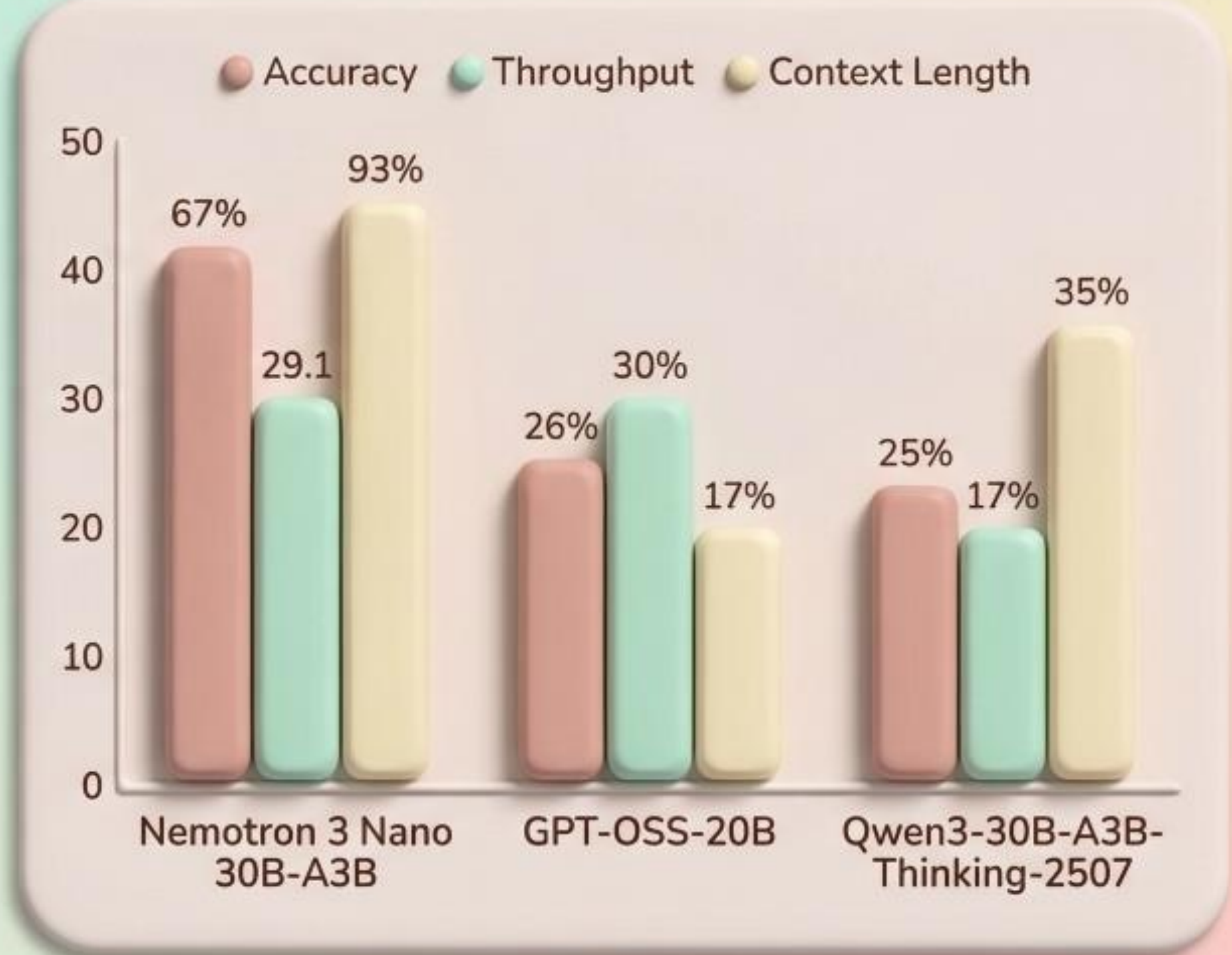

Nemotron 3
Nano
Open, Efficient Mixture-of-Experts
Hybrid Mamba-Transformer Model
Focus on Data
Mixture & Categories

Executive Summary & Key Achievements

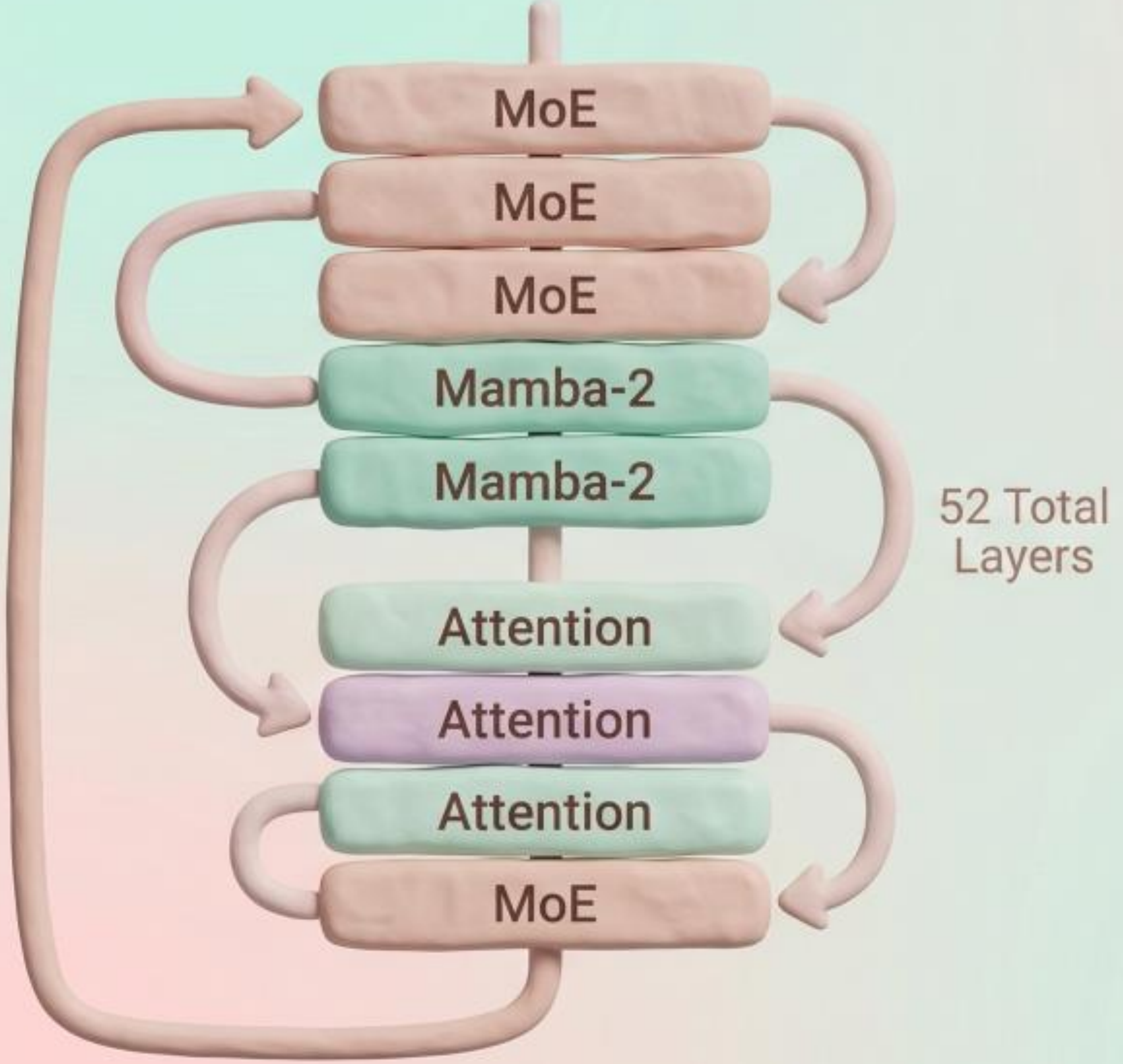
Nemotron 3 Nano 30B-A3B Core Achievements

- Pretrained on 25 Trillion Tokens (3T New Unique Tokens)
- Better Accuracy, <1/2 Activated Parameters
- 3.3x Higher Inference Throughput
- Supports 1M Context Length
- Enhanced Agentic, Reasoning, & Chat Abilities

Performance Comparison



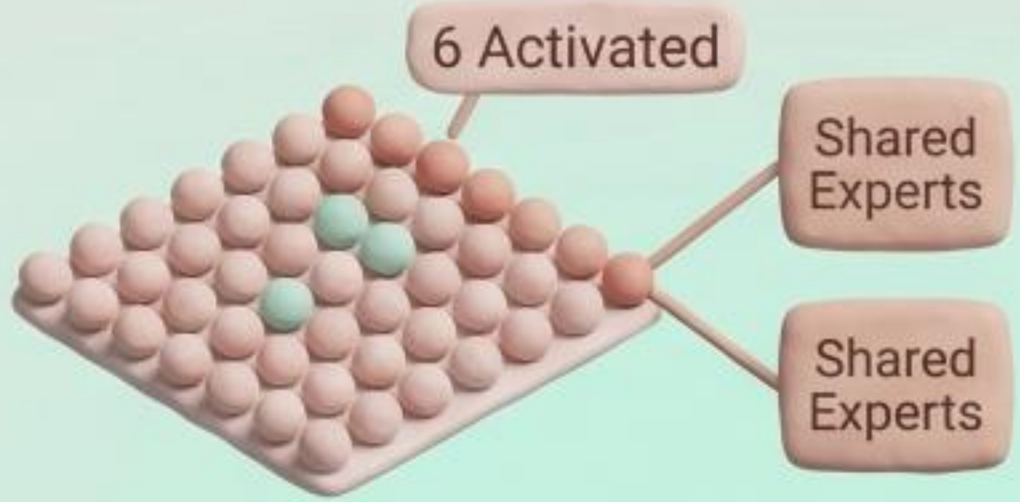
Model Architecture Overview



31.6B
Total Parameters

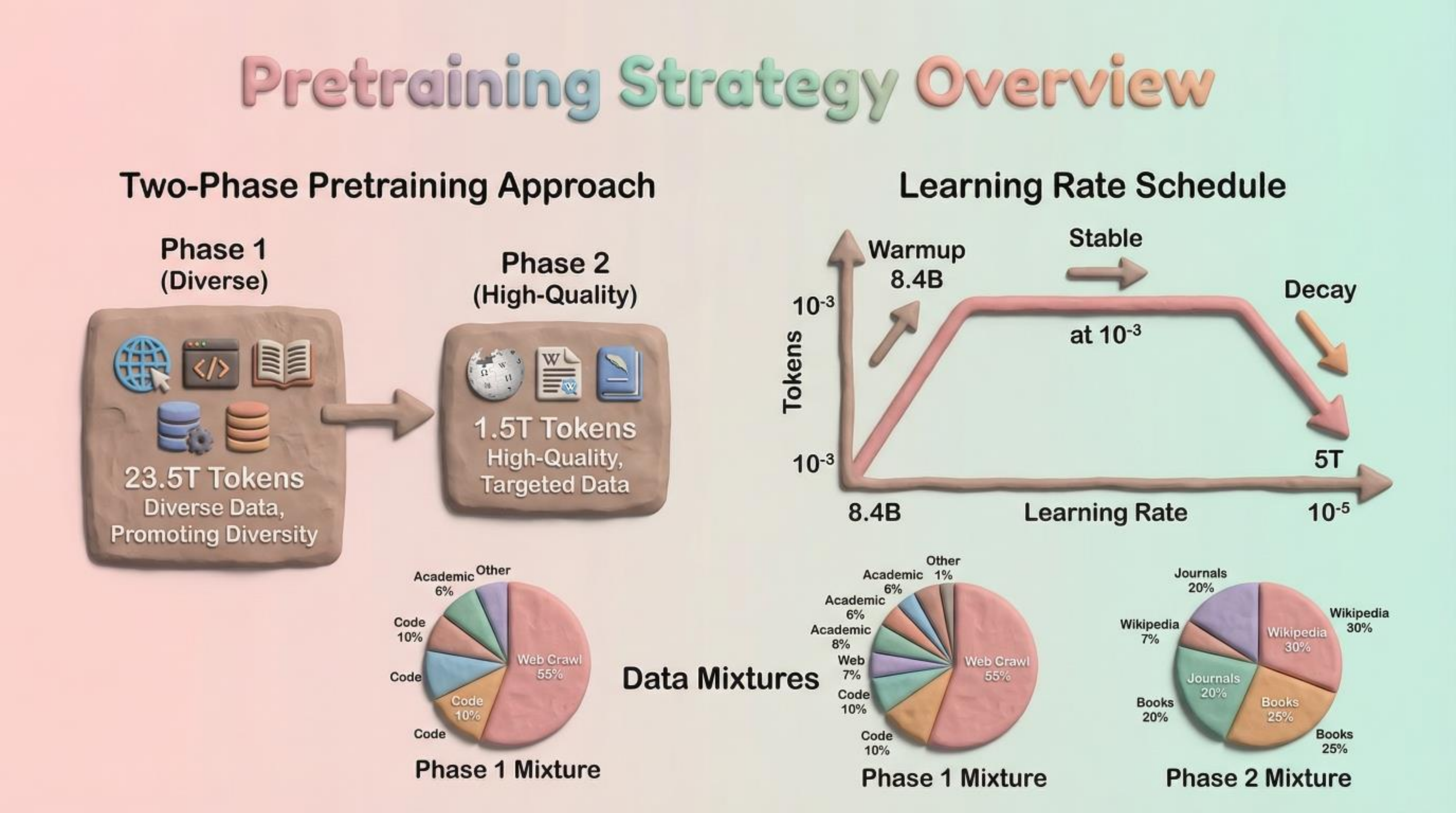
3.2B
Activated (3.6B with embeddings)

Granular MoE



128 Experts, 6 Activated + 2 Shared

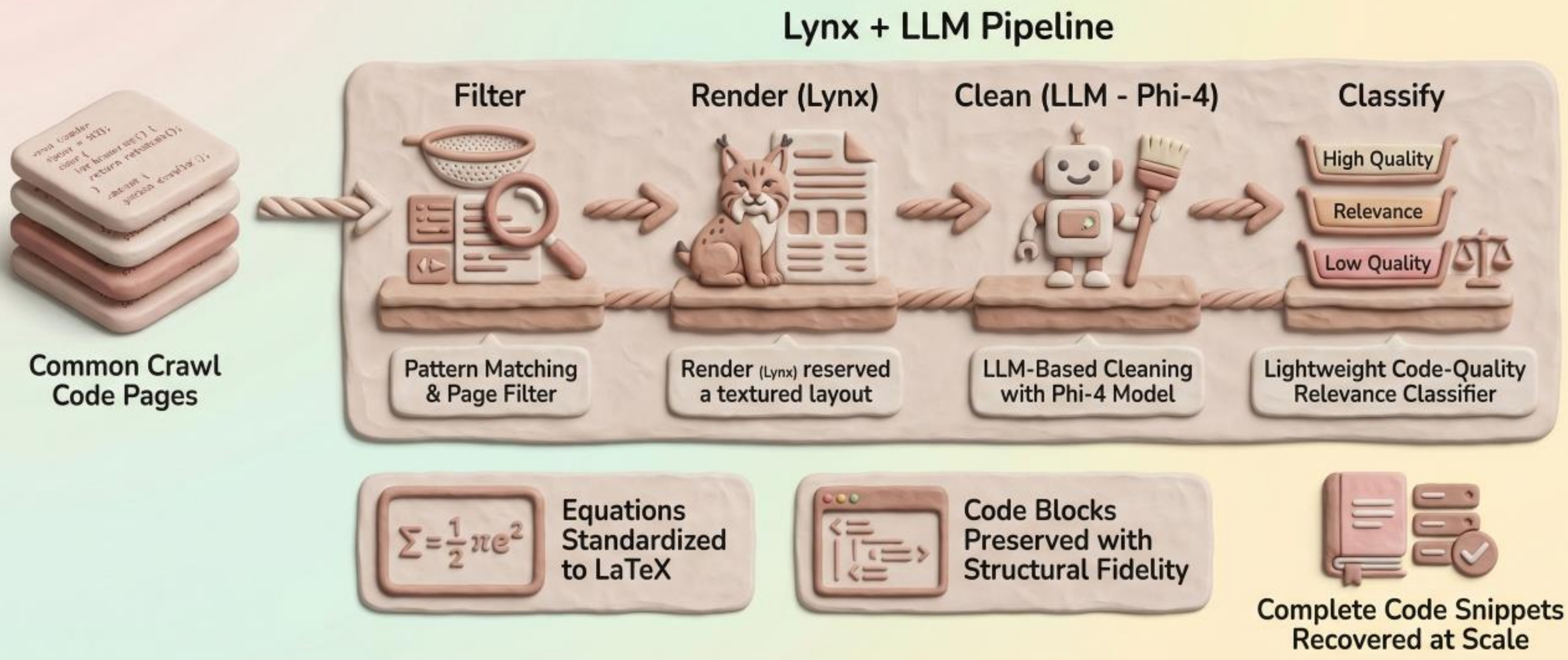
Architecture Specifications & Hyperparameters

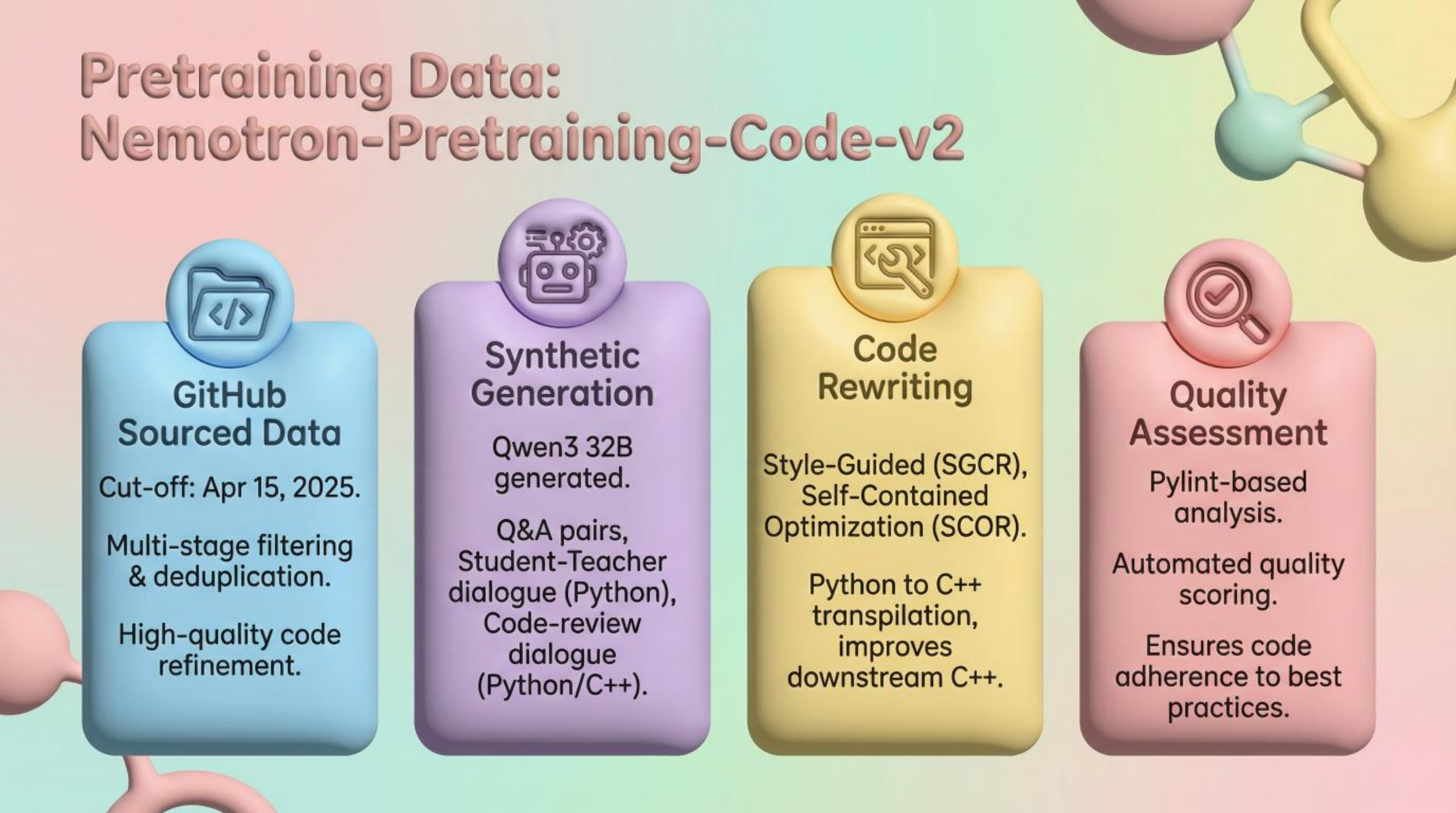


Pretraining Strategy Overview
Two-Phase Pretraining Approach
Phase 1
(Diverse)
23.5T Tokens
Diverse Data,
Promoting Diversity
Phase 2
(High-Quality)
1.5T Tokens
High-Quality,
Targeted Data
Learning Rate Schedule
Warmup
8.4B
Stable
at 10-3
Decay
Tokens
10-3
10-3
5T
8.4B
Learning Rate
10-5
Data Mixtures
Academic 6%
Other
Code 10%
Code
Web Crawl 55%
Code 10%
Code
Phase 1 Mixture
Academic 6%
Other 1%
Academic 6%
Academic 8%
Web 7%
Code 10%
Code 10%
Web Crawl 55%
Phase 1 Mixture
Journals 20%
Wikipedia 30%
Wikipedia 7%
Wikipedia 30%
Journals 20%
Books 20%
Books 25%
Books 25%
Phase 2 Mixture

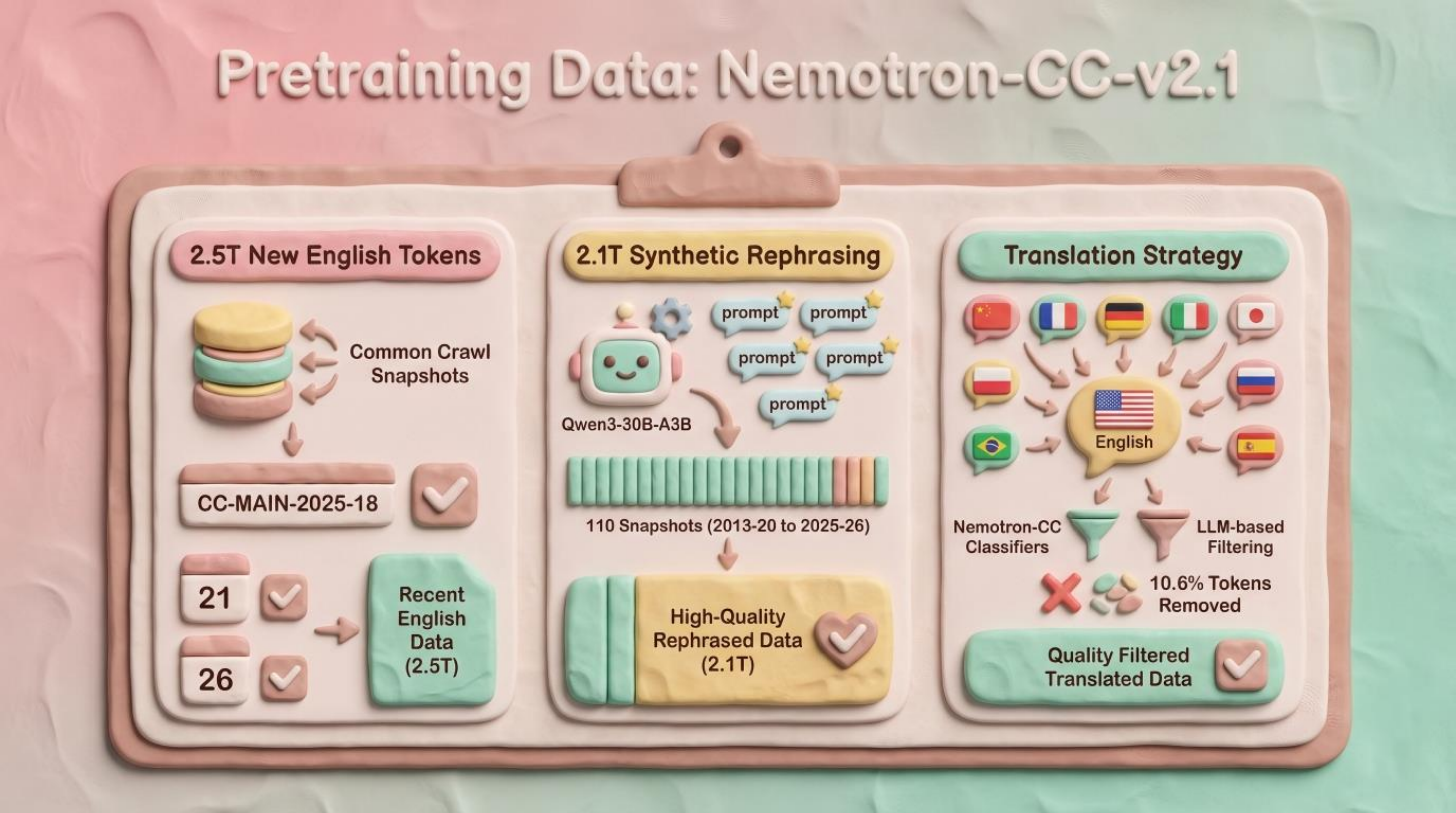
Pretraining Data: Nemotron-CC-Code-v1

428B-Token High-Quality Code Dataset from Common Crawl





Pretraining Data:
Nemotron-Pretraining-Code-v2
GitHub Sourced Data
Cut-off: Apr 15, 2025.
Multi-stage filtering & deduplication.
High-quality code refinement.
Synthetic Generation
Qwen3 32B generated.
Q&A pairs, Student-Teacher dialogue (Python), Code-review dialogue (Python/C++).
Code Rewriting
Style-Guided (SGCR), Self-Contained Optimization (SCOR).
Python to C++ transpilation, improves downstream C++.
Quality Assessment
Pylint-based analysis.
Automated quality scoring.
Ensures code adherence to best practices.


Pretraining Data: Nemotron-CC-v2.1
2.5T New English Tokens
Common Crawl Snapshots
CC-MAIN-2025-18
21
26
Recent English Data (2.5T)
2.1T Synthetic Rephrasing
Qwen3-30B-A3B
prompt
prompt
prompt
prompt
prompt
110 Snapshots (2013-20 to 2025-26)
High-Quality Rephrased Data (2.1T)
Translation Strategy
English
Nemotron-CC Classifiers
LLM-based Filtering
10.6% Tokens Removed
Quality Filtered Translated Data

Pretraining Data: Nemotron-Pretraining-Specialized-v1

Comprehensive synthetic specialized datasets for STEM and scientific domains.

Synthetic Wikipedia & Educational

Synthetic Wikipedia:
Revised for clarity & formatting.

Synthetic Math Textbook:
Undergraduate+ level,
educational features.

Scientific & Cross-Domain Coding

Synthetic Scientific Coding:
Code-embedded articles &
computational problems.

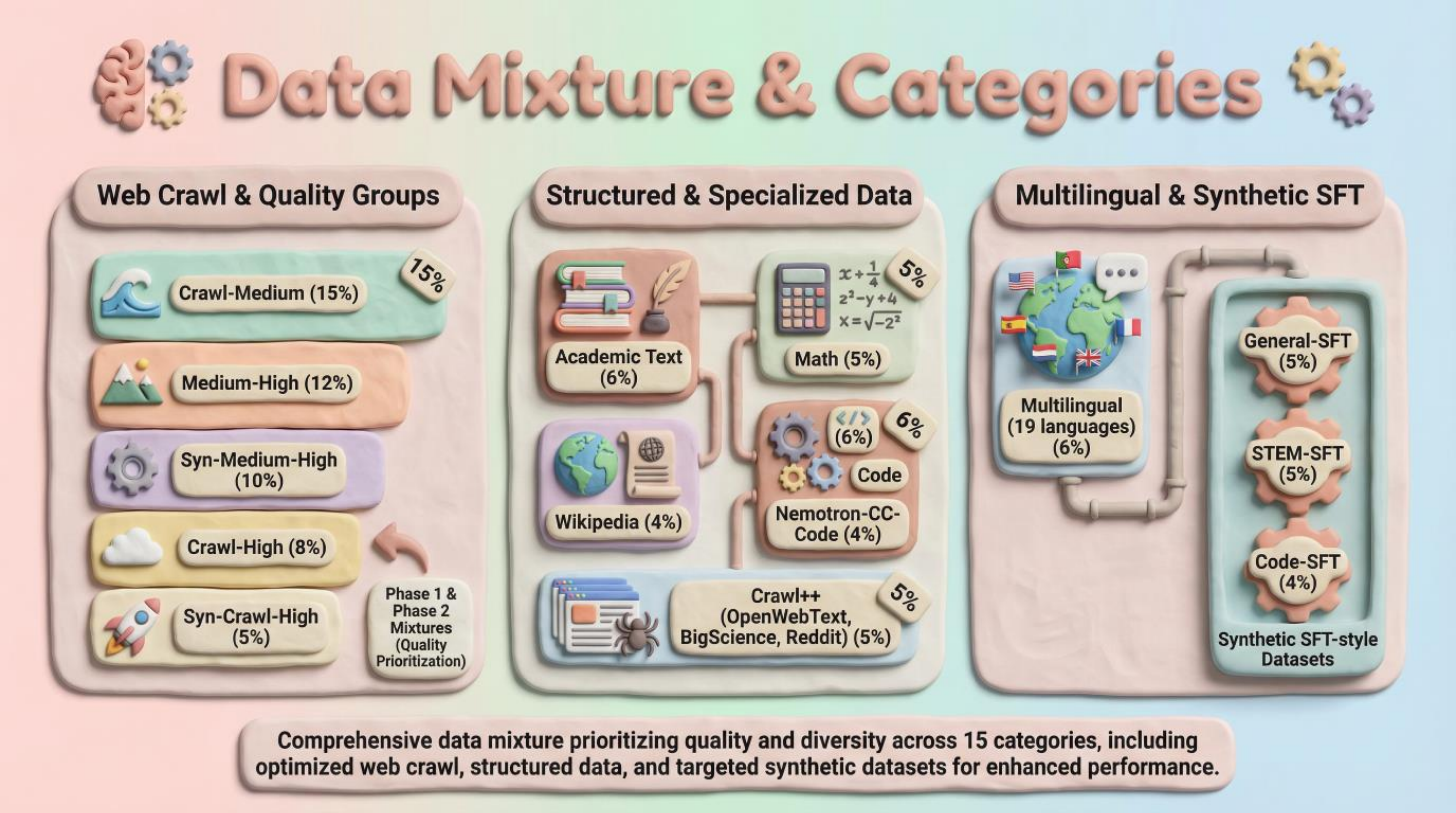
Synthetic Cross-Domain Code:
InfiniByte approach, cross-
breeding datasets.

STEM Reasoning & RQA

Synthetic STEM Reasoning:
RQA dataset with 4.3M
demonstrations, 31.7B tokens.

Additional SFT-Style Data

SFT-style data for code,
math, and STEM to enhance
performance.


Data Mixture & Categories
Web Crawl & Quality Groups
15%
Crawl-Medium (15%)
Medium-High (12%)
Syn-Medium-High (10%)
Crawl-High (8%)
Syn-Crawl-High (5%)
Phase 1 & Phase 2 Mixtures (Quality Prioritization)
Structured & Specialized Data
Academic Text (6%)
5%
x + 1/4
2²−y+4
x=√−2²
Math (5%)
Wikipedia (4%)
6%
</>
(6%)
Code
Nemotron-CC-Code (4%)
5%
Crawl++ (OpenWebText, BigScience, Reddit) (5%)
Multilingual & Synthetic SFT
Multilingual (19 languages) (6%)
General-SFT (5%)
STEM-SFT (5%)
Code-SFT (4%)
Synthetic SFT-style Datasets
Comprehensive data mixture prioritizing quality and diversity across 15 categories, including optimized web crawl, structured data, and targeted synthetic datasets for enhanced performance.

Long-Context Extension Phase

Training Overview

Continuous Pretraining

121B Tokens

Constant Learning Rate: 10^{-5}

Parallelism Strategy on H100

8-way Context, Tensor, Expert Parallelism

H100 GPUs

4-way Pipeline Parallelism

Data Mixture for Long Context

79% Phase 2 Data

20% – Document QA

–1% Synthetic Retrieval (256k max)

Performance Goal

4k

512k

Improves both short (4k) and long-context (512k) performance.

Base Model Evaluation Results

Comprehensive Benchmark Comparison: Nemotron 3 Nano vs. Qwen3

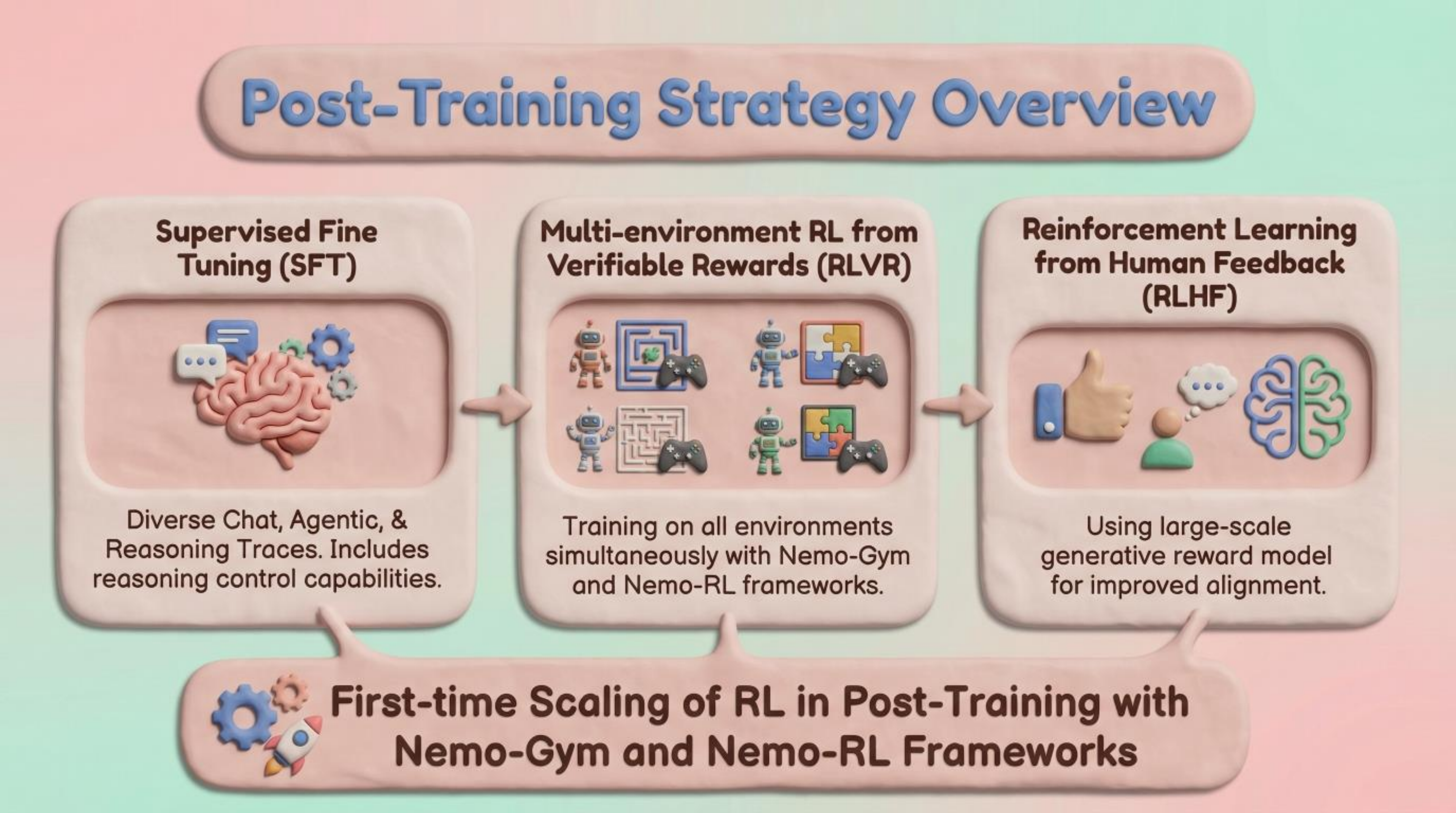
Nemotron 3 Nano 30B-A3B Base

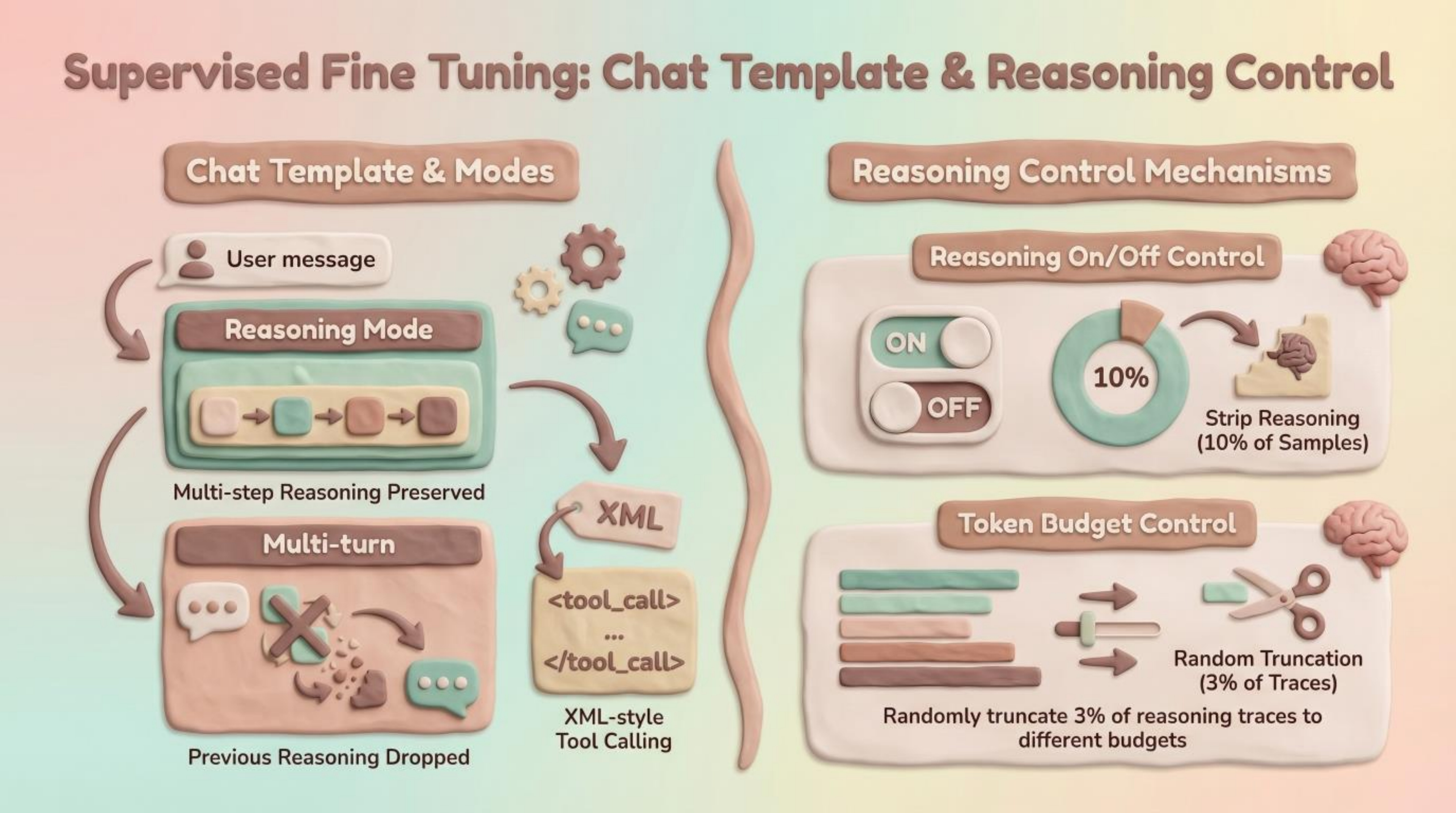
General Knowledge	MMLU 78.5% > 75.1%
Code	HumanEval 82.0% > 79.5%
Math	GSM8K 85.2% > 81.9%
Commonsense	ARC-Challenge > 72.0%
Reading	RACE > 74.7%
Multilingual	MGSM > 79.7%
Long Context	RULER > 77.9%

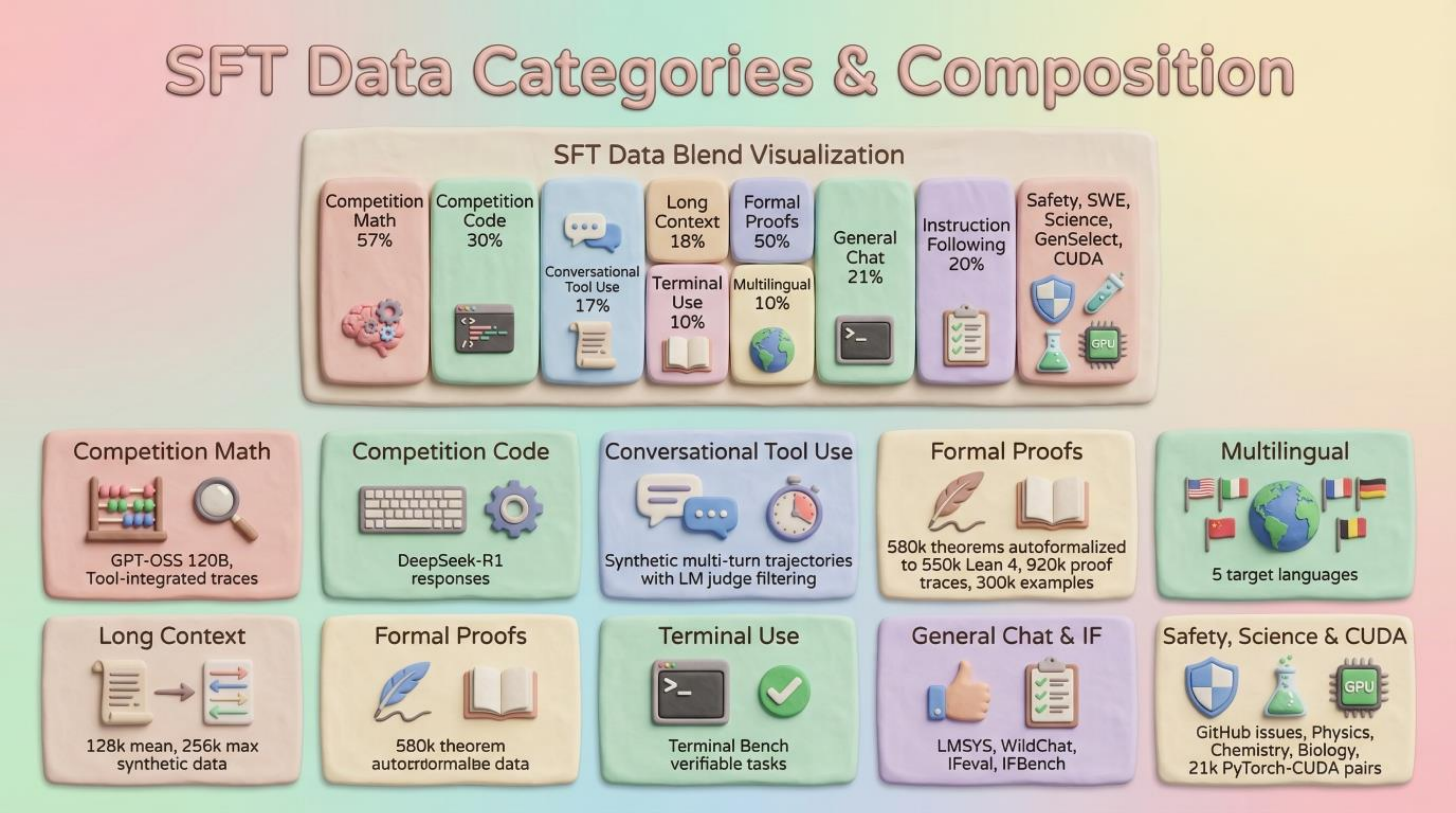
Qwen3-30B-A3B-Base

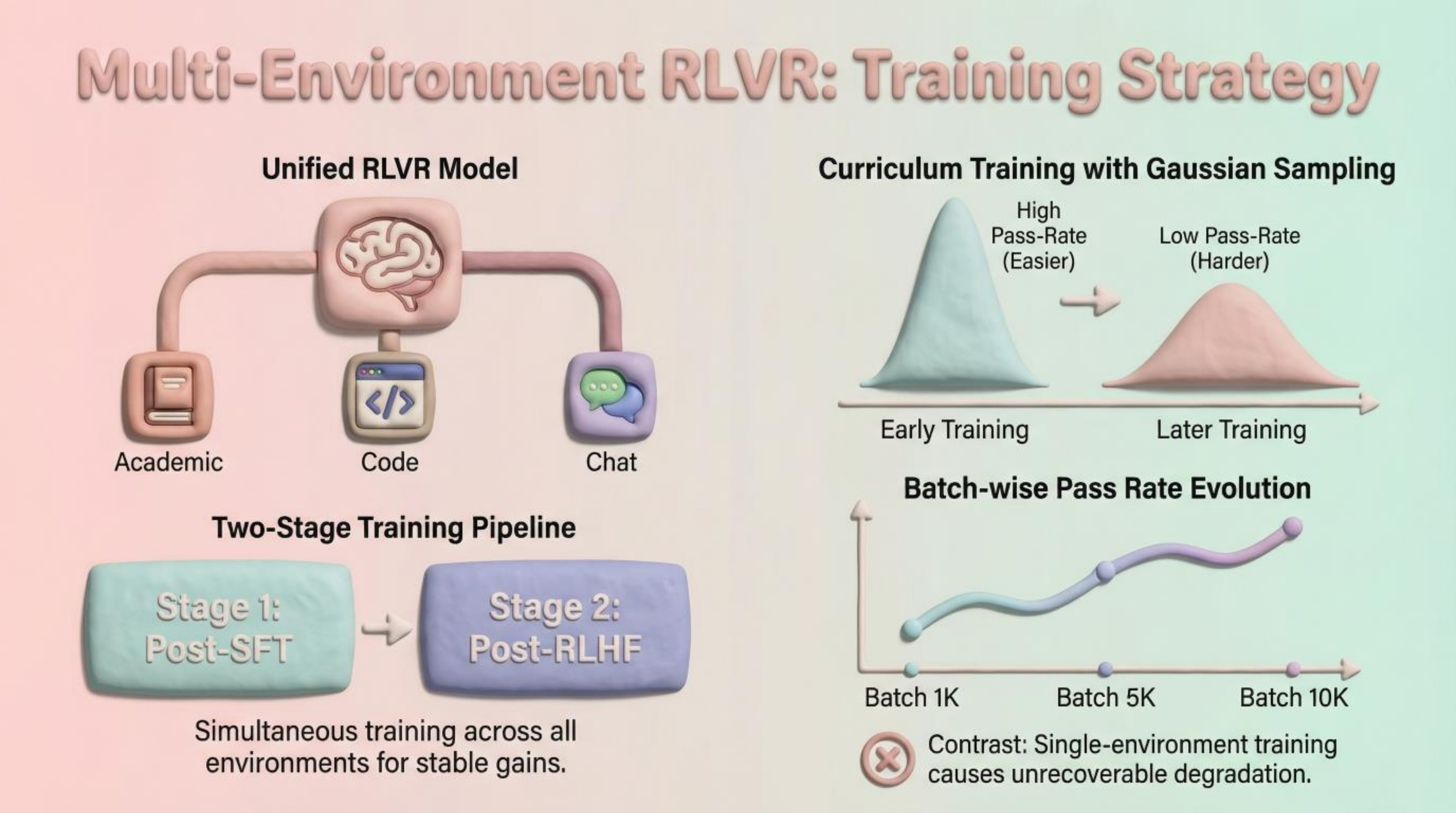
General Knowledge	MMLU 78.5% > 75.1%
Code	HumanEval 82.0% > 79.5%
Math	GSM8K 85.2% > 81.9%
Commonsense	ARC-Challenge > 62.3%
Reading	RACE > 69.6%
Multilingual	MGSM 83.8% > 68.3%
Long Context	RULER > 93.3%

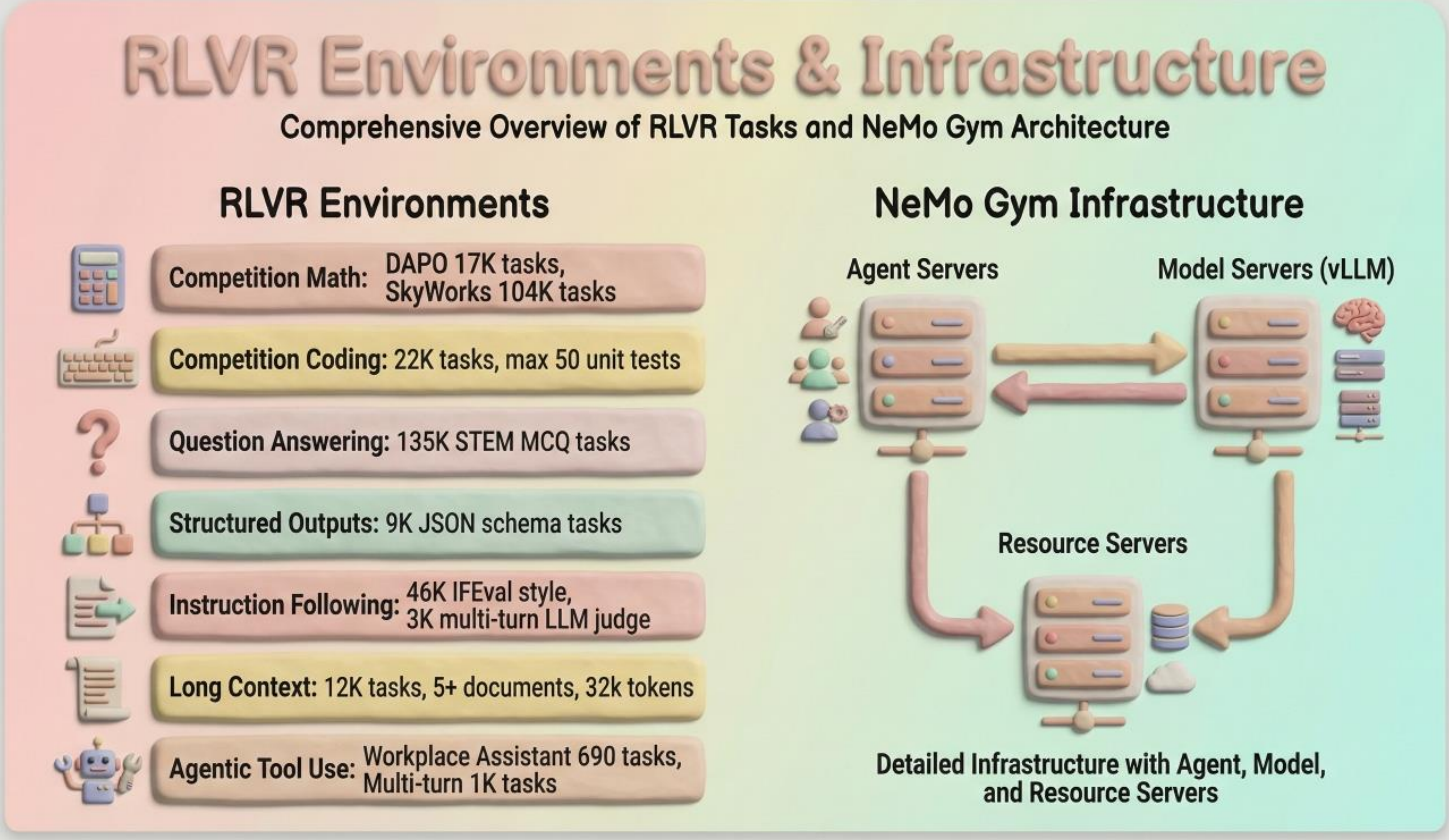
Highlight: Nemotron 3 Nano demonstrates superior performance in most categories, leading in General, Code, and Math benchmarks.


Post-Training Strategy Overview
Supervised Fine Tuning (SFT)
Diverse Chat, Agentic, & Reasoning Traces. Includes reasoning control capabilities.
Multi-environment RL from Verifiable Rewards (RLVR)
Training on all environments simultaneously with Nemo-Gym and Nemo-RL frameworks.
Reinforcement Learning from Human Feedback (RLHF)
Using large-scale generative reward model for improved alignment.
First-time Scaling of RL in Post-Training with Nemo-Gym and Nemo-RL Frameworks


Supervised Fine Tuning: Chat Template & Reasoning Control
Chat Template & Modes
User message
Reasoning Mode
Multi-step Reasoning Preserved
Multi-turn
Previous Reasoning Dropped
XML
<tool_call>
...
</tool_call>
XML-style
Tool Calling
Reasoning Control Mechanisms
Reasoning On/Off Control
ON
OFF
10%
Strip Reasoning
(10% of Samples)
Token Budget Control
Random Truncation
(3% of Traces)
Randomly truncate 3% of reasoning traces to
different budgets


SFT Data Categories & Composition
SFT Data Blend Visualization
Competition Math 57%
Competition Code 30%
Conversational Tool Use 17%
Long Context 18%
Terminal Use 10%
Formal Proofs 50%
Multilingual 10%
General Chat 21%
Instruction Following 20%
Safety, SWE, Science, GenSelect, CUDA
GPU
Competition Math
GPT-OSS 120B, Tool-integrated traces
Competition Code
DeepSeek-R1 responses
Conversational Tool Use
Synthetic multi-turn trajectories with LM judge filtering
Formal Proofs
580k theorems autoformalized to 550k Lean 4, 920k proof traces, 300k examples
Multilingual
5 target languages
Long Context
128k mean, 256k max synthetic data
Formal Proofs
580k theorem autotrdormalbe data
Terminal Use
Terminal Bench verifiable tasks
General Chat & IF
LMSYS, WildChat, IFeval, IFBench
Safety, Science & CUDA
GPU
GitHub issues, Physics, Chemistry, Biology, 21k PyTorch-CUDA pairs


Multi-Environment RLVR: Training Strategy
Unified RLVR Model
Academic
Code
Chat
Two-Stage Training Pipeline
Stage 1: Post-SFT
Stage 2: Post-RLHF
Simultaneous training across all environments for stable gains.
Curriculum Training with Gaussian Sampling
High Pass-Rate (Easier)
Low Pass-Rate (Harder)
Early Training
Later Training
Batch-wise Pass Rate Evolution
Batch 1K
Batch 5K
Batch 10K
Contrast: Single-environment training causes unrecoverable degradation.

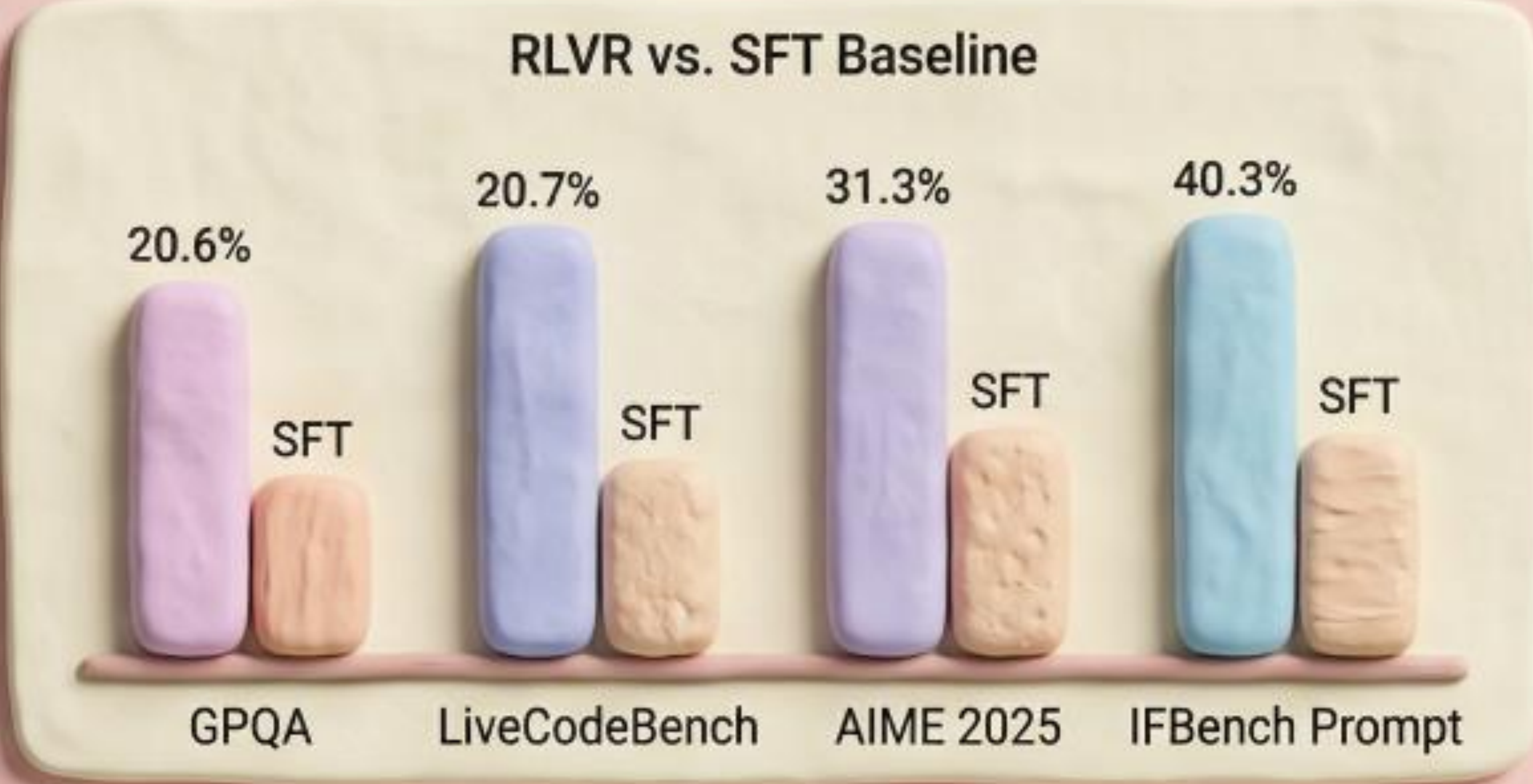

RLVR Environments & Infrastructure
Comprehensive Overview of RLVR Tasks and NeMo Gym Architecture
RLVR Environments
Competition Math: DAPO 17K tasks, SkyWorks 104K tasks
Competition Coding: 22K tasks, max 50 unit tests
Question Answering: 135K STEM MCQ tasks
Structured Outputs: 9K JSON schema tasks
Instruction Following: 46K IFEval style, 3K multi-turn LLM judge
Long Context: 12K tasks, 5+ documents, 32k tokens
Agentic Tool Use: Workplace Assistant 690 tasks, Multi-turn 1K tasks
NeMo Gym Infrastructure
Agent Servers
Model Servers (vLLM)
Resource Servers
Detailed Infrastructure with Agent, Model, and Resource Servers

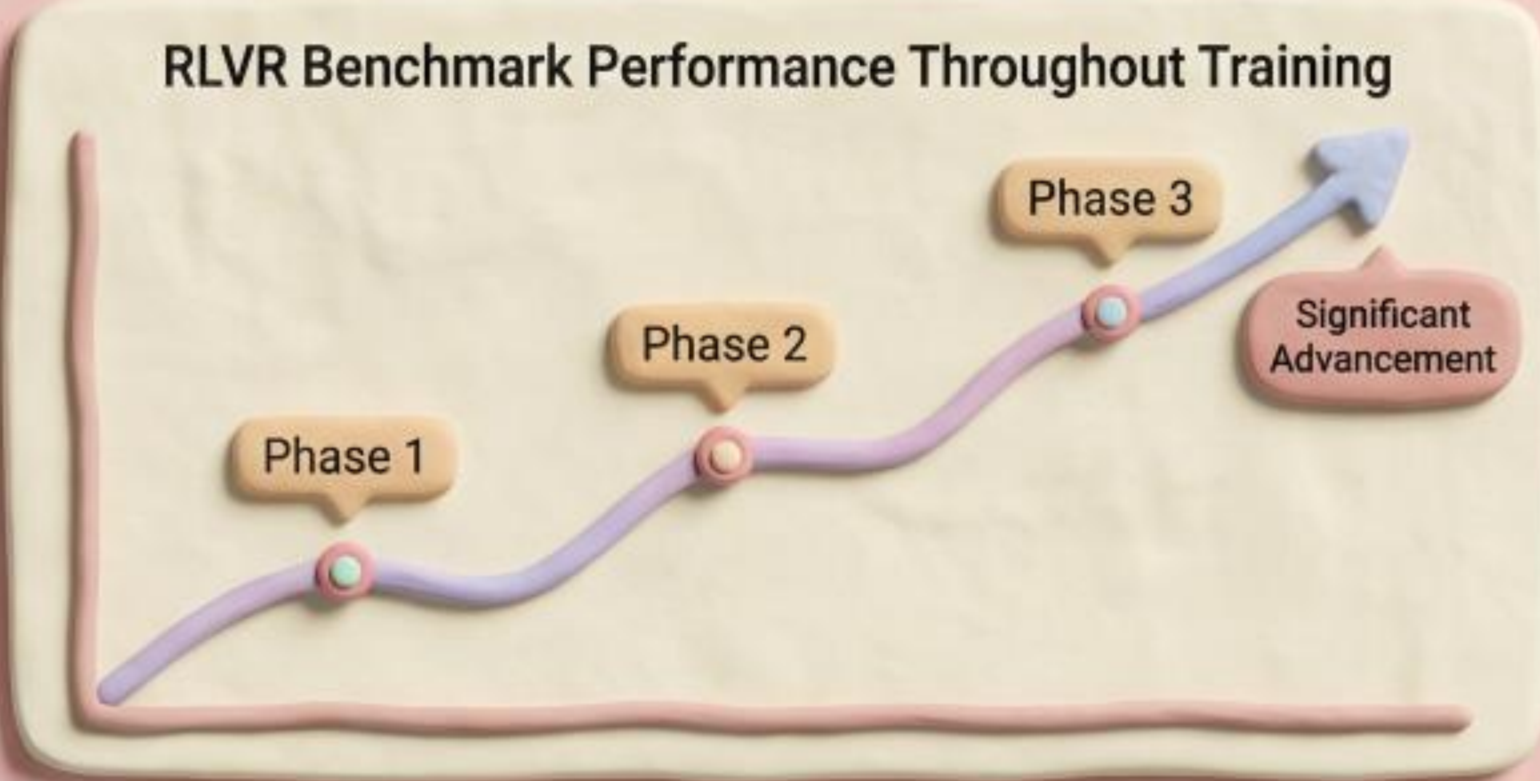
RLVR Algorithm & Performance

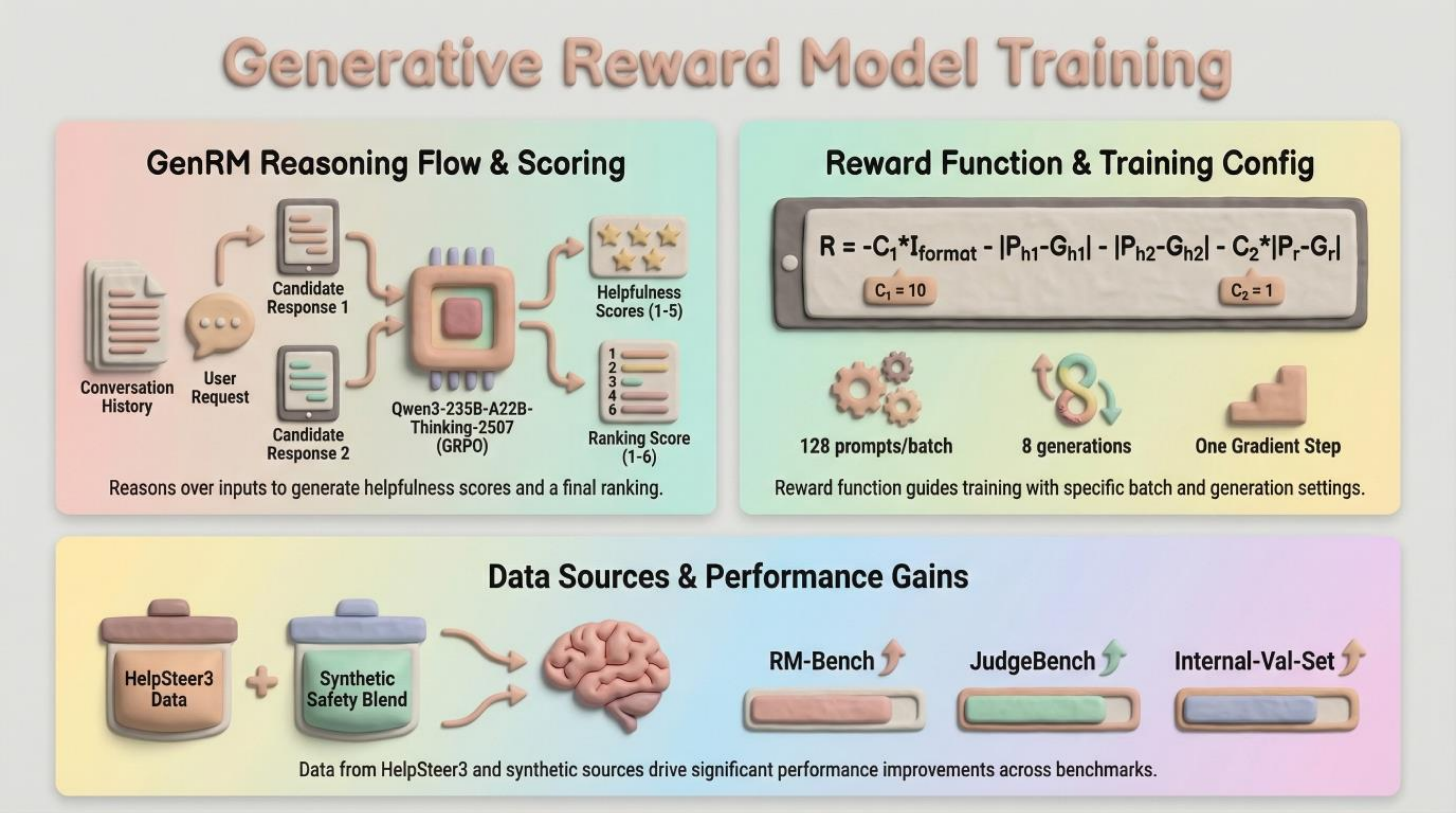
Algorithm & Training Pipeline

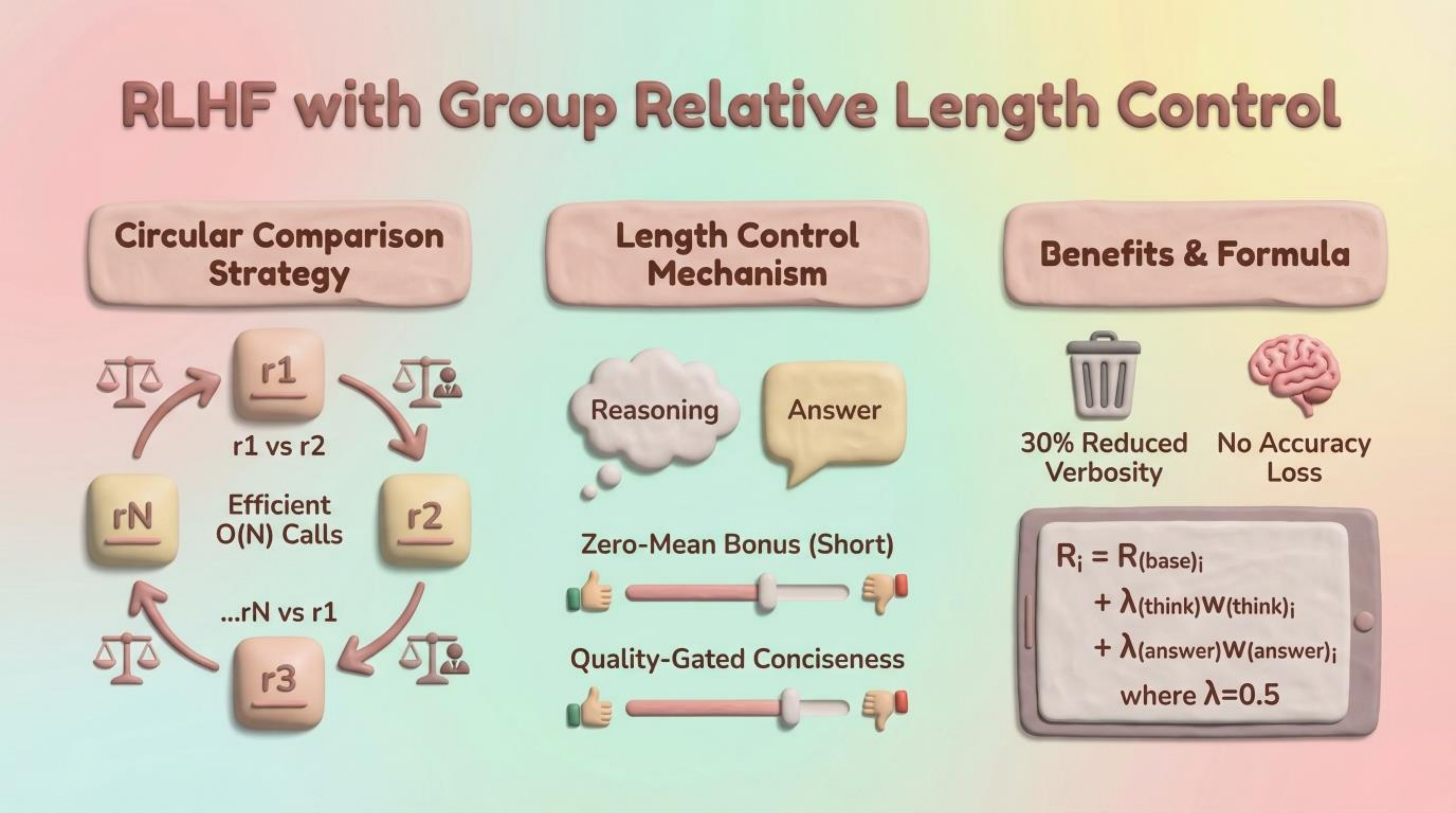
- Synchronous GRPO with masked importance sampling
- 128 Prompts/Step, 16 Gens/Prompt, Batch Size 2048 — 16
- On-Policy Updates, MoE Router Weights Frozen, Aux-loss-free Load Balancing
- Maximum Generation Length 49K, Overlong Filtering Enabled

Performance Benchmarks

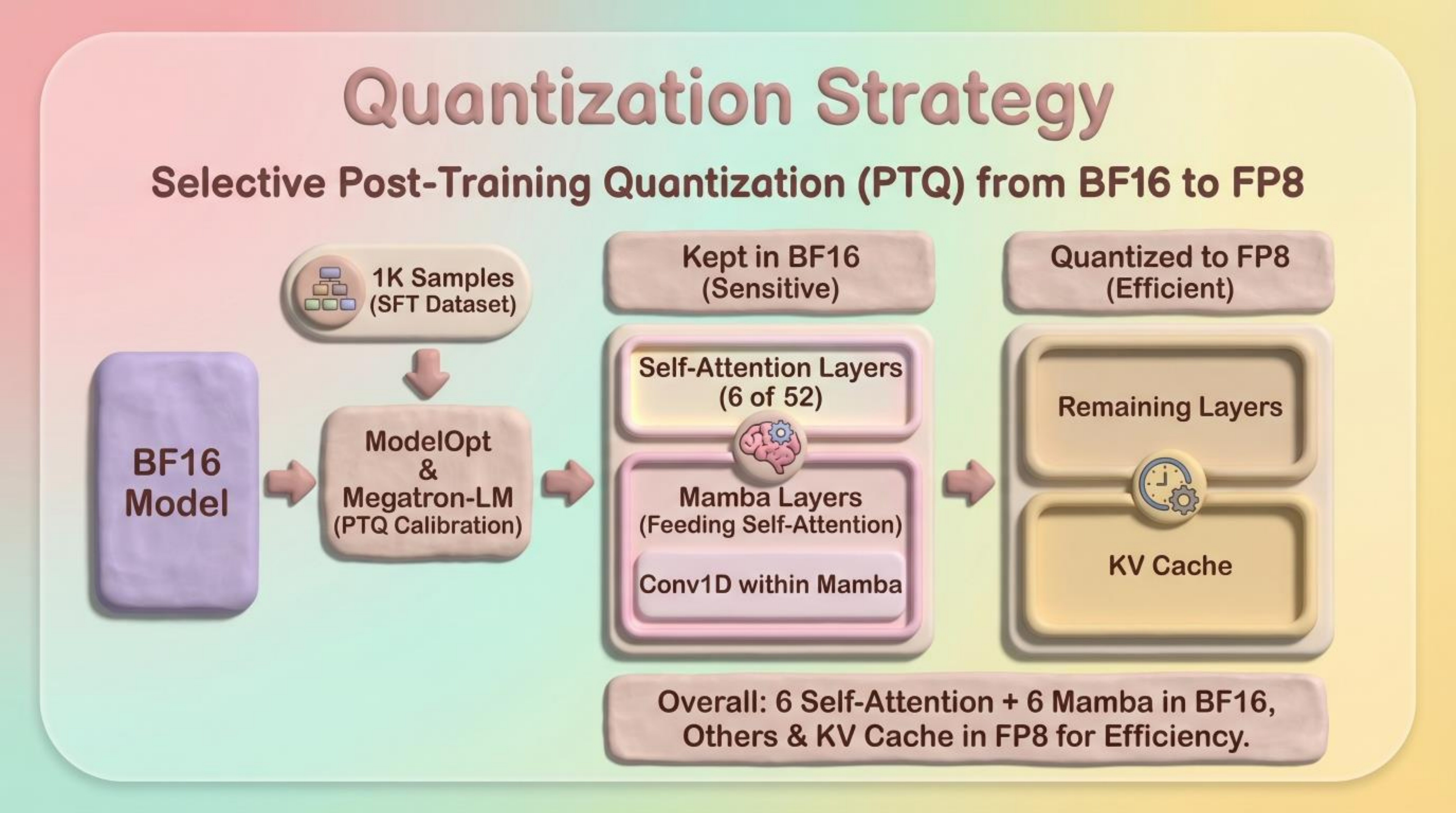





Generative Reward Model Training
GenRM Reasoning Flow & Scoring
Conversation History
User Request
Candidate Response 1
Candidate Response 2
Qwen3-235B-A22B-Thinking-2507 (GRPO)
Helpfulness Scores (1-5)
1
2
3
4
6
Ranking Score (1-6)
Reasons over inputs to generate helpfulness scores and a final ranking.
Reward Function & Training Config
$R = -C_1 * I_{format} - |P_{h1}-G_{h1}| - |P_{h2}-G_{h2}| - C_2 * |P_r-G_r|$
$C_1 = 10$
$C_2 = 1$
128 prompts/batch
8 generations
One Gradient Step
Reward function guides training with specific batch and generation settings.
Data Sources & Performance Gains
HelpSteer3 Data
Synthetic Safety Blend
RM-Bench
JudgeBench
Internal-Val-Set
Data from HelpSteer3 and synthetic sources drive significant performance improvements across benchmarks.


RLHF with Group Relative Length Control
Circular Comparison Strategy
r1
r1 vs r2
r2
Efficient O(N) Calls
rN
...rN vs r1
r3
Length Control Mechanism
Reasoning
Answer
Zero-Mean Bonus (Short)
Quality-Gated Conciseness
Benefits & Formula
30% Reduced Verbosity
No Accuracy Loss
$R_i = R_{(base)_i} + \lambda_{(think)} w_{(think)_i} + \lambda_{(answer)} w_{(answer)_i}$
where $\lambda=0.5$

Post-Training Evaluation Results
General Knowledge (MMLU-Pro)
78.30
66.00
57.8
Nemotron 3 Nano
Qwen3
GPT-OSS 20B
Reasoning
80.90
33.00
10.90
75.00
34.00
10.90
73.04
38.00
54.80
AIME25
GPQA
LiveCodeBench
Agentic
48.00
22.00
5.00
38.76
10.00
12.05
TauBench V2 avg
SWE-Bench
Chat & Instruction Following
67.65
48.55
51.00
72.10
25.90
43.03
Arena-Hard-V2 Avg
IFBench
Long Context
59.50
N/A
34.00
RULER @ 1M
Multilingual
86.20
65.00
77.60
MMLU-ProX
Nemotron 3 Nano
Qwen3
GPT-OSS 20B


Quantization Strategy
Selective Post-Training Quantization (PTQ) from BF16 to FP8
1K Samples (SFT Dataset)
BF16 Model
ModelOpt & Megatron-LM (PTQ Calibration)
Kept in BF16 (Sensitive)
Self-Attention Layers (6 of 52)
Mamba Layers (Feeding Self-Attention)
Conv1D within Mamba
Quantized to FP8 (Efficient)
Remaining Layers
KV Cache
Overall: 6 Self-Attention + 6 Mamba in BF16, Others & KV Cache in FP8 for Efficiency.

Quantization Results & Trade-offs

99%

Median Accuracy Recovery vs BF16

Benchmark	%	FP8 BF16
MMLU-Pro	78.30	78.30 vs 77.48
AIME25 no tools	89.06	89.06 vs 87.71
AIME25 with tools	99.17	99.17 vs 98.80
GPQA no tools	73.04	73.04 vs 72.47
GPQA with tools	75.00	75.00 vs 73.40
LiveCodeBench	68.25	68.25 vs 67.62
TauBench average	48.00	48.00 vs 44.79
IFBench	35.85	35.85 vs 36.06
AA-LCR	59.50	59.50 vs 59.63
MMLU-ProX	78.10	78.10 vs 77.48

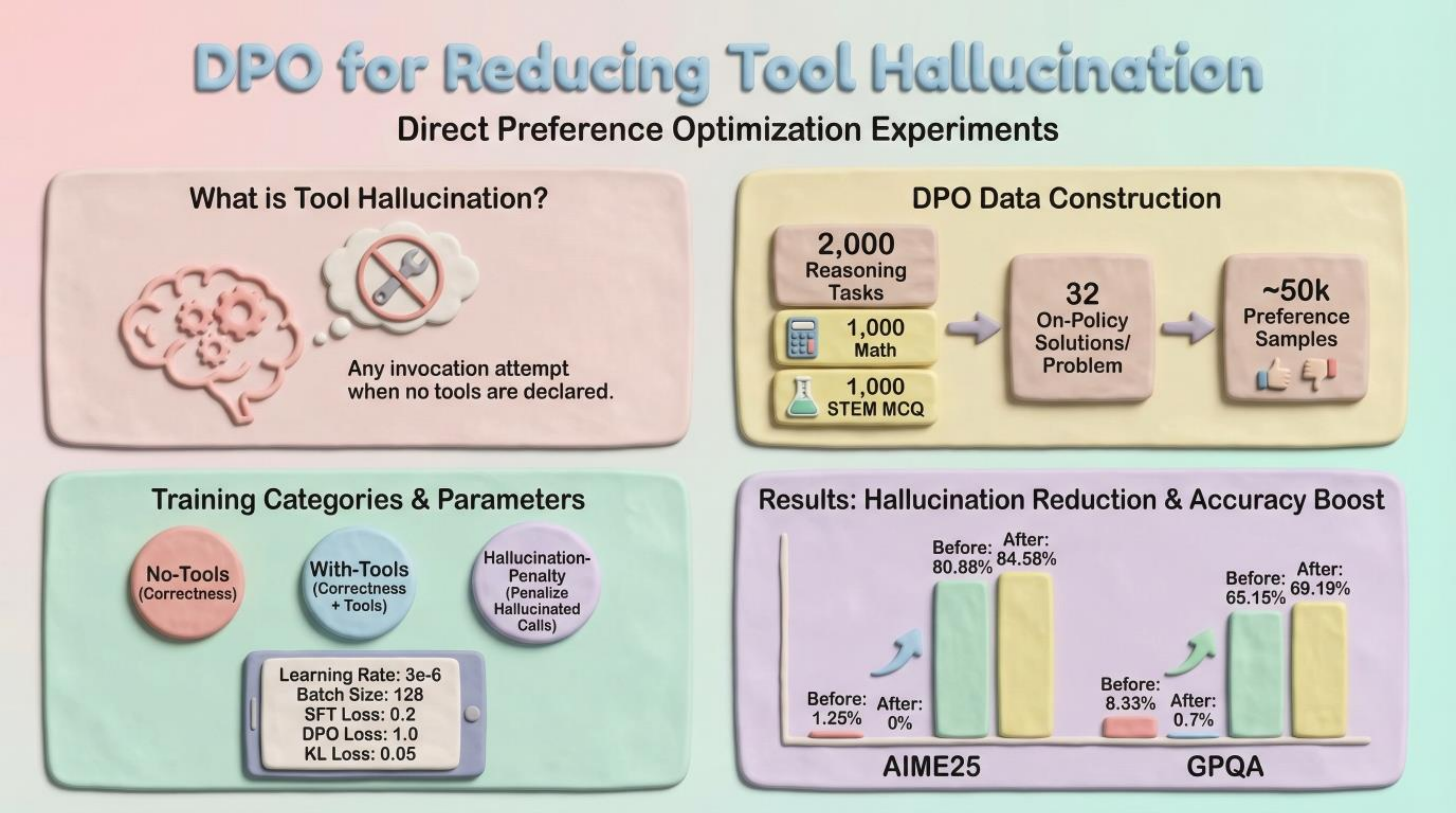
Throughput Improvement

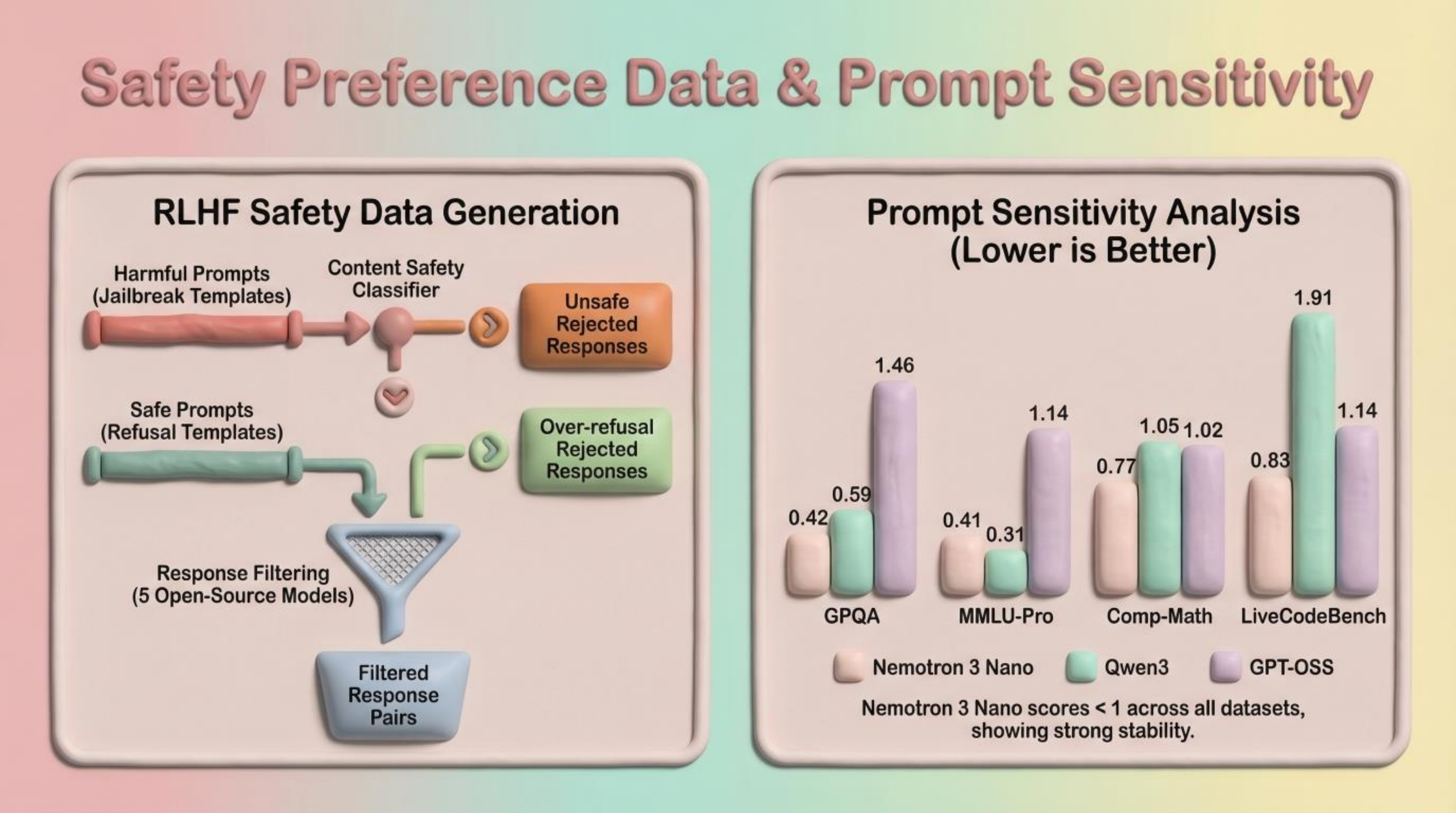
KV Cache FP8 Quantization

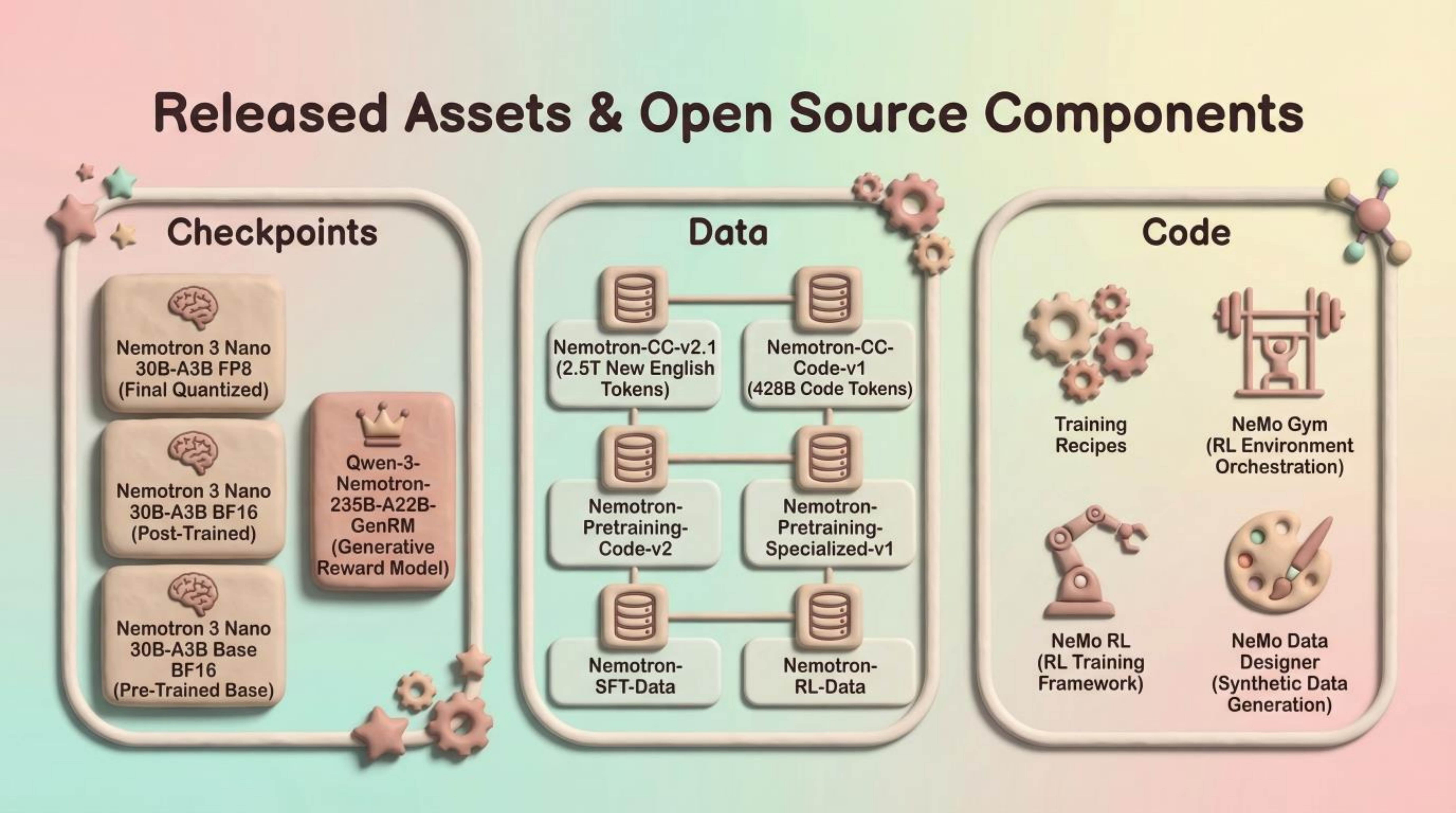
Significant Throughput Gains

Ablation Study Visualization

Visualization of Different Quantization Configurations & Results


DPO for Reducing Tool Hallucination
Direct Preference Optimization Experiments
What is Tool Hallucination?
Any invocation attempt when no tools are declared.
DPO Data Construction
2,000 Reasoning Tasks
1,000 Math
1,000 STEM MCQ
32 On-Policy Solutions/ Problem
~50k Preference Samples
Training Categories & Parameters
No-Tools (Correctness)
With-Tools (Correctness + Tools)
Hallucination-Penalty (Penalize Hallucinated Calls)
Learning Rate: 3e-6
Batch Size: 128
SFT Loss: 0.2
DPO Loss: 1.0
KL Loss: 0.05
Results: Hallucination Reduction & Accuracy Boost
Before: 1.25%
After: 0%
Before: 80.88%
After: 84.58%
AIME25
Before: 8.33%
After: 0.7%
Before: 65.15%
After: 69.19%
GPQA


Safety Preference Data & Prompt Sensitivity
RLHF Safety Data Generation
Harmful Prompts (Jailbreak Templates)
Content Safety Classifier
Unsafe Rejected Responses
Safe Prompts (Refusal Templates)
Over-refusal Rejected Responses
Response Filtering (5 Open-Source Models)
Filtered Response Pairs
Prompt Sensitivity Analysis (Lower is Better)
1.91
1.46
1.14
1.05
1.02
1.14
0.77
0.83
0.59
0.42
0.41
0.31
GPQA
MMLU-Pro
Comp-Math
LiveCodeBench
Nemotron 3 Nano
Qwen3
GPT-OSS
Nemotron 3 Nano scores < 1 across all datasets, showing strong stability.


Released Assets & Open Source Components
Checkpoints
Nemotron 3 Nano 30B-A3B FP8 (Final Quantized)
Nemotron 3 Nano 30B-A3B BF16 (Post-Trained)
Nemotron 3 Nano 30B-A3B Base BF16 (Pre-Trained Base)
Qwen-3-Nemotron-235B-A22B-GenRM (Generative Reward Model)
Data
Nemotron-CC-v2.1 (2.5T New English Tokens)
Nemotron-CC-Code-v1 (428B Code Tokens)
Nemotron-Pretraining-Code-v2
Nemotron-Pretraining-Specialized-v1
Nemotron-SFT-Data
Nemotron-RL-Data
Code
Training Recipes
NeMo Gym (RL Environment Orchestration)
NeMo RL (RL Training Framework)
NeMo Data Designer (Synthetic Data Generation)

Conclusion & Key Contributions

Nemotron 3 Nano Overview

Open, Efficient MoE Hybrid Mamba-Transformer for agentic reasoning

Better/on-par accuracy with up to 3.3x higher inference throughput.

Supports 1M context length.

Key Innovations: Architecture

Granular MoE architecture with shared experts.

Two-phase pretraining on 25T tokens (diverse then high-quality data).

Key Innovations: Training & Optimization

Multi-environment RLVR training simultaneously.

GenRM-based RLHF with length control.

Selective FP8 quantization preserving 99% accuracy.

Released Assets & Impact

All model weights, training recipes, data, and code released on HuggingFace.

Significant advancement in efficient yet capable language models for agentic applications.