Technical Report

Solar Open

102B-Parameter Bilingual Mixture-of-Experts Language Model for Underserved Languages

Upstage Solar Team

Parameters
102B Total

Training Data
20T Tokens

Release
Jan 5, 2026

Overview

Contents

Introduction: The Open Model Ecosystem Gap

Current State: Incomplete Democratization

The open model ecosystem is pivotal for democratizing access to Large Language Models, fostering transparency and community-driven innovation. However, this democratization is far from complete, failing most of the world's languages.

Leading open models reflect this asymmetry: Qwen, DeepSeek, and Kimi prioritize English and Chinese; OLMo focuses solely on English. For other languages, neither large-scale datasets nor frontier models exist in comparable quantity or quality.

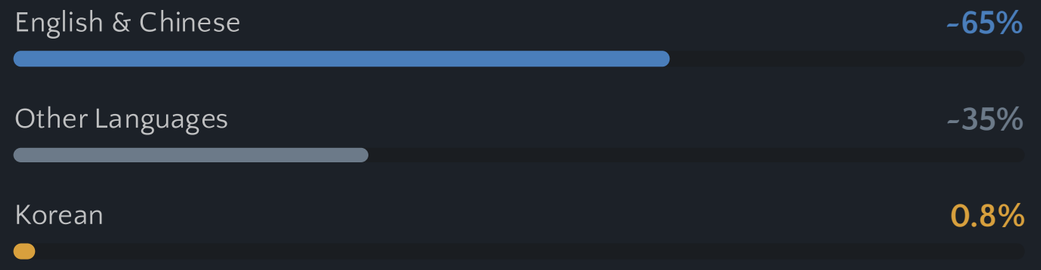
The Korean Context: Data Scarcity

Without language-specific considerations, models suffer degraded performance. Korean exemplifies this challenge, occupying merely 0.8% of indexed web content and ranking 17th in FineWeb 2 by byte count. Language is intrinsically linked to its speakers and cultural context, fundamentally reshaping knowledge and task definitions.

Solar Open: A Systematic Solution

Solar Open is Upstage's flagship open-weight LLM trained on 20 trillion tokens. Built on MoE architecture with 102B total parameters and 12B active parameters per token, the model addresses data scarcity and reasoning challenges through a systematic methodology.

Web Content Distribution



Key Challenges

- ⚠ Suboptimal tokenization with byte-level fallback inflating sequence lengths
- ⚠ Cultural blind spots impacting model performance in downstream tasks
- ⚠ Lack of domain-specific datasets for specialized knowledge areas

Solar Open Highlights

102B	**12B**
Total Parameters	Active per Token
20T	**131k**
Training Tokens	Context Length

02 · Methodology

Three Interconnected Challenges

Unified by a curriculum-coordinated approach to bilingual data and reasoning development

A Synthetic Data Generation

Korean exemplifies data scarcity—occupying only 0.8% of ind exed web and ranking 17th in FineWeb 2, lacking both quantity and quality.

4.5T Tokens Generated

High-quality, domain-specific synthetic data through divers e augmentation, filtering, and transformation pipelines

Mid-Training: Queries + diverse reasoning trajectories

SFT: Successful problem-solving trajectories

RL: Queries for compositional reasoning

B Bilingual Curriculum

Pre-training a 102B-parameter model requires sophisticated d ata curriculum, and bilingual optimization introduces compoun ding complexity.

20T Token Coordination

Jointly optimizing composition, quality thresholds, and dom ain coverage

Balance: Korean and English across stages

Quality: Language-aware thresholds

Domain: Coverage in both languages

C Scalable RL Framework

Traditional online RL tightly couples data generation, reward c omputation, and training, limiting scalability for diverse objecti ves.

SnapPO Framework

Cyclic off-policy framework decoupling three stages

Linear Scaling: Direct throughput increase

Flexible: Multi-domain composition

Independent: Reward computation per domain

Interconnections: A Unified Framework

These challenges are deeply interconnected. **Data scarcity drives aggressive synthetic generation**, which the progressive curriculum coordinates across stages—balancing quality, domains, and langua ges while preparing for RL through reasoning trajectory synthesis. This RL-oriented data strategy feeds directly into **SnapPO**, whose decoupled architecture makes multi-domain, multi-objective traini ng tractable at scale. The result is a cohesive methodology where **data generation, curriculum design, and RL framework mutually reinforce** to overcome challenges facing underserved languages.

4.5T	20T	3
Synthetic Tokens Generated	Total Tokens Coordinated	Decoupled RL Stages

Solar Open Model Architecture

Sparse Mixture-of-Experts Transformer prioritizing efficiency and bilingual capability

Core Specifications

Total Parameters	102.6B
Active per Token	12.0B
Context Length	131,072
Vocabulary Size	196,608
Layers	48
Hidden Size	4,096

MoE Configuration

Total Experts	129
Routed Experts	128
Shared Experts	1
Experts per Token	8
Activation	SiLU
Positional Embeddings	RoPE

Design Philosophy

Design prioritizes efficiency within resource constraints of **480 B200 GPUs** and **three-month timeline**. Approximately 100B total parameters and 10B active parameters represents feasible and effective configuration.

Future Scaling: Anticipate up-scaling to ~200B model based on Depth-Up Scaling

Training Stability

Comparative experiments between 10B-A1B MoE prototype and 3B dense baseline across 50B-100B tokens determine architectural decisions.

Block-masked attention adopted

Systematic variation of **hidden dimension, MoE FFN, layer count**

Final config: **102B total, 12B active**

Architectural Design Process

Benchmarked against contemporary MoE architectures (gpt-oss-120b, Qwen3-235B-A22B, GLM-4.5, DeepSeek-V3). Expert configuration validated through ablation studies on 10B-A1B prototype using MMLU and HellaSwag.

Shared Expert Impact

Incorporating a single shared expert improves performance while maintaining equivalent throughput. This assists when routing paths are uncertain.

Load Balancing

Expert bias (1e-3) combined with sequence-wise load balancing loss (1e-4) provides stability without dense layers.

Key Metrics

Active Ratio	11.7%
Expert Utilization	6.2%

Distinctive Design Choices

SiLU Activation: Only component universally shared across reference architectures

No Dense Layers: Complete absence prioritizing simplicity

Intermediate Size: 10,240 optimized for capacity allocation

MoE Dimension: 1,280 for efficient allocation strategy

Solar Open Tokenizer Design

Custom BPE tokenizer optimized for Korean-English bilingual modeling

Design Overview

Custom-built byte-level BPE tokenizer with **196,608 vocabulary** trained on large corpus oversampling Korean and target domains.

Training Corpus Composition

English	40%
Korean	22%
Code	12%
Math	10%
Multilingual	8%
Domain Specific	4%

Pre-Tokenization Rules

Regex-based pre-tokenization rules enhance performance in reasoning and code generation:

Digit Splitting

All digits treated as individual tokens (pattern `\p{N}`). Prevents number fragmentation, improving arithmetic and scientific formula parsing.

Whitespace Preservation

Preserves whitespace patterns critical for programming languages relying on indentation (e.g., Python), ensuring high fidelity in code generation.

Chat Template Design

Structured message protocol balancing compatibility with modern API patterns while optimizing training stability and reasoning controllability.

Role Types

Supports four role types: **system, user, assistant, tool** with native parallel tool calling for agentic workflows.

Reasoning Separation

Explicit separation using `<|think|>` token facilitates precise reward modeling and efficient reasoning path management.

Compression Rate Benchmarks

Preliminary evaluations show competitive compression rates measured by Bytes per Token. Korean group demonstrates importance of oversampling target languages.

Training-Time Efficiency

Solar Open, A.X-K1, and K-EXAONE achieve best performance on Korean, closely followed by KORMo. All Korean models outperform global models.

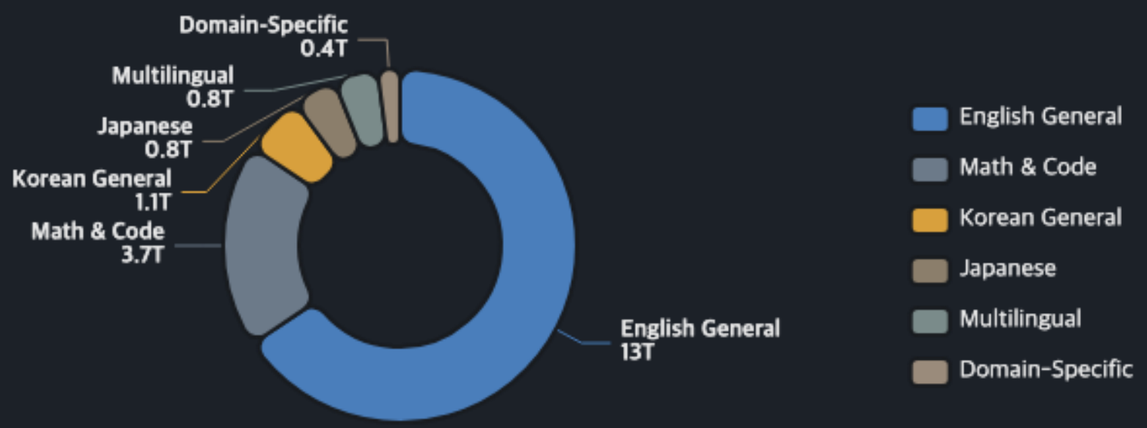
Inference-Time Efficiency

Korean non-reasoning: **4.69 Bytes/Token** (+36% vs gpt-oss, +47% vs DeepSeek V3). Korean reasoning: **4.83 Bytes/Token** (+34% vs gpt-oss).

Pre-Training: Data Construction Strategy

19.7 trillion tokens balancing scale, quality, and Korean representation

Final Corpus Composition



Multi-Pronged Approach

License-Compatible Aggregation

All available openly licensed datasets establish broad foundation

Domain-Specific Curation

Manually curated 0.4T tokens in finance, legal, and medical for enhanced expertise

Key Innovations

Custom PDF Parsing

Extracted 0.4T tokens preserving formatting and semantic structure critical for technical content

Synthetic Data Generation

4.5T tokens using Solar Pro 2 and permissive open-source models ensuring diversity and compliance

Scale Achievement

19.7T

Total tokens successfully prepared and curated for pre-training stage

Korean Representation

5.6%

Korean content including general (1.1T) and domain-specific portions

Synthetic Innovation

22.8%

Synthetic data ratio reaching 64% in later curriculum phases

Progressive Curriculum Learning Strategy

Low-to-high quality progression with increasing synthetic data ratios

Three-Phase Curriculum Overview

Phase 1: Foundation

Tokens: **10.8T**

Synthetic: **10%**

Korean: **3.3%**

Diverse, noisy corpus with basic general quality filtering

Phase 2.A: Refinement I

Tokens: **5.3T**

Synthetic: **32%**

Korean: **3.2%**

Level 1 threshold, three-method filtering framework

Phase 2.B: Refinement II

Tokens: **2.7T**

Synthetic: **36%**

Korean: **16.4%**

Level 2 threshold, top ~50% educational content

Phase 2.C: Refinement III

Tokens: **1.2T**

Synthetic: **64%**

Korean: **13.6%**

Level 3 threshold, top 35% educational content

Phase 3: Specialization

Final **1.5T tokens** of highly curated data targeting specific capabilities:

Korean Cultural Knowledge

0.5T manually curated Korean content for cultural and historical knowledge

Advanced Mathematics & Code

Structured repositories for advanced reasoning and long-form generation (0.5T repository-level code)

Multi-Layer Quality Filtering

Three-method framework operating dynamically during data loading:

1. General Quality Filtering

Lightweight classifier removing noisy text (garbled encoding, repetition, low coherence)

2. Educational Quality Scoring

Regression model scoring educational suitability 0–5, retaining top 35–50% in later phases

3. Embedding-Based Topic Filtering

Text embeddings cluster corpus, selectively sampling aligned domains (science, technical, reasoning)

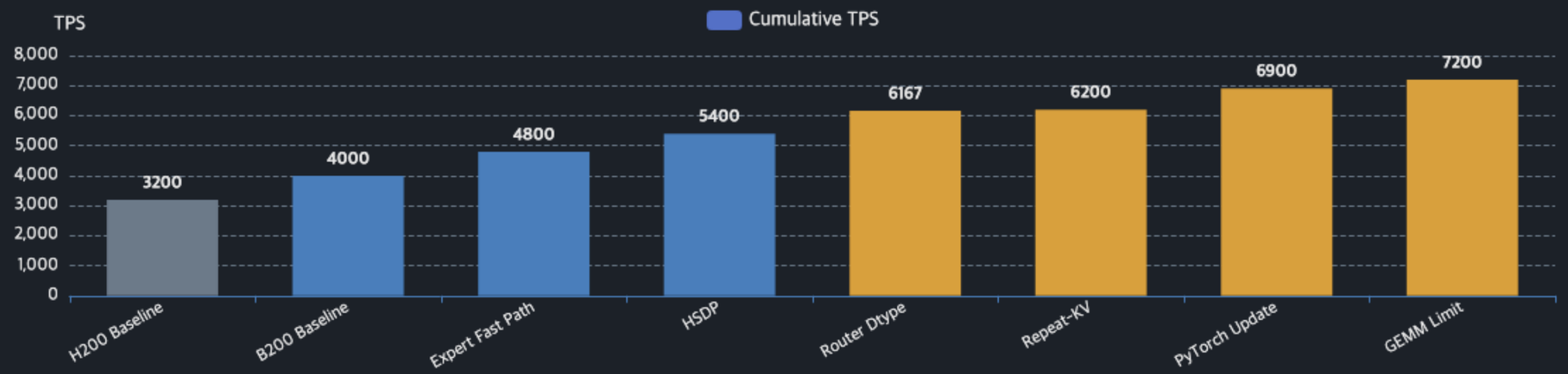
10%	64%	35%	32K
Initial Synthetic Ratio	Peak Synthetic Ratio	Top Content Retained	Final Context Window

04 · Pre-Training

Engineering Optimization Achievements

Systematic optimization: 3,200 TPS → 7,200 TPS (80% improvement)

Throughput Progression Timeline



Framework Selection

DeepSpeed: Stable but slow

Megatron-LM: High potential, complex

TorchTune: 30-100% speedup but for fine-tuning

TorchTitan: +50% TPS improvement

Key Insights

Larger batch sizes enabled by full activation checkpointing outweigh selective checkpointing savings.

Multi-Node Scaling

Standard FSDP2 performance degrades from 16 to 60 nodes (5,500 → 4,267 TPS).

HSDP Solution

Hybrid Sharding divides global pool into smaller sharding groups (10 nodes). Runs FSDP within groups, synchronizes across 6 replicas. **+26.5% throughput**, reaching 5,400 TPS at 60 nodes.

MoE Optimizations

Router Dtype Restoration

Cast back after sigmoid: **+13.7% speedup**

Load Balancing

Histogram-based computation: **-20% routing time**

Expert Parallel Fast Path

Bypass padding when EP disabled: **+14.5% TPS**

Hardware Adaptation

Early B200 deployment challenges:

- Triton CUDA 13.0 support
- ScaledDotProductAttention backend

Data Loading Optimization

Initial bottleneck: 8+ hours initialization.

Arrow File Sharding

Final Achievement

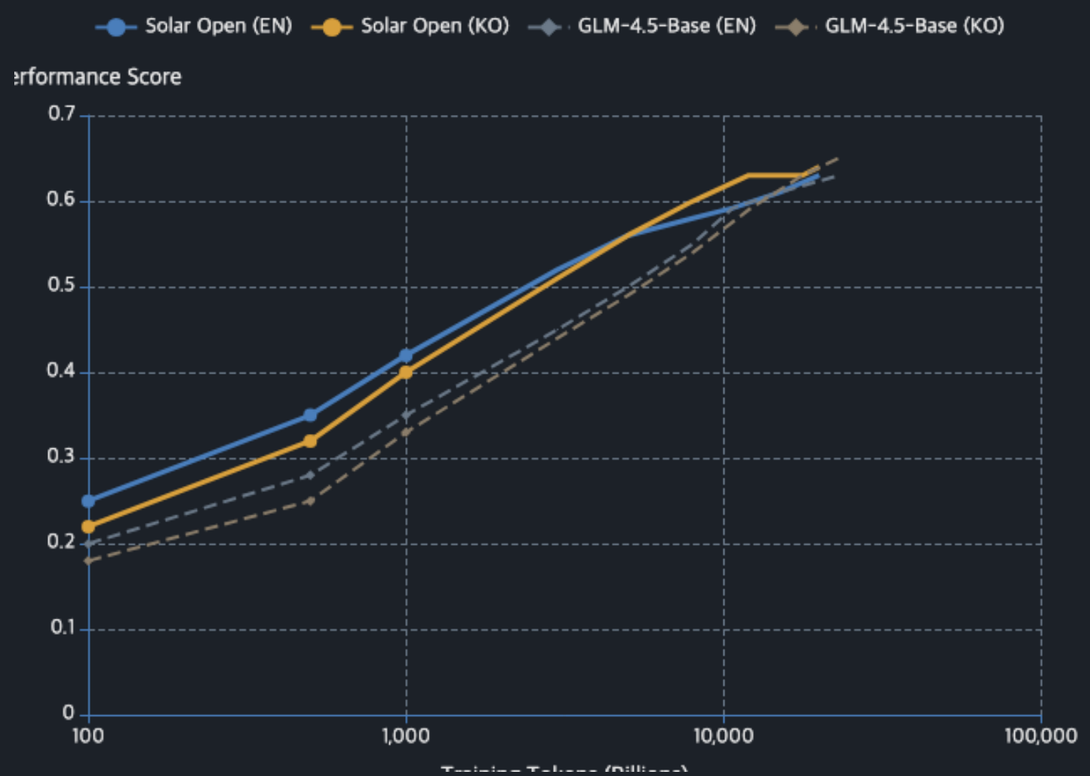
7,200

TPS Final Throughput

Pre-Training Results: Training Efficiency Validation

Comparable performance at significantly reduced token budgets

Training Trajectory Comparison



Efficiency Validation

Preliminary evaluations track performance on standard benchmarks throughout pre-training, comparing against GLM-4.5-Base (106B) trained on 23T tokens.

English Benchmarks

Solar Open achieves GLM-4.5-Base performance at:

At Token Budget **10.7T**

Only **48%** of GLM's total training budget

Korean Benchmarks

Solar Open achieves GLM-4.5-Base performance at:

At Token Budget **17.8T**

Only **77%** of GLM's total training budget

Key Enablers

Progressive Refinement: Synthetic ratio increases from 10% to 64% while raising quality filtering thresholds

Bilingual Curriculum: Carefully designed data curriculum and synthesis strategies

Quality Optimization: Aggressive filtering ensuring exposure to high-quality patterns

Evaluation Benchmarks

Employs MMLU, MMLU-Pro, and HellaSwag measuring performance on both English and Korean versions:

MMLU
Massive Multitask

MMLU-Pro
Robust & Challenging

Mid-Training: RL-Oriented Reasoning Enhancement

1.15T token intermediate stage connecting pre-training and post-training

Stage Purpose

Following pre-training, mid-training enhances reasoning capabilities through **RL-oriented data synthesis** while preventing catastrophic forgetting. Shares same engineering infrastructure as pre-training (next-token prediction).

Key Objectives

- Bridge broad knowledge acquisition and capability specialization
- Provide diverse logical step sequences for atomic reasoning
- Generate flexible reasoning patterns for later RL recombination

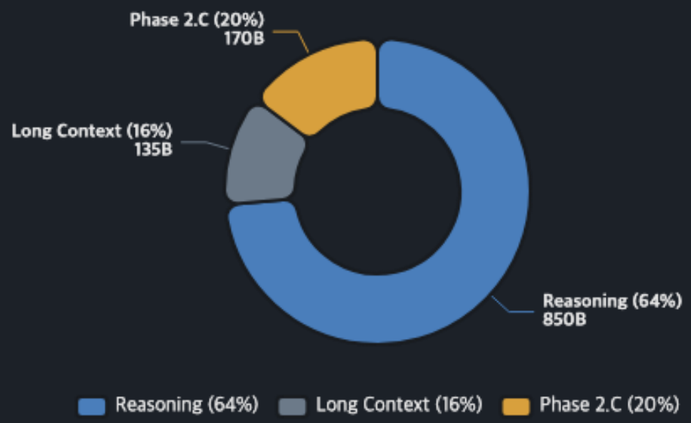
RL-Oriented Reasoning Trajectory Synthesis

For challenging queries (identified by difficulty estimator), generate multiple diverse reasoning trajectories from open-source models. Structure as **coherent documents** (not chat) containing query followed by 2-5 solution approaches.

Generation Details

Total Generated	% of Mid-Training
850B	64%

Data Composition (1,150B Total)



Training Strategy Details

Language Balance

Korean and English each:

~50%

Synthetic Ratio

Synthetic data comprises:

80%

Reasoning Category Breakdown

General Reasoning	37%
Code	16%
Mathematics	47%

06 · Supervised Fine-Tuning

SFT: Difficulty-Aware Data Curation

Strategic data allocation through automated difficulty estimation

Difficulty Estimator Design

Not all queries contribute equally to capability development. Training on queries that are too easy provides diminishing returns, while overly difficult queries provide insufficient learning signal.

Core Insight

Difficult queries elicit **divergent responses** even from strong models, while easy queries produce consistent answers across varying capabilities.

Training Approach

Dedicated classifier trained on self-consistency patterns across **5 model configurations**:

- GLM-4.5-Air (±CoT)
- Qwen-3-30B (±CoT)
- Solar Pro 2 (no reasoning)

Training Data Construction

Query Sampling

Sample queries from math, code, science, and medical domains, generating **~140K query-response tuples**.

Construct pairwise comparisons with **LLM-based labeling** determining difficulty ranking.

Estimator Applications

1. **Difficulty-Based Filtering:** Apply thresholds curating training data appropriate for each stage
2. **Enhanced Quality Control:** High-difficulty queries require stricter validation with manual inspection
3. **Balanced Sampling:** Stratify data by difficulty preventing over-representation of elementary queries

Complex Query Generation Pipeline

Step 1: Generator Fine-Tuning

Aggregate existing query data with related seed texts. Fine-tune **Solar Pro 2** as generator leveraging base model's general capabilities.

Challenge: Model initially generates predominantly trivial queries ineffective for reasoning development.

Step 2: Cyclic Optimization

Use difficulty classifier to assess and stratify generated queries. Iteratively retrain generator toward progressively higher complexity.

Assess → Stratify → Retrain → Repeat

Step 3: Domain Expansion

Apply optimized generator across diverse domain seeds. Construct robust dataset demanding reasoning capabilities across wide spectrum.

Result: Balanced query distribution across difficulty spectrum with sufficient coverage.

06 · Supervised Fine-Tuning

SFT: Agent Capability Development

Two complementary simulation pipelines for multi-turn tool-use trajectories

Task-Oriented Simulation

Generates complex single-task scenarios requiring multi-step reasoning and tool use. Synthesizes task specifications with explicit success criteria and expected tool-use patterns.

Complexity Levels

Easy: Single-Tool with Constraints
Direct tool application with bounded parameters

Medium: Multi-Tool with Dependencies
Sequential tool execution with output references

Hard: Extended Multi-Step with Branching
Complex workflows with conditional logic and error recovery

Output Statistics

Samples	Tokens
161,608	1.3B

User-Oriented Simulation

Models interactive process where simulated user iteratively refines requirements through dialogue. Generates high-level task then decomposes into sub-tasks across multiple turns.

Key Features

- Multi-turn conversations with **3.16 avg turns**
- **2.48 avg sub-tasks** per conversation
- Context maintenance across turns
- Incremental requirement refinement

Output Statistics

Samples	Tokens
177,375	3.0B

Tool Synthesis and API Graph Construction

Synthesize diverse tool sets by expanding API specifications. Utilize structured output format prediction and semantic expansion to model tool dependencies, enabling realistic multi-step workflows.

60
Tau2-Bench Score
Without dedicated RL training

06 · Supervised Fine-Tuning

SFT: Korean Knowledge Integration & Safety Framework

Multi-hop QA construction and comprehensive 38-category safety system

Korean Knowledge Integration

Korean cultural knowledge is typically fragmented across sources. Use **embedding-based matching** to construct multi-hop QA pairs linking related knowledge.

Comparative QA
Pairs related documents for comparison or contrast reasoning

Causal QA
Chains multiple documents into cause-effect sequences

Multi-Hop QA
Builds on information from multiple sources through multi-turn dialogue

Theme Inference QA
Requires identifying common patterns or hidden intentions

Integration Pipeline

This multi-hop construction enables the model to develop **relational understanding** beyond memorization of isolated facts.

SFT & DPO
Knowledge-augmented data trains general Korean capabilities

RL Stage
Targeted alignment for Korean cultural sensitivities using culturally-informed reward models

Comprehensive Safety Framework

Addresses **38 risk categories** with appropriate response strategies. Categories span child safety, violence, privacy, weapons, psychological harm, misinformation, political manipulation, and more.

Response Strategies

Refuse with Redirection
For harmful requests (bomb-making, hacking, fraud): explicit refusal with educational context about risks

Safe Completion
For sensitive topics (self-harm, suicide): supportive responses with professional resources and crisis hotlines

Preventing Over-Refusal

Construct **adversarial safety data**—queries designed to appear unsafe but are contextually appropriate.

Examples

- Educational discussions of historical atrocities
- Medical research questions
- Academic analysis of controversial topics

Critical Distinction

For self-harm queries, providing **empathetic support and professional resources** is safer than simple refusal, which may prevent users from seeking necessary help.

SnapPO: Decoupled RL Framework

Snapshot Sampling for Policy Optimization enabling scalable multi-objective training

Core Challenge: Traditional RL Limitations

Traditional online RL tightly couples data generation, reward computation, and training. When targeting multiple capabilities simultaneously (reasoning, safety, preference alignment, cultural nuances), this requires expensive infrastructure retuning for each objective, limiting scalability.

SnapPO Solution

A **cyclic off-policy framework** that decouples these three steps into independent processes with cached intermediate results. This enables **linear scaling** and **flexible multi-domain composition**.

1 Generation Step

Employ **vLLM** for efficient response generation at scale.

Process
For each prompt: 8-16 response candidates per prompt

Caching
Cache behavior policy's log probabilities for off-policy learning

Async Operation
Asynchronous from training, enabling pre-generation of large pools

2 Reward Computation

Compute domain-specific rewards in separate batch processing.

STEM
Verifiable correctness for closed-ended problems

Agent Sim
Multi-dimensional scoring for task completion

Open-Ended
Reward model-based evaluation for writing/reasoning

Degeneration
Pattern-based detection for repetition/errors

3 Training Step

Employ **Group Sequence Policy Optimization (GSPO)**

MoE Stability
Superior stability for sparse MoE architectures vs conventional methods

Memory Efficiency
No additional KL divergence term needed, enhancing memory efficiency

Infrastructure

Leverages same TorchTitan with HSDP and compilation optimizations

Iterative Cycles

Training proceeds through iterative cycles: model trained on prompt batch → updated policy generates new responses for next batch. Decoupled architecture enables **independent tuning** of generation throughput, reward complexity, and training hyperparameters, plus **rapid testing** and **resource flexibility**.

SnapPO: Key Advantages and Implementation

Concrete engineering benefits from decoupled architecture

Linear Scalability

By separating generation from training, SnapPO achieves **near-linear scaling** with compute resources.

Benefit
Adding nodes increases throughput proportionally

No Redesign
Without infrastructure redesign or hyperparameter re-tuning

Contrast
Coupled online RL competes for GPU memory and computation

Consistent scaling across
Hundreds of GPUs

TorchTitan Integration

Implemented using **TorchTitan** providing optimized FSDP and model parallelism.

Performance
Significantly faster than verl in benchmarks

Shared Infrastructure
HSDP, compilation optimizations shared between pre-training and RL

Rapid Transfer
Engineering improvements transfer quickly across training steps

Enabling rapid iteration on
Infrastructure Improvements

Multi-Domain Composition

Cached intermediate representation (responses + rewards) enables flexible mixing of data from diverse sources.

Dynamic Balancing
Math RL, code RL, agent simulation, safety data balanced dynamically

No Regeneration
Without regenerating responses or recomputing rewards

Two-Phase RL
Essential for reasoning optimization and preference alignment approach

Similar to approaches in
PRIME-RL Framework

Independent Optimization Dimensions

Generation Throughput
Scale independently without affecting training

Reward Complexity
Iterate on reward design without retraining

Training Hyperparams
Tune without affecting generation pipeline

Resource Allocation
Specialized hardware for each step

RL Phase A: Reasoning Optimization

Maximizing reasoning capabilities across STEM, agent workflows, and complex problem-solving

Phase Overview

First RL phase focuses **exclusively on maximizing reasoning capabilities** across STEM domains, agent workflows, and complex problem-solving. Trains on ~200K prompts sampled to emphasize challenging scenarios where exploration significantly improves over supervised learning.

Prompt Sources

- **Mid-training:** Query generation with difficulty-aware sampling
- **SFT:** Agent simulation with diverse tool-use scenarios
- **Synthesized:** Additional open-ended reasoning queries

Challenging Scenario Selection

Prompts filtered using **difficulty estimator** ensuring appropriate challenge levels. Multiple response candidates per prompt enable exploration of diverse reasoning strategies.

Training Method

Model explores diverse reasoning strategies and consolidates effective approaches through **reward-weighted gradient updates**.

Data Composition

STEM Reasoning

Closed-ended and open-ended mathematical and scientific problems:

- **Verifiable correctness rewards**
- **Process-based evaluation**
- Korean problems entirely synthetic (~50% of category)

Code Generation

Programming tasks with comprehensive evaluation:

- **Execution-based rewards**
- **Style and efficiency scoring**

Agent Simulation

Multi-turn tool-use scenarios with composite rewards:

- **Task completion quality**
- **Interaction quality metrics**
- **Error recovery capability**

Multi-Domain Composition

SnapPO enables efficient multi-domain composition, reflecting capabilities targeted during RL training. Prompt corpus demonstrates **reasoning optimization focus** with specialized reward functions per domain.

RL Phase B: DPO for Preference Alignment

Human preference alignment while maintaining reasoning capabilities

Phase Focus

Second RL phase shifts focus to **human preference alignment** while maintaining reasoning capabilities established in Phase A. Addresses writing quality, safety, and degeneration handling through **cyclic DPO**.

KL Divergence Regularization

KL divergence regularization in DPO loss prevents degradation of reasoning capabilities during preference optimization.

Conservative Exploration

This phase employs **more conservative exploration** compared to Phase A, prioritizing stable alignment over aggressive capability expansion.

Reasoning Maintenance

Inclusion of **Phase A data subset** ensures the model retains mathematical and agentic capabilities while optimizing for human preferences.

Data Composition

Human Preference

Preference pairs covering STEM explanations, creative writing, and conversational quality with model-based reward estimation

Safety Alignment

Scenarios spanning 38 risk categories with dedicated rewards for appropriate refusal and safe completion strategies

Degeneration Handling

Detection and penalization of repetition, language errors, and formatting issues

Korean Cultural Sensitivity

Targeted alignment for culturally sensitive Korean topics through specialized reward models

Agent Data

Scenarios for model's self-correction behavior on agentic workflows

Phase B Strategy Summary

Evaluation: Comprehensive Benchmark Suite

Korean and English benchmarks spanning knowledge, reasoning, and alignment

Korean Benchmarks

General Knowledge

- KMMLU & KMMLU-Pro
- CLIcK & HAE-RAE v1.1
- KoBALT

Domain-Specific

- **KBankMMLU** (Finance) – Derived from national qualification exams
- **KBL** (Law) – Korean legal language understanding
- **KorMedMCQA** (Medical) – Healthcare professional licensing

Mathematical Reasoning

- **Ko-AIME 2024/2025** – Translated official AIME exams
- **HRM8K** – Korean mathematical reasoning

Other Capabilities

- Ko-IFEval (IF)
- Ko Arena Hard v2 (Preference)

English Benchmarks

General Knowledge & Science

- MMLU & MMLU-Pro
- GPQA-Diamond
- HLE (text only)

Mathematical Reasoning

- **AIME 2024/2025** – Official American Invitational Math Exam
- **HMMT 2025** (Feb/Nov) – Harvard-MIT Math Tournament

Code & Instruction Following

- LiveCodeBench v6
- IFBench & IFEval

Preference, Agent & Long Context

- Arena Hard v2
- Writing Bench
- Tau2 (Air/Tel/Retail)
- AA-LCR

Evaluation Scope

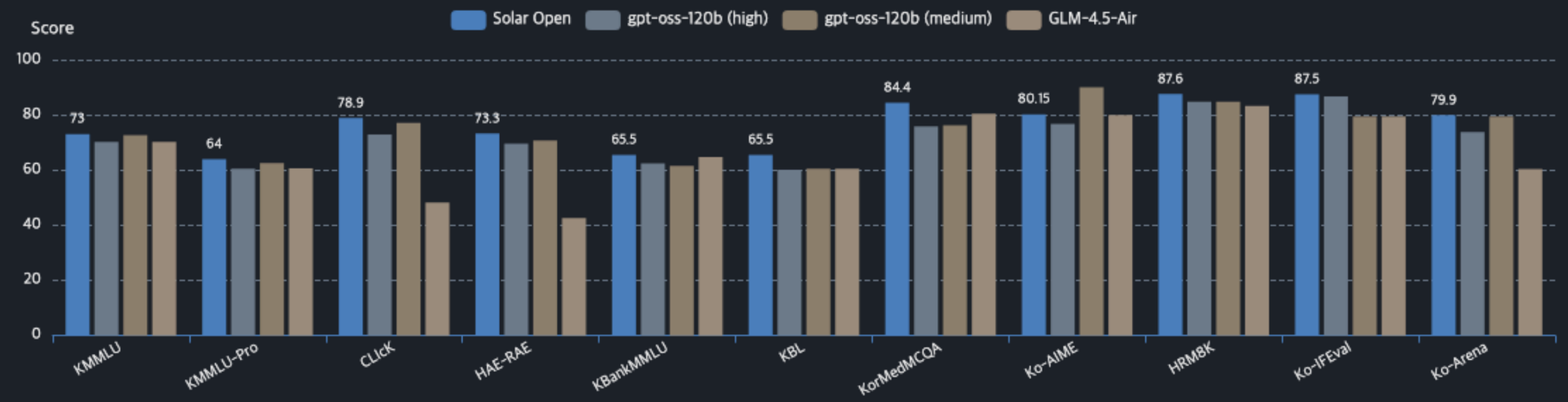
Comprehensive evaluation spans **general knowledge, domain expertise, reasoning, instruction following, and preference alignment** in both Korean and English. Benchmarks selected to validat

08 · Evaluation

Korean Benchmark Results: Domain Excellence

Leading performance across finance, law, and medical domains

Korean Benchmark Performance



General Knowledge Leadership

KMMLU	73.0 (+2.7pp)
KMMLU-Pro	64.0 (+1.4pp)
CLIcK	78.9 (+1.7pp)
HAE-RAE v1.1	73.3 (+2.5pp)

vs gpt-oss-120b-high baseline

Domain-Specific Excellence

Finance: KBankMMLU

65.5

Leading all baselines by +0.8–4.0pp

Law: KBL

65.5 (+2.7pp)

Medical: KorMedMCQA

84.4 (+8.6pp)

Other Capabilities

Preference (Ko-Arena Hard v2)	79.9
Math (Ko-AIME 2024)	80.3
Math (Ko-AIME 2025)	80.0
Math (HRM8K)	87.6
IF (Ko-IFEval)	87.5

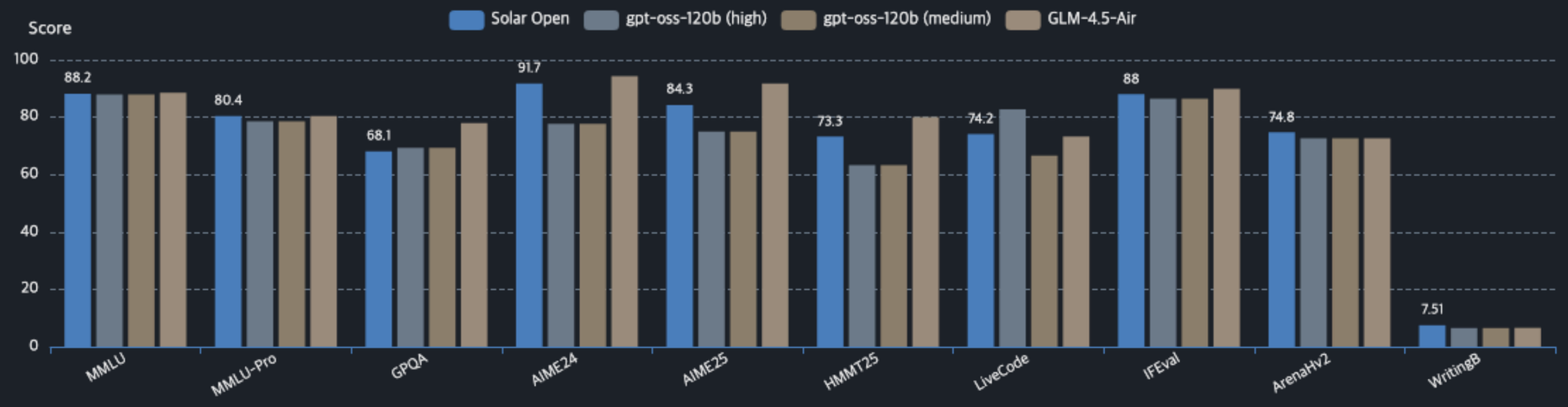
Performance Analysis

08 · Evaluation

English Benchmark Results: Competitive Performance

Strong English capabilities while advancing Korean language AI

English Benchmark Performance



General Knowledge & Science

MMLU	88.2
MMLU-Pro	80.4 (+1.8pp)
GPQA-Diamond	68.1

Mathematical Excellence

AIME 2024
91.7 (+14.0pp)

AIME 2025
84.3 (+9.3pp)

HMMT 2025 Feb
73.3 (+10.0pp)

vs gpt-oss-120b-medium

Other Capabilities

Code (LiveCodeBench v6)	74.2
Preference (Arena Hard v2)	74.8
Writing (Writing Bench)	7.51
Agent (Tau2 Avg)	55.8
Long Context (AA-LCR)	35.0

Performance Analysis: Data Strategy Impact

How design choices drive performance profile

Data Composition Strategy

Performance profile directly reflects **data composition strategy** detailed in Section 3 and Section 6. Key design choices:

Korean Synthetic Investment

Pre-training allocated **4.5T of 20T total tokens** (22.5%) to Korean synthetic data generation—substantial investment addressing data scarcity.

Natural Text Prioritization

Prioritized **natural text over mathematical content**, reflecting strategic focus on domain expertise and cultural knowledge.

Preference Alignment Focus

Two-phase RL emphasized **preference alignment** in Phase B rather than aggressive reasoning expansion.

Strategic Trade-offs

These design choices produced **strong domain expertise and human preference alignment for Korean**—capabilities particularly relevant for practical deployment in underserved language contexts.

Mathematical Reasoning

While behind leading specialized models, mathematical reasoning remains **sufficient for many applications** and reflects deliberate trade-off. Could be improved through targeted continual training for specific use cases.

Performance Profile Analysis

Korean Domain Leadership

Finance: +3.0pp | **Law:** +2.7pp | **Medical:** +8.6pp

Strong domain expertise directly results from 0.4T manually curated domain-specific content and targeted synthetic generation.

Preference Alignment Success

Ko-Arena Hard v2: 79.9 | **Arena Hard v2:** 74.8

Strong preference alignment reflects Phase B focus on writing quality, safety, and cultural sensitivity through DPO training.

Maintained English Competitiveness

Performance comparable to GLM-4.5-Air across most categories

Bilingual curriculum successfully balances language capabilities without catastrophic forgetting of English knowledge.

Validation of Methodology

Performance profile validates **core thesis**: Carefully designed data curriculum and synthesis strategies can dramatically improve training efficiency for underserved languages without compromising English performance. **Solar Open achieves domain leadership where it matters most for Korean users**—practical domain expertise and preference alignment—while maintaining broadly competitive general capabilities.

Conclusion: Methodological Innovations

Three breakthrough innovations addressing underserved language challenges

1 Synthetic Data Generation

4.5T
Tokens Generated

Aggressive synthetic data generation overcomes Korean data s carcity through diverse augmentation, filtering, and transforma tion pipelines.

Mid-Training
Queries + diverse reasoning trajectories

SFT
Successful problem-solving trajectories

RL
Queries for compositional reasoning

2 Bilingual Curriculum

20T
Tokens Coordinated

Bilingual curriculum optimization with language-aware quality filtering, educational scoring, and topic clustering.

Progressive Quality
10% → 64% synthetic ratio

Language Balance
Korean-English coordination

Efficiency Gain
48% English, 77% Korean tokens

3 SnapPO Framework

3
Decoupled Stages

Decoupled RL framework enabling scalable multi-objective tra ining through generation-reward-training separation.

Linear Scalability
Direct throughput increase

Flexible Composition
Multi-domain mixing

Independent Optimization
Separate tuning dimensions

Unified Methodology

These three innovations are **deeply interconnected**, unified by curriculum-coordinated approach to bilingual data and reasoning development. Data scarcity drives aggressive synthetic generation, whic h progressive curriculum coordinates across stages. This RL-oriented data strategy feeds directly into SnapPO, whose decoupled architecture makes multi-domain, multi-objective training tractable at s cale. The result is **cohesive methodology where data generation, curriculum design, and RL framework mutually reinforce** to overcome challenges facing underserved languages.

Validation Results: Domain Leadership & Efficiency

Empirical validation of systematic methodology

Korean Domain Expertise

Korean domain expertise—the **primary target of data strategy**—shows substantial advantages over gpt-oss-120b-high:

Finance	Law	Medical
+3.0pp	+2.7pp	+8.6pp
KBankMMLU	KBL	KorMedMCQA

General Knowledge & Preference

Ko-Arena Hard v2: **79.9**

Arena Hard v2: **74.8**

English Performance

English performance is **competitive with GLM-4.5-Air** across most categories:

General Knowledge

MMLU: **88.2**

MMLU-Pro: **80.4**

Math Excellence

AIME 2024: **91.7**

AIME 2025: **84.3**

Training Efficiency Achievement

Solar Open demonstrates **substantial efficiency gains** compared to GLM-4.5-Base (106B, 23T tokens):

English Benchmarks

At Token Budget **10.7T**

Only **48%** of GLM's training budget

Korean Benchmarks

At Token Budget **17.8T**

Only **77%** of GLM's training budget

Validation Summary

These results validate the **core thesis**: Carefully designed data curriculum and synthesis strategies can dramatically improve training efficiency for underserved languages without compromising English performance.

Key Achievement

Solar Open achieves **domain leadership where it matters most for Korean users**—practical domain expertise and preference alignment—while maintaining broadly competitive general capabilities.

Future Directions & Open Questions

Research opportunities extending beyond Solar Open

Lower-Resource Languages

While methodology effectively addresses Korean's data scarcity, its applicability to **even lower-resource languages** remains an open question requiring empirical validation.

Research Questions

- How do techniques scale with **extreme data scarcity**?
- What **language-specific adaptations** are necessary?
- How to handle **diverse scripts and linguistic structures**?

Assumption-Free Curriculum

Current data curriculum relies on ML-based filtering models trained on specific assumptions. Exploring **more assumption-free approaches** would reduce complexity.

Goals

- Reduce infrastructure complexity
- Improve accessibility for resource-constrained communities
- Maintain comparable curriculum quality

Fundamental RL Challenges

While SnapPO enables multi-objective RL framework, **fundamental challenges** require continued research.

Key Areas

- **Reward Design:** Better reward functions capturing human preferences
- **Exploration Efficiency:** More effective discovery of reasoning strategies
- **Alignment Robustness:** Ensuring stable behavior across diverse inputs

Two-phase approach (reasoning then preference) provides one pattern, but alternative decompositions warrant exploration.

Language Scaling Laws

Establishing **principled language scaling laws** would provide critical insights for multilingual model development.

Fundamental Questions

- How does adding target language affect **performance across existing languages**?
- What are optimal **data allocation strategies** given fixed budgets?
- How to guide **continual training vs. fresh training** decisions?

Solar Open was built from scratch, but **continual training paradigm** deserves systematic study across architectures.

Solar Open Technical Report

Democratizing AI for Underserved Languages

Solar Open establishes a systematic methodology for building competitive LLMs in data-scarce language contexts, demonstrating that carefully designed data curriculum, aggressive synthetic data generation, and decoupled RL frameworks can achieve domain-leading performance while maintaining training efficiency.

102B Parameters
Mixture-of-Experts Architecture

20T Tokens
Bilingual Curriculum Training

Domain Leadership
Korean Language Excellence

As a case study in methodological development for underserved languages, Solar Open provides a **blueprint for expanding AI access** to languages beyond the dominant English-Chinese duopoly.

- Open-Weight Model
- Community-Driven Innovation
- Democratized Access

Upstage Solar Team | January 5, 2026