM-based methods but does not claim to introduce a specific agent system that would refute InternGeometry's novelty.

2. Knowledge crosswords: Geometric reasoning over structured knowledge with large language models

URL: View paper

Brief Assessment

Knowledge Crosswords[58] focuses on multi-blank QA over knowledge graphs with geometric constraints (entity networks), not geometry theorem proving with symbolic verification and auxiliary constructions for IMO problems.

3. GeoSketch: A Neural-Symbolic Approach to Geometric Multimodal Reasoning with Auxiliary Line Construction and Affine Transformation

URL: View paper

Brief Assessment

GeoSketch[51] focuses on multimodal reasoning with auxiliary line construction and affine transformations using a neural-symbolic approach with visual perception modules. InternGeometry emphasizes iterative proposition proving and auxiliary construction with symbolic verification through DDAR engines, representing a distinct architectural approach to geometry problem solving.

4. Geox: Geometric problem solving through unified formalized vision-language pre-training

URL: View paper

Brief Assessment

GeoX[54] focuses on unified formalized vision-language pre-training for geometric problem solving, using a multi-stage pre-training approach with diagram encoders and formal language alignment. It does not employ iterative proposition-construction-verification cycles with symbolic engine feedback or dynamic memory for extended interactions, which are core to InternGeometry's agent-based approach.

5. LeanGeo: Formalizing Competitional Geometry problems in Lean

URL: View paper

Brief Assessment

LeanGeo[8] focuses on formalizing geometry problems in Lean 4 theorem prover with a declarative framework and benchmark, not on building an LLM agent that iteratively proposes constructions and verifies them with symbolic engines for solving IMO-level problems.

6. Enhancing the geometric problem-solving ability of multimodal llms via symbolic-neural integration

URL: View paper

Brief Assessment

Symbolic Neural Integration[52] focuses on data synthesis and training MLLMs for geometry problem solving, not on building an iterative LLM agent with dynamic memory for 200+ interactions with symbolic engines.

7. Machine assisted proof

URL: View paper

Brief Assessment

Machine Assisted Proof[57] is a general survey discussing proof assistants, machine learning, and LLMs across mathematics broadly. It does not present a specific geometry-solving agent system with iterative proposition-construction-verification loops and symbolic engines like InternGeometry.

8. Seed-Prover: Deep and Broad Reasoning for Automated Theorem Proving

URL: View paper

Brief Assessment

Seed Prover[4] focuses on formal theorem proving in Lean with a geometry reasoning engine (Seed-Geometry) as a component, while the original paper presents an LLM agent that iteratively interacts with a symbolic DDAR engine for geometry-specific problem solving with dynamic memory and complexity-boosting RL.

9. GeoFM: Enhancing Geometric Reasoning of MLLMs via Synthetic Data Generation through Formal Language

URL: View paper

Brief Assessment

GeoFM[53] focuses on synthetic data generation for training MLLMs on geometry problems using formal language to explore metric space combinations. It does not present an LLM agent that iteratively proposes propositions and auxiliary constructions with symbolic verification, which is the core novelty of InternGeometry.

10. Towards Reliable Proof Generation with LLMs: A Neuro-Symbolic Approach

URL: View paper

Brief Assessment

Neuro Symbolic Proofs[55] focuses on guiding LLMs with analogous problems and symbolic verification for proof generation, not on building an autonomous agent with iterative proposition-construction cycles and dynamic memory for IMO-level geometry solving.

Contribution 2: Complexity-Boosting Reinforcement Learning (CBRL)

Description: The authors propose CBRL, a multi-stage curriculum reinforcement learning framework that progressively increases the difficulty of synthesized geometry problems during training. This approach accelerates learning by adapting problem complexity to the current model capability.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Formal mathematics statement curriculum learning

URL: View paper

Brief Assessment

Formal Mathematics Curriculum[71] focuses on formal theorem proving with expert iteration on statement curricula, not geometry problem solving. The candidate's curriculum emerges from proof search difficulty on formal statements, while CBRL explicitly synthesizes geometry problems with controllable complexity levels and adapts difficulty based on RL training dynamics.

2. Ghpo: Adaptive guidance for stable and efficient llm reinforcement learning

URL: View paper

Brief Assessment

GHPO[73] focuses on adaptive guidance through prompt refinement to address reward sparsity in RL training, not on progressive difficulty adaptation through curriculum learning for mathematical problem solving as in CBRL.

3. Self-Evolving Curriculum for LLM Reasoning

URL: View paper

Prior Art Analysis

Self Evolving Curriculum[70] demonstrates that similar curriculum reinforcement learning approaches with progressive difficulty adaptation existed prior to the ORIGINAL paper. Both papers propose multi-stage curriculum RL frameworks that adaptively adjust problem complexity during training based on the model's current capability. Self Evolving Curriculum[70] formulates curriculum selection as a non-stationary multi-armed bandit problem and uses absolute advantage as a reward signal to guide difficulty progression, which is conceptually similar to CBRL's approach of using absolute advantage to determine optimal complexity levels (as stated in the ORIGINAL's Theorem 1 and Equation 9).

Evidence

Evidence 1 - **Rationale**: Both papers propose curriculum RL frameworks that adaptively adjust problem difficulty during training. Self Evolving Curriculum[70] uses absolute advantage as a reward signal for curriculum selection, which directly corresponds to CBRL's use of absolute advantage to determine optimal complexity (Equation 9 in ORIGINAL). - **Original**: we introduce complexity-boosting reinforcement learning (cbrl), a multi-stage curriculum rl pipeline (wang et al., 2025b; chen et al., 2025b; zhang et al., 2025b; parashar et al., 2025), to further improve training efficiency. specifically, we build a data synthesis pipeline that can generate geomet... - **Candidate**: we proposeselfevolving curriculum (sec), an automatic curriculum learning method that learns a curriculum policy concurrently with the rl fine-tuning process. our approach formulates curriculum selection as a non-stationary multi-armed bandit problem, treating each problem category (e.g., difficulty...

Evidence 2 - **Rationale**: Both papers use absolute advantage as the core signal for curriculum optimization. The ORIGINAL paper's Theorem 2 states that maximum absolute advantage of 0.5 indicates optimal difficulty, while Self Evolving Curriculum[70] uses absolute advantage to guide curriculum selection, demonstrating the same underlying principle was already established. - **Original**: where the goal of κ optimization is to maximize the average absolute advantage during learning, and a(x, y)is the advantage of outcome reward. we then present properties of cbrl. as shown in wang et al. (2025b) and chen et al. (2025b), maximum absolute advantage has the following properties: theorem... - **Candidate**: specifically, we operationalize the concept of learning outcomes using the gradient norm, noting that, in policy gradient methods, the gradient norm is weighted by the absolute value of the advantage function. leveraging this observation, we define the absolute advantage as the reward for each arm. ...

Evidence 3 - **Rationale**: Both methods iteratively sample training data based on adaptive difficulty/complexity parameters and update these parameters during training. Self Evolving Curriculum[70]'s approach of sampling categories and updating curriculum policy mirrors CBRL's iterative complexity adjustment. - **Original**: in practice, in each cbrl round, we sample data conditioned on complexity κ, perform rl training to the agent, and finally updateκaccording to learning rateα. - **Candidate**: at each rl training step, a batch of problems is generated as follows. first, categories are sampled according to a boltzmann distribution defined by the current values of qt(c): p(c) = eqt(c)/τ pn i=1 eqt(ci)/τ , where τ is the temperature parameter controlling the exploration-exploitation trade-of...

---

4. Efficient reinforcement finetuning via adaptive curriculum learning

URL: View paper

Prior Art Analysis

Adaptive Curriculum Reinforcement[67] demonstrates that adaptive curriculum learning with progressive difficulty adjustment for mathematical problem solving was proposed prior to the original paper's CBRL framework. Both papers present multi-stage curriculum RL frameworks that dynamically adjust problem difficulty based on model performance. The candidate paper's ADAaRFT method (published as arxiv:2504.05520v2 in April 2025, before the original paper's ICLR 2026 submission) explicitly describes adaptive curriculum learning that 'dynamically adjusts the difficulty of training problems based on the model's recent reward signals' and maintains 'an optimal difficulty range' - the same core mechanism as CBRL. The candidate also provides theoretical justification and implements difficulty-based sampling, directly overlapping with the original paper's claimed novelty of complexity-based curriculum RL.

Evidence

Evidence 1 - **Rationale**: Both papers describe the same core mechanism: a curriculum RL framework that dynamically adjusts problem difficulty based on model performance feedback. The candidate's ADAaRFT predates the original submission and explicitly claims this adaptive curriculum approach. - **Original**: we introduce complexity-boosting reinforcement learning (cbrl), a multi-stage curriculum rl pipeline (wang et al., 2025b; chen et al., 2025b; zhang et al., 2025b; parashar et al., 2025), to further improve training efficiency. specifically, we build a data synthesis pipeline that can generate geomet... - **Candidate**: we propose adarft , a reinforcement finetuning method based on adaptive curriculum learning (bengio et al., 2009), which dynamically adjusts training set difficulty to match the model's evolving skill level. the intuition is simple: learning is most effective when tasks are neither too easy nor too ...

Evidence 2 - **Rationale**: The candidate paper's abstract explicitly describes the same adaptive difficulty adjustment mechanism that the original paper claims as novel in CBRL, including the key insight of maintaining optimal difficulty range. - **Original**: to train interngeometry, we first apply cold start training using 7k examples created by formalizing existing geometry problems and constructing trajectory data. after the cold start, we introduce a complexity-boosting reinforcement learning (cbrl) framework, a multi-stage curriculum rl pipeline - **Candidate**: adarft dynamically adjusts the difficulty of training problems based on the model's recent reward signals, ensuring that the model consistently trains on tasks that are challenging but solvable. this adaptive sampling strategy accelerates learning by maintaining an optimal difficulty range, avoiding...

Evidence 3 - **Rationale**: Both papers formalize difficulty adjustment based on reward signals. The candidate provides an explicit update rule for target difficulty based on average reward, implementing the same adaptive curriculum concept that the original claims as novel. - **Original**: κ∗ = arg max κ ex∼x(κ)ey∼gθ |a(x, y)| where the goal of κ optimization is to maximize the average absolute advantage during learning, and a(x, y)is the advantage of outcome reward. - **Candidate**: to adapt the curriculum dynamically, the target difficulty is updated based on the average reward. if the model performs well on the current difficulty level (high reward), the target difficulty increases, making the training problems harder. conversely, if the model performs poorly, the target diff...

Evidence 4 - **Rationale**: Both papers arrive at the same theoretical insight that 0.5 reward (50% success rate) represents optimal difficulty for learning with binary rewards, suggesting the same underlying curriculum learning principle. - **Original**: theorem 2.for binary rewards, the maximum average absolute advantage is 0.5, which indicates that the task is of moderate difficulty for the model-neither too difficult nor too trivial. - **Candidate**: in this light, setting β = 0.5, corresponding to a success rate of roughly 50%, naturally aligns with this goal. this section formalizes that intuition by analyzing the relationship between reward variance and learnability in rft with binary rewards.

---

5. Light-r1: Curriculum sft, dpo and rl for long cot from scratch and beyond

URL: View paper

Brief Assessment

Light R1[69] focuses on curriculum training for long chain-of-thought reasoning in language models, using progressive difficulty in supervised fine-tuning and DPO stages. The original paper's CBRL is specifically designed for geometry problem solving with dynamic complexity adjustment based on DDAR proof steps, which is fundamentally different from Light R1's approach to mathematical reasoning through multi-stage SFT.

---

6. Learning like humans: Advancing llm reasoning capabilities via adaptive difficulty curriculum learning and expert-guided self-reformulation

URL: View paper

Brief Assessment

Adaptive Difficulty Learning[72] focuses on curriculum learning for general mathematical reasoning using dynamic difficulty re-estimation within batches, while CBRL specifically targets geometry problem synthesis with controllable proof-step complexity for expert-level geometric theorem proving.

---

7. SATURN: SAT-based Reinforcement Learning to Unleash Language Model Reasoning

URL: View paper

Brief Assessment

SATURN[74] focuses on SAT-based curriculum learning with difficulty estimation via formula parameters (n,k,l), while CBRL targets geometry problems with proof-step-based complexity. The domains and difficulty metrics differ fundamentally.

---

8. Vl-cogito: Progressive curriculum reinforcement learning for advanced multimodal reasoning

URL: View paper

Brief Assessment

VL Cogito[68] focuses on multimodal reasoning tasks with progressive curriculum RL (PCURL) using online difficulty soft weighting, while CBRL targets geometry problem solving with complexity-conditioned data synthesis based on proof step counts. The domains and technical implementations differ substantially.

---

9. Progressive Mastery: Customized Curriculum Learning with Guided Prompting for Mathematical Reasoning

URL: View paper

Brief Assessment

Progressive Mastery[66] focuses on curriculum learning for mathematical reasoning using model-adaptive difficulty definitions and guided prompting, but does not address geometry problem solving or multi-stage curriculum RL with synthesized geometry problems of controllable complexity.

Contribution 3: InternGeometry-DDAR: an interactive geometric proof engine

Description: The authors develop InternGeometry-DDAR, an enhanced interactive geometric proof engine based on the open-source DDAR system. It includes advanced definition strategies and a rich theorem library whose search space theoretically covers complete solutions for most IMO geometry problems.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. ArgoTriCSâautomated triangle construction solver

URL: View paper

Brief Assessment

ArgoTriCS[61] focuses on automated triangle construction problems using straightedge-and-compass, not interactive geometric proof engines for general IMO geometry problems with theorem proving capabilities.

2. Proof exploration using dynamic geometry systems with integrated automated deduction capabilities

URL: View paper

Brief Assessment

Dynamic Geometry Exploration[24] focuses on combining dynamic geometry systems with automated theorem provers for educational exploration, not on developing an interactive proof engine with advanced definition strategies and theorem libraries for solving IMO-level problems.

3. Evolution of automated deduction and dynamic constructions in geometry

URL: View paper

Brief Assessment

Automated Deduction Evolution[59] discusses the general evolution of automated deduction and dynamic constructions in geometry but does not provide specific technical details about interactive geometric proof engines that would challenge InternGeometry-DDAR's novelty claims regarding its enhanced definition strategies and rich theorem library.

4. Computer-assisted theorem proving in synthetic geometry

URL: View paper

Brief Assessment

Computer Assisted Synthetic[60] focuses on formal verification and axiom systems in synthetic geometry using interactive theorem provers (Coq, Isabelle), not on interactive proof engines for IMO-level problems with automated construction search and DDAR systems.

5. Formalgeo: An extensible formalized framework for olympiad geometric problem solving

URL: View paper

Prior Art Analysis

FormalGeo[62] demonstrates that interactive geometric proof engines with automated theorem proving capabilities existed prior to the original paper's work. The candidate paper presents FGPS (Formal Geometry Problem Solver), which serves as an interactive assistant for verifying problem-solving processes and an automated problem solver. Like InternGeometry-DDAR, FGPS is built upon open-source DDAR systems and includes advanced features such as automatic diagram construction, condition validity checks, and a rich theorem library. Both systems support interactive proof verification and automated geometric reasoning, indicating that the core concept of an interactive geometric proof engine with enhanced DDAR capabilities was already established in FormalGeo[62].

Evidence

Evidence 1 - **Rationale**: Both papers describe interactive geometric proof engines built on DDAR systems. FormalGeo[62]'s FGPS serves as an interactive assistant and automated solver, demonstrating that such interactive proof engines existed before the original paper's InternGeometry-DDAR. - **Original**: we build interngeometry-ddar, an interactive geometric proof engine based on the open-source ddar system newclid (sicca et al., 2024). to support more complex geometric structures, we introduce several advanced definition strategies, such as globally optimizing point placements to satisfy constraint... - **Candidate**: we have crafted the formal geometry problem solver (fgps) in python. it serves as both an interactive assistant for verifying problem-solving processes and an automated problem solver, utilizing various methods such as forward search, backward search and ai-assisted search

Evidence 2 - **Rationale**: FormalGeo[62] demonstrates a rich theorem library (196 theorems) capable of solving IMO-level problems, showing that comprehensive theorem libraries for IMO-level geometry existed prior to the original paper's work. - **Original**: contains a rich theorem library whose search space theoretically covers the complete solution of most of the imo geometry problems. - **Candidate**: based on the gft, we have established theformalgeo, which consists of 88 geometric predicates and 196 theorems. it can represent, validate, and solve imo-level geometry problems.

6. Automated theorem proving in GeoGebra: Current achievements

URL: View paper

Brief Assessment

GeoGebra Proving[23] focuses on visual dynamic presentation of proofs in an educational context, while InternGeometry-DDAR is designed as an interactive proof engine for solving IMO-level geometry problems with advanced definition strategies and theorem libraries.

7. Considerations on approaches and metrics in automated theorem generation/finding in geometry

URL: View paper

Brief Assessment

Theorem Generation Metrics[20] focuses on metrics for evaluating the interestingness of automatically generated theorems, not on building interactive proof engines. The candidate discusses theoretical frameworks for assessing theorem quality rather than developing proof systems with construction search capabilities.

8. Automatic discovery of theorems in elementary geometry

URL: View paper

Brief Assessment

Automatic Discovery Elementary[32] focuses on algorithmic commutative algebra and algebraic geometry approaches to theorem proving, not on interactive proof engines with agent-based interaction paradigms and dynamic memory mechanisms as described in the original paper.

9. Combining pencil/paper proofs and formal proofs, a challenge for Artificial Intelligence and mathematics education

URL: View paper

Brief Assessment

Pencil Paper Formal[63] focuses on comparing pencil/paper proofs with formal proofs in educational contexts using existing systems like Coq and newclid. It does not present a novel interactive geometric proof engine with advanced definition strategies and rich theorem libraries as InternGeometry-DDAR does.

---

10. Geoint-R1: Formalizing Multimodal Geometric Reasoning with Dynamic Auxiliary Constructions

URL: View paper

Brief Assessment

GeoInt R1[64] focuses on multimodal geometric reasoning with visual diagram processing and Lean4 formalization, while the original paper develops an interactive DDAR-based proof engine for symbolic geometric theorem proving without multimodal components.

Appendix: Text Similarity Detection

Textual similarity detection checked 31 papers and found 2 similarity segment(s) across 1 paper(s).

The following **1 paper(s)** were detected to have high textual similarity with the original paper. These may represent different versions of the same work, duplicate submissions, or papers with substantial textual overlap. Readers are advised to verify these relationships independently.

1. Self-Evolving Curriculum for LLM Reasoning

Detected in: Contribution: contribution_2

⚠ **Note**: This paper shows substantial textual similarity with the original paper. It may be a different version, a duplicate submission, or contain significant overlapping content. Please review carefully to determine the nature of the relationship.

References

- [0] Achieving Expert-Level Agent from Foundation Model via Complexity Curriculum Reinforcement Learning with Synthetic Data View paper
- [1] Automatic geometry theorem proving View paper
- [2] Aristotle: IMO-level Automated Theorem Proving View paper
- [3] Automated theorem proving View paper
- [4] Seed-Prover: Deep and Broad Reasoning for Automated Theorem Proving View paper
- [5] Automated geometric theorem proving: Wu's method View paper
- [6] A matroid-based automatic prover and coq proof generator for projective incidence geometry View paper
- [7] Autoformalizing euclidean geometry View paper
- [8] LeanGeo: Formalizing Competitional Geometry problems in Lean View paper
- [9] Self-evident Automated Geometric Theorem Proving Based on Complex Number Identity View paper
- [10] Automatic proving of geometric theorems View paper
- [11] Automated generation of illustrated proofs in geometry and beyond View paper
- [12] Automated Generation of Geometry Proof Problems Based on Point Geometry Identity View paper
- [13] A machine proof system of point geometry based on Coq View paper
- [14] A Review on Mechanical Proving and Formalization of Mathematical Theorems View paper
- [15] Automated geometry theorem proving by vector calculation View paper
- [16] Automated reasoning in geometry theorem proving with Prolog View paper
- [17] Automated Geometry Theorem Proving for Human-Readable Proofs View paper
- [18] Automated geometry theorem proving using Buchberger's algorithm View paper
- [19] Automated Completion of Statements and Proofs in Synthetic Geometry: an Approach based on Constraint Solving View paper
- [20] Considerations on approaches and metrics in automated theorem generation/finding in geometry View paper
- [21] Geometry Theorem Proving View paper
- [22] Machine proofs in geometry: Automated production of readable proofs for geometry theorems View paper
- [23] Automated theorem proving in GeoGebra: Current achievements View paper
- [24] Proof exploration using dynamic geometry systems with integrated automated deduction capabilities View paper
- [25] Language Models for Verifiable Mathematical Automation Interaction, Integration, and Autoformalization View paper
- [26] Euclid Machines View paper
- [27] A machine-checked direct proof of the Steiner-lehmus theorem View paper
- [28] Automated production of traditional proofs for constructive geometry theorems View paper
- [29] Wu's method for automated geometry theorem proving and discovering View paper
- [30] Dealing with Degeneracies in Automated Theorem Proving in Geometry View paper
- [31] Partial Label Learning for Automated Theorem Proving View paper
- [32] Automatic discovery of theorems in elementary geometry View paper
- [33] Mechanical theorem proving in geometries: Basic principles View paper
- [34] A mechanical geometer View paper
- [35] Automatic theorem generation in plane geometry View paper
- [36] A Deductive Database Approach to Automated Geometry Theorem Proving and Discovering View paper
- [37] Automated theorem proving for elementary geometry View paper
- [38] On the Application of Buchberger's Algorithm to Automated Geometry Theorem Proving View paper
- [39] Towards automated proving in solid geometry View paper
- [40] Solving geometric construction problems supported by theorem proving View paper
- [41] Self-evident automated proving based on point geometry from the perspective of Wu's method identity View paper
- [42] Mechanical geometry theorem proving View paper
- [43] Automated Generation of Triangle Geometry Theorems View paper
- [44] Automated Theorem Proving Practice with Null Geometric Algebra View paper
- [45] What can formal systems do for mathematics? A discussion through the lens of proof assistants View paper

- [46] A new approach for automatic theorem proving in real geometry View paper
- [47] Solving olympiad geometry without human demonstrations. View paper
- [48] Verification as learning geometric concepts View paper
- [49] Ritt-Wu's decomposition algorithm and geometry theorem proving View paper
- [50] Standardizing the specification, verification, and exchange of product geometry: Research, status and trends View paper
- [51] GeoSketch: A Neural-Symbolic Approach to Geometric Multimodal Reasoning with Auxiliary Line Construction and Affine Transformation View paper
- [52] Enhancing the geometric problem-solving ability of multimodal llms via symbolic-neural integration View paper
- [53] GeoFM: Enhancing Geometric Reasoning of MLLMs via Synthetic Data Generation through Formal Language View paper
- [54] Geox: Geometric problem solving through unified formalized vision-language pre-training View paper
- [55] Towards Reliable Proof Generation with LLMs: A Neuro-Symbolic Approach View paper
- [56] Towards Geometry Problem Solving in the Large Model Era: A Survey View paper
- [57] Machine assisted proof View paper
- [58] Knowledge crosswords: Geometric reasoning over structured knowledge with large language models View paper
- [59] Evolution of automated deduction and dynamic constructions in geometry View paper
- [60] Computer-assisted theorem proving in synthetic geometry View paper
- [61] ArgoTriCSâ□□automated triangle construction solver View paper
- [62] Formalgeo: An extensible formalized framework for olympiad geometric problem solving View paper
- [63] Combining pencil/paper proofs and formal proofs, a challenge for Artificial Intelligence and mathematics education View paper
- [64] Geoint-R1: Formalizing Multimodal Geometric Reasoning with Dynamic Auxiliary Constructions View paper
- [65] Curriculum Reinforcement Learning from Easy to Hard Tasks Improves LLM Reasoning View paper
- [66] Progressive Mastery: Customized Curriculum Learning with Guided Prompting for Mathematical Reasoning View paper
- [67] Efficient reinforcement finetuning via adaptive curriculum learning View paper
- [68] Vl-cogito: Progressive curriculum reinforcement learning for advanced multimodal reasoning View paper
- [69] Light-r1: Curriculum sft, dpo and rl for long cot from scratch and beyond View paper
- [70] Self-Evolving Curriculum for LLM Reasoning View paper
- [71] Formal mathematics statement curriculum learning View paper
- [72] Learning like humans: Advancing llm reasoning capabilities via adaptive difficulty curriculum learning and expert-guided self-reformulation View paper
- [73] Ghpo: Adaptive guidance for stable and efficient llm reinforcement learning View paper
- [74] SATURN: SAT-based Reinforcement Learning to Unleash Language Model Reasoning View paper