FRONT-LOADING REASONING:

The Synergy between Pretraining and Post-Training Data

A Comprehensive University Lecture on
Data Allocation Strategies and LLM Reasoning Development

RESEARCH OVERVIEW & MOTIVATION

CORE RESEARCH QUESTIONS

- **ADDING REASONING DATA:**
  PRETRAINING VS. POST-TRAINING?
  (Better?)

- **EARLY INCLUSION:**
  OVERFITTING RISKS OR DURABLE FOUNDATIONS?

KEY FINDINGS SUMMARY

- **FRONT-LOADING CRITICAL:**
  +19% GAIN
  (vs. Baseline)

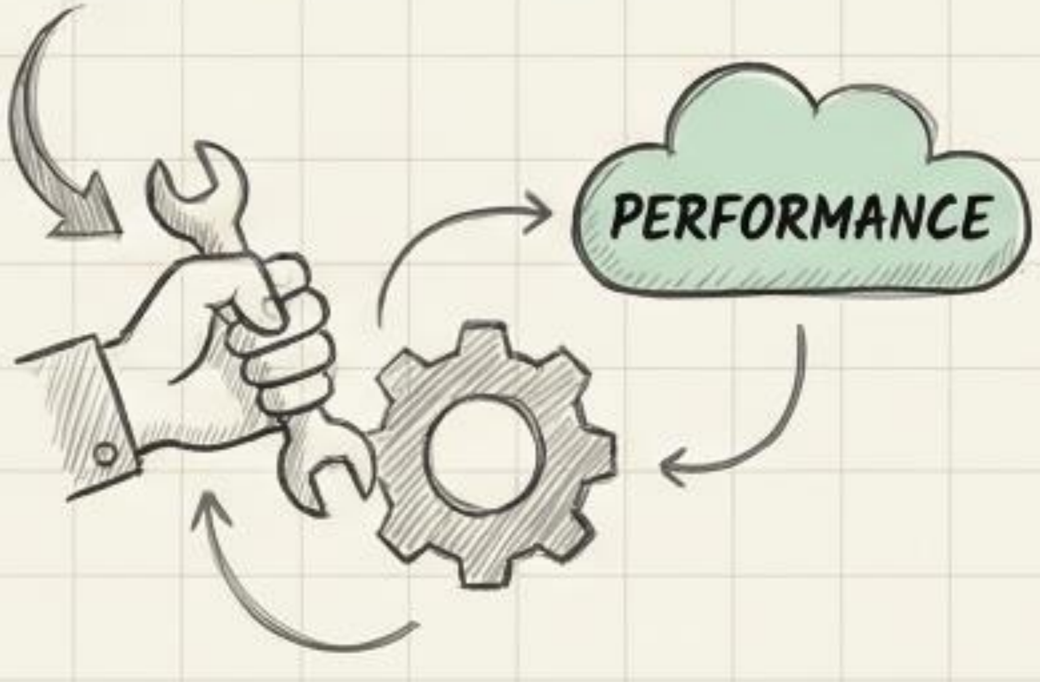
- **ASYMMETRIC PRINCIPLE:**
  DIVERSITY IN PRETRAINING, QUALITY IN SFT

- **LATENT EFFECTS OF HIGH-QUALITY DATA**

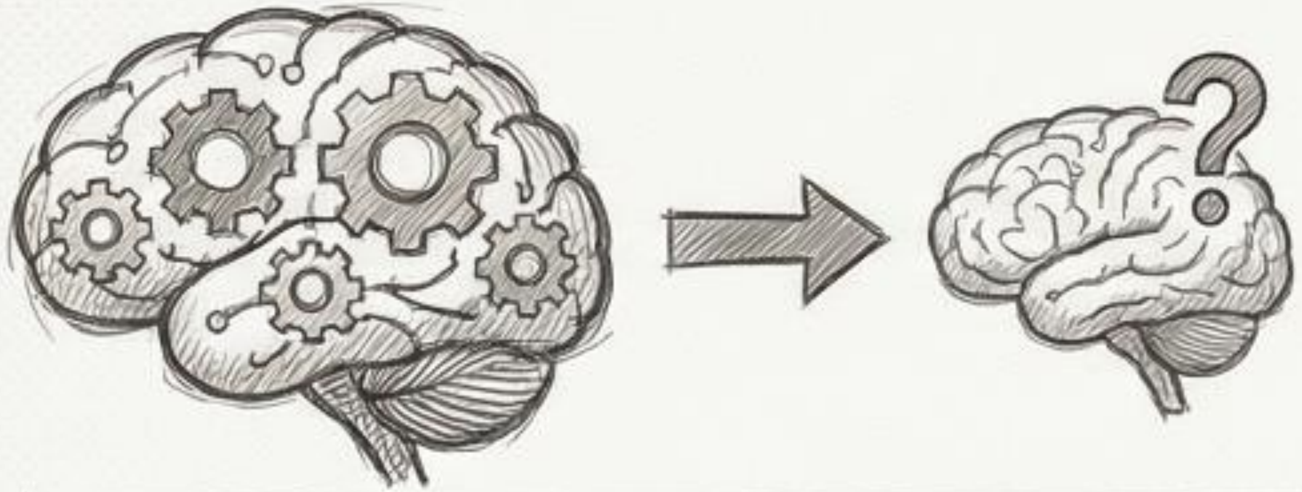
OPTIMIZATION PROBLEM

$$\boxed{\mathrm{MAX}\ P\,(D_{PT\_RES},\ D_{SFT\_RES})}$$

BALANCING DATA ALLOCATION & QUALITY

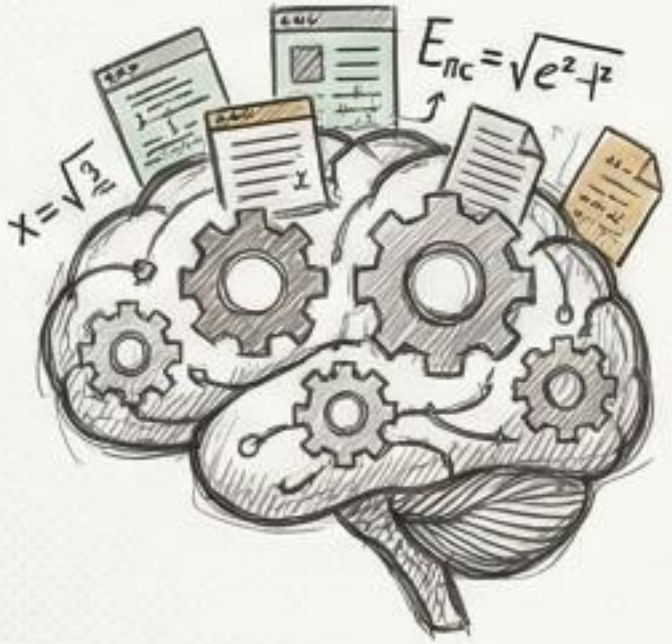


KEY RESEARCH QUESTIONS



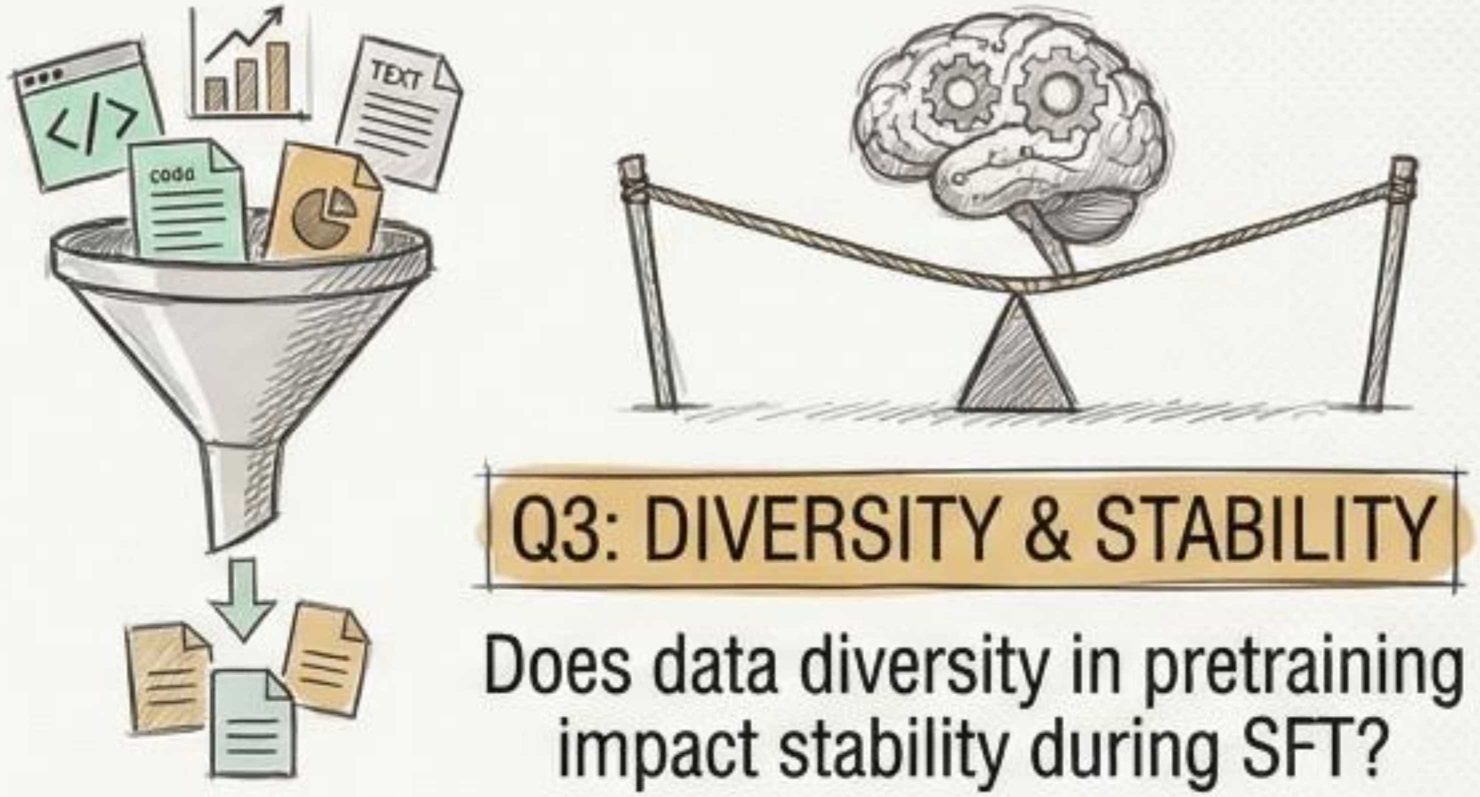
Q1: REASONING-RICH PRETRAINING

Is reasoning-rich pretraining essential, or can models catch up later?



Q2: OVERFITTING & GENERALIZATION

Does reasoning data make base LLMs overfitted and less generalizable?

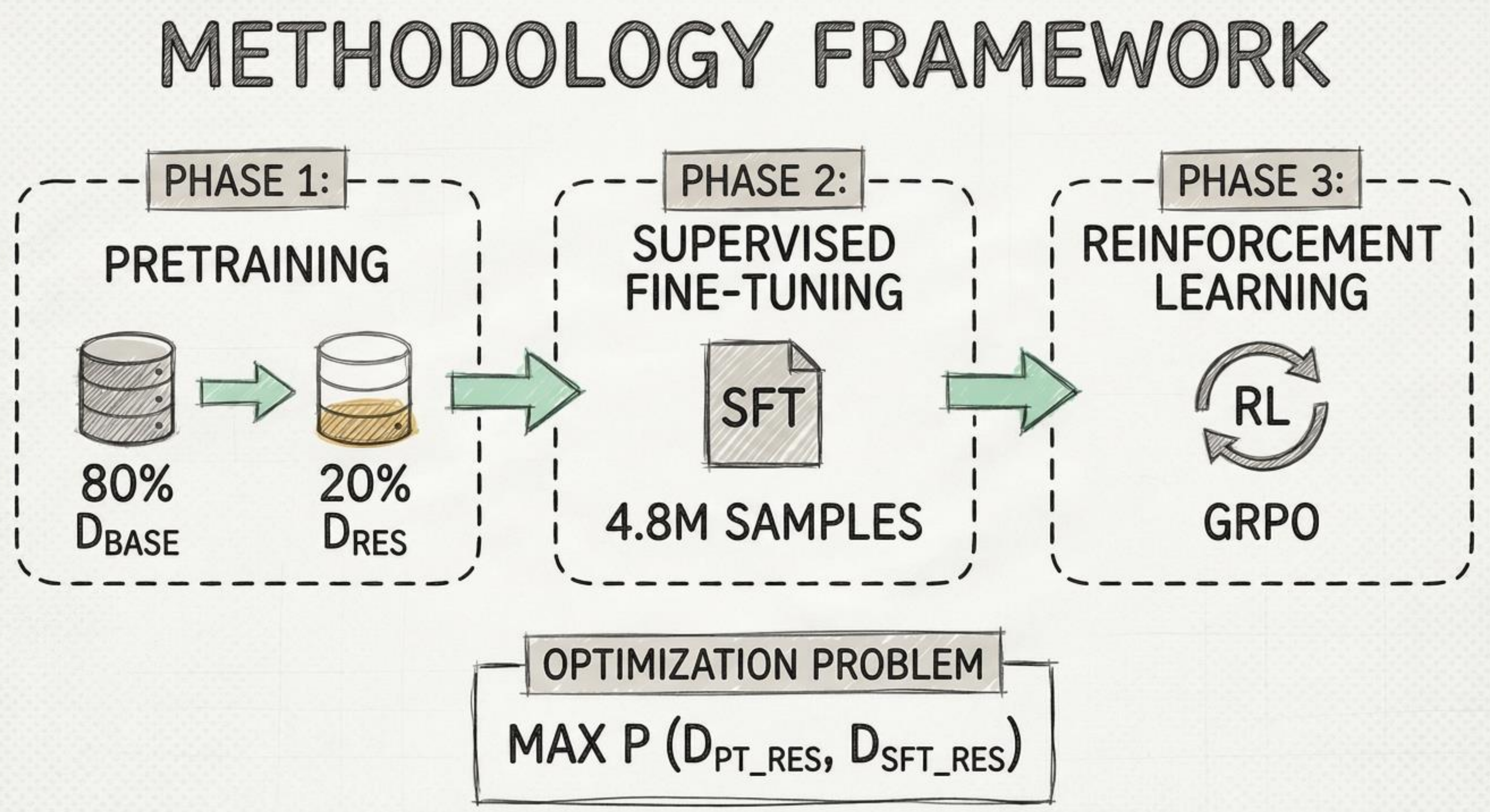


Q3: DIVERSITY & STABILITY

Does data diversity in pretraining impact stability during SFT?

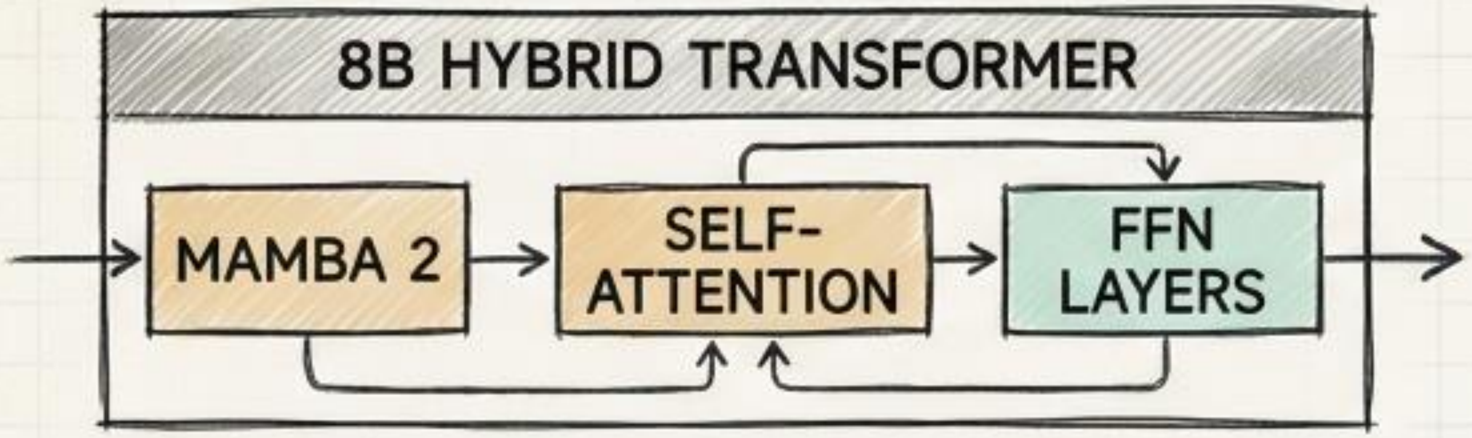
Q4: COMPLEXITY & QUALITY

Does complexity and quality of reasoning data matter during pretraining?


METHODOLOGY FRAMEWORK
PHASE 1:
PRETRAINING
80%
D_{BASE}
20%
D_{RES}
PHASE 2:
SUPERVISED FINE-TUNING
SFT
4.8M SAMPLES
PHASE 3:
REINFORCEMENT LEARNING
RL
GRPO
OPTIMIZATION PROBLEM
MAX P (D_{PT_RES}, D_{SFT_RES})

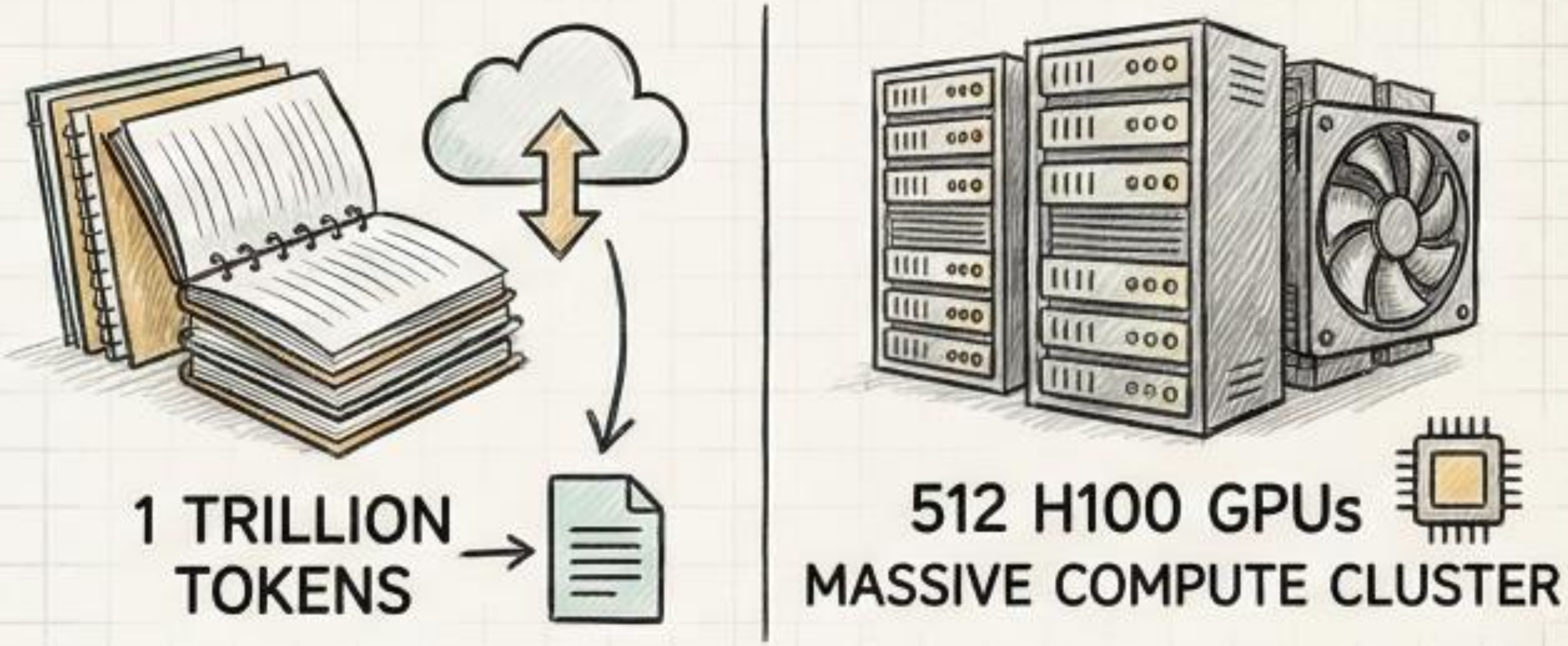
MODEL ARCHITECTURE & TRAINING SETUP

PANEL 1: HYBRID ARCHITECTURE OVERVIEW

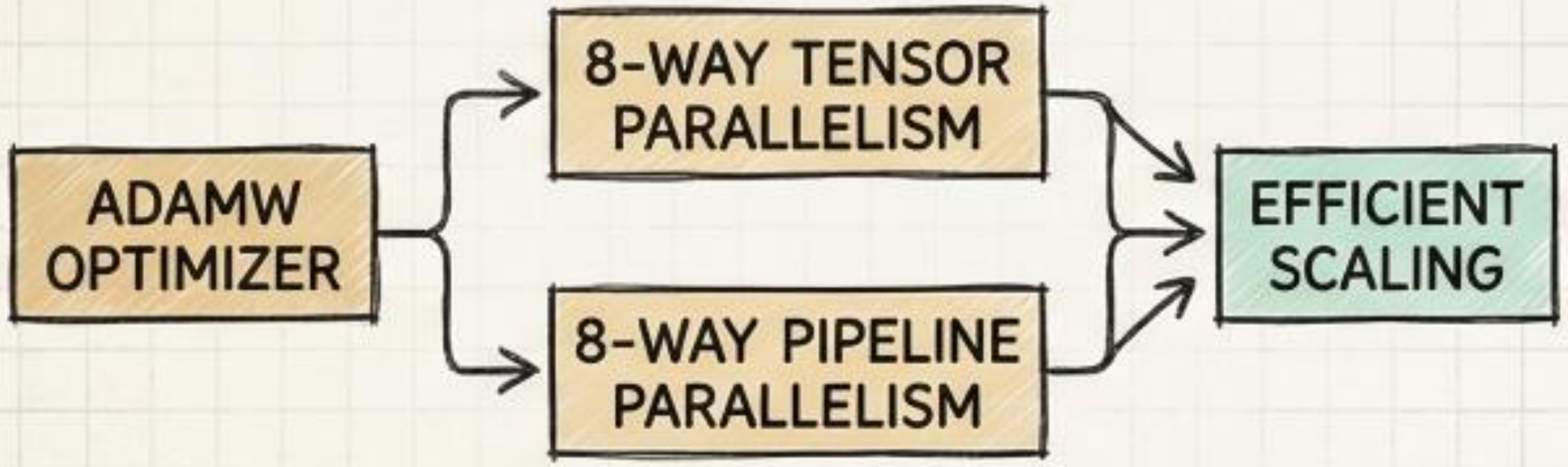


- Integrates Mamba 2 for efficiency
- Self-attention for global context
- FFN for refinement

PANEL 2: PRETRAINING SCALE & HARDWARE



PANEL 3: OPTIMIZATION & PARALLELISM



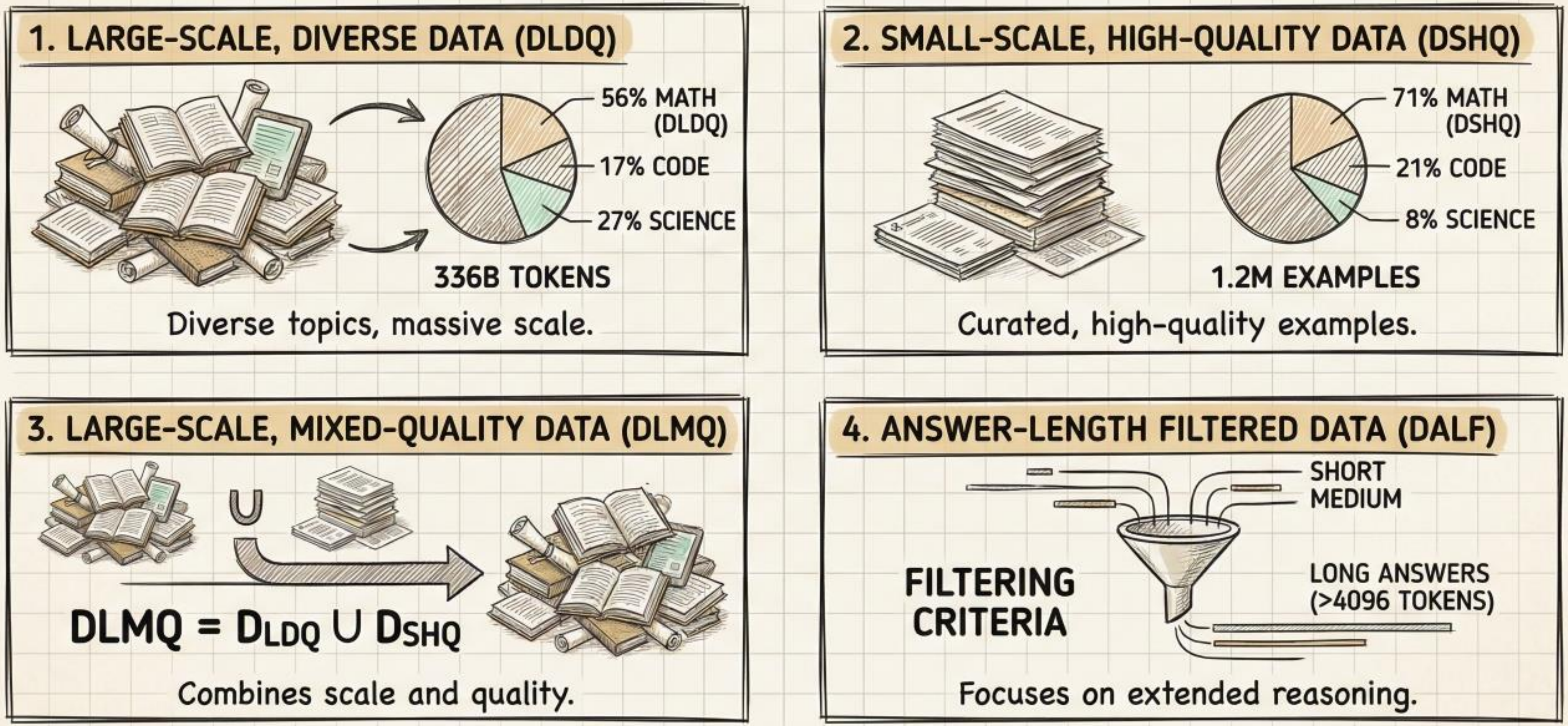
Distributed training across GPUs
Optimized gradient updates

PANEL 4: HYPERPARAMETERS & CONTEXT

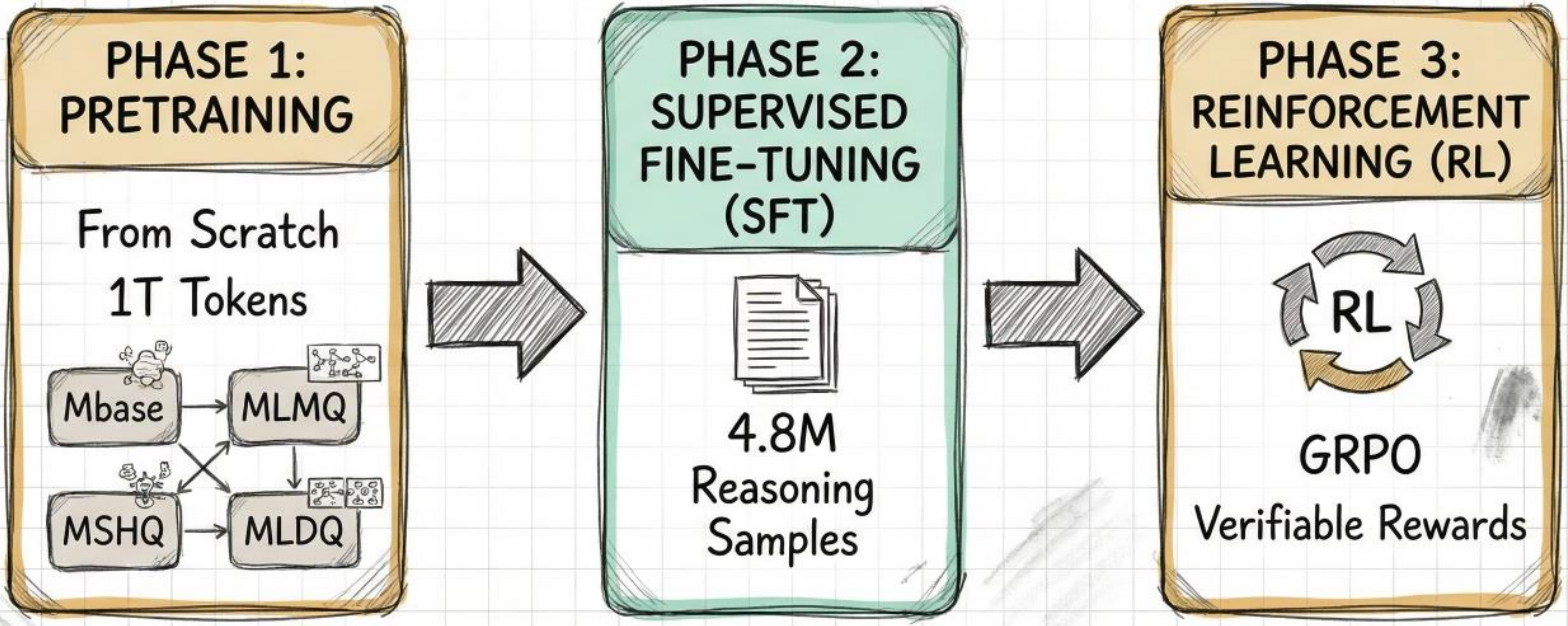
PARAMETER	VALUE
LEARNING RATE	3e-4 to 3e-6 (Scheduled)
BATCH SIZE	6M TOKENS (Global)
CONTEXT LENGTH	8192 TOKENS

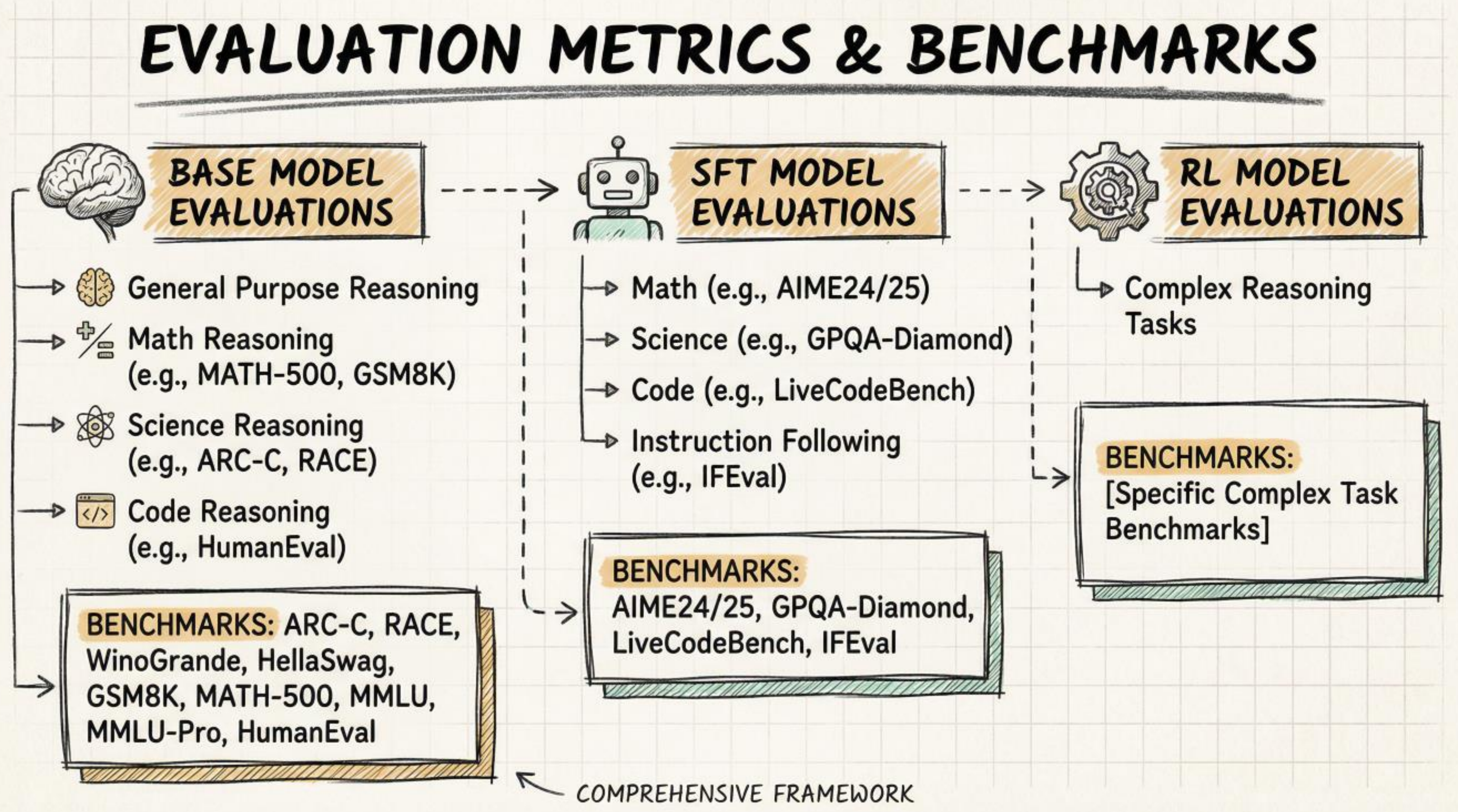
Systematic tuning for stable training

DATA PIPELINE: REASONING DATASETS

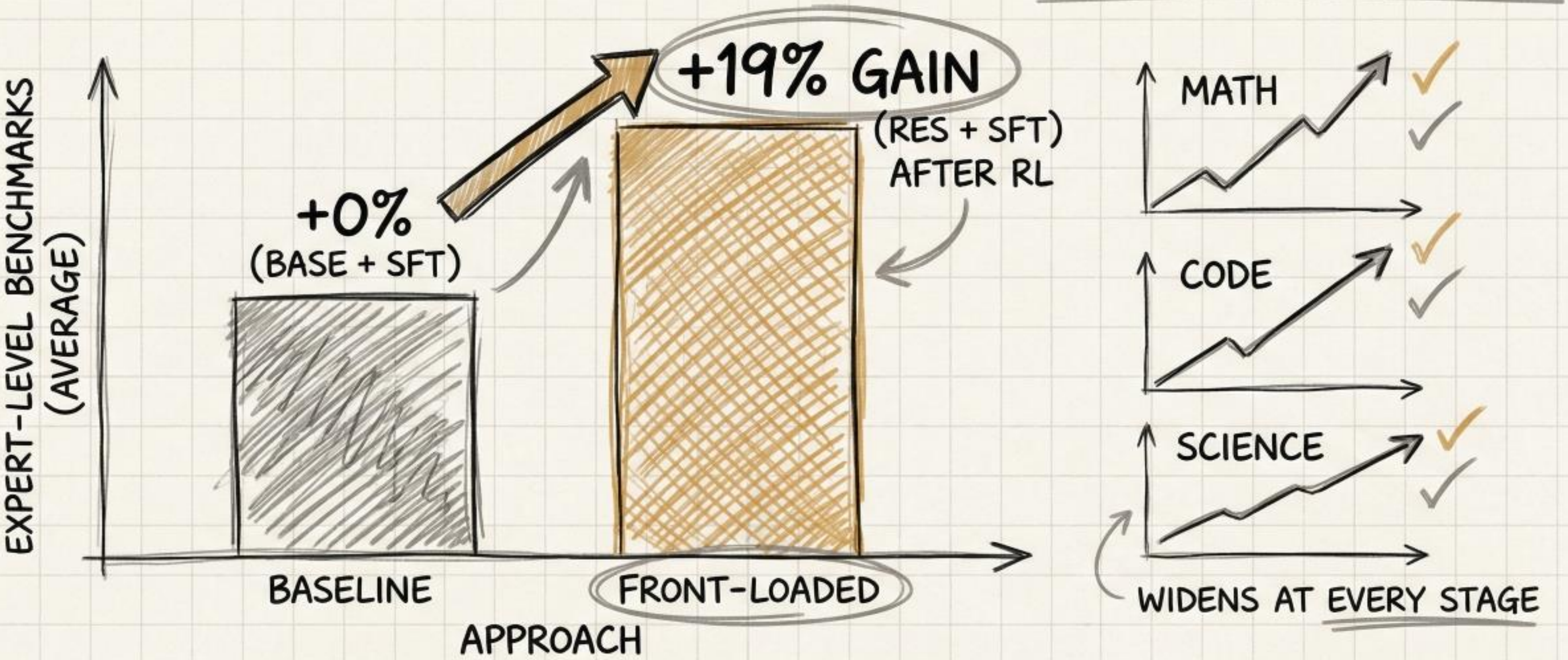


EXPERIMENTAL DESIGN: THREE TRAINING PHASES

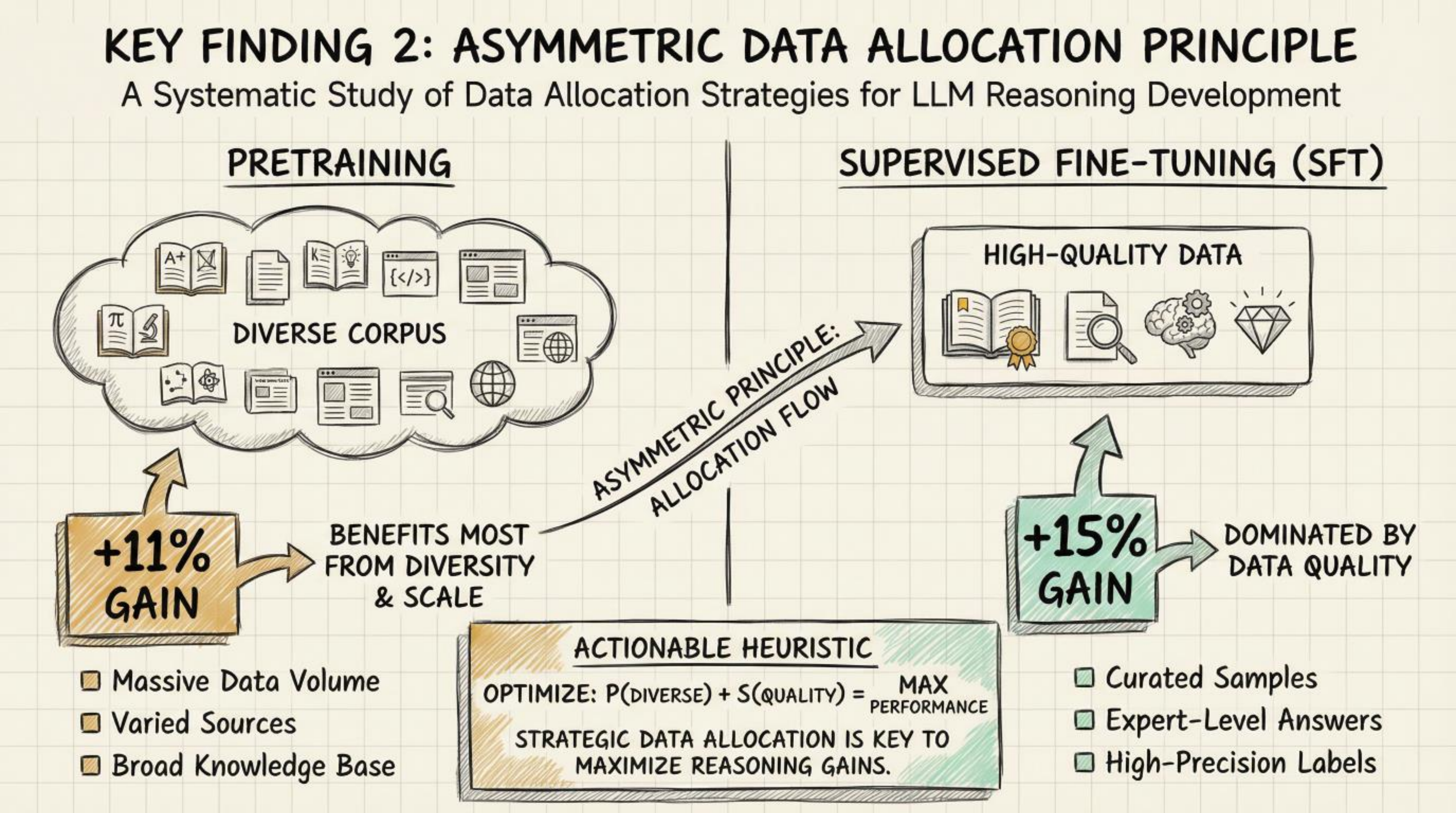



EVALUATION METRICS & BENCHMARKS
BASE MODEL EVALUATIONS
General Purpose Reasoning
Math Reasoning (e.g., MATH-500, GSM8K)
Science Reasoning (e.g., ARC-C, RACE)
Code Reasoning (e.g., HumanEval)
BENCHMARKS: ARC-C, RACE, WinoGrande, HellaSwag, GSM8K, MATH-500, MMLU, MMLU-Pro, HumanEval
SFT MODEL EVALUATIONS
Math (e.g., AIME24/25)
Science (e.g., GPQA-Diamond)
Code (e.g., LiveCodeBench)
Instruction Following (e.g., IFEval)
BENCHMARKS: AIME24/25, GPQA-Diamond, LiveCodeBench, IFEval
RL MODEL EVALUATIONS
Complex Reasoning Tasks
BENCHMARKS: [Specific Complex Task Benchmarks]
COMPREHENSIVE FRAMEWORK

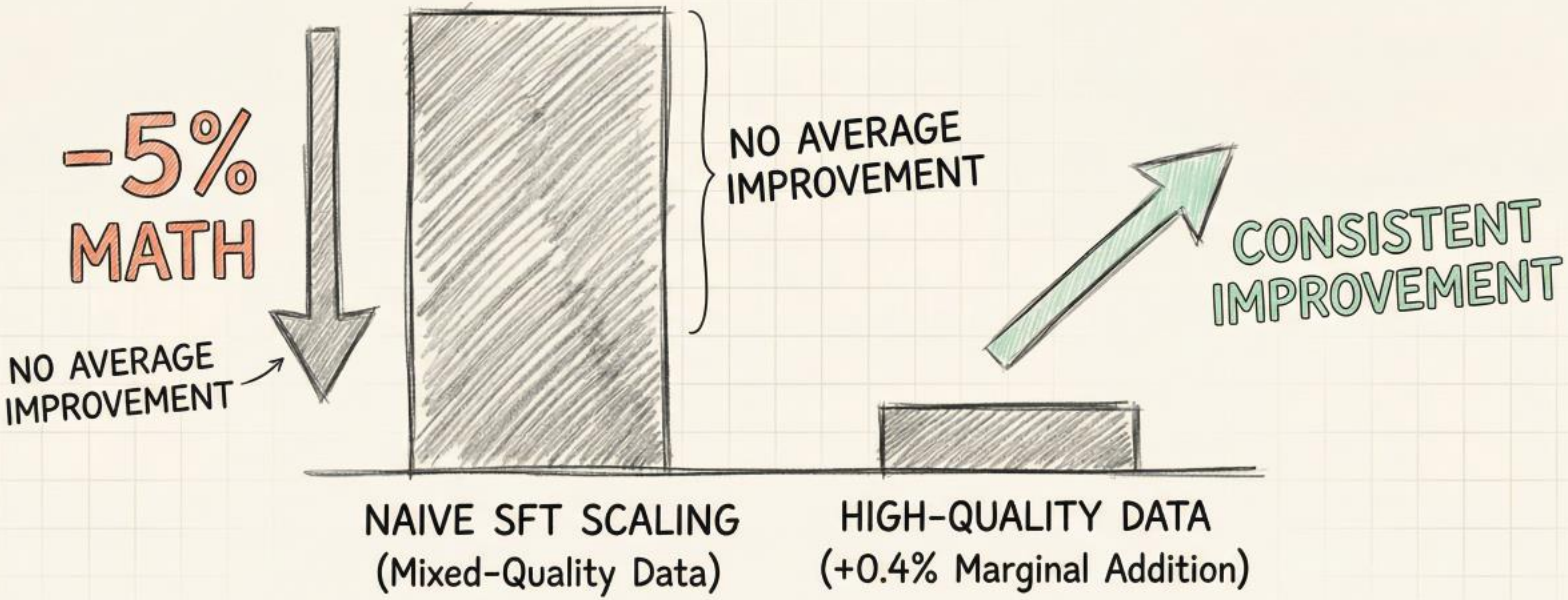
KEY FINDING 1: FRONT-LOADING CREATES DURABLE ADVANTAGE



CORE INSIGHT: Injecting reasoning data during PRETRAINING establishes a SUPERIOR FOUNDATION. Widely refutes catch-up & overfitting hypotheses. SFT CANNOT COMPENSATE for weak foundations.

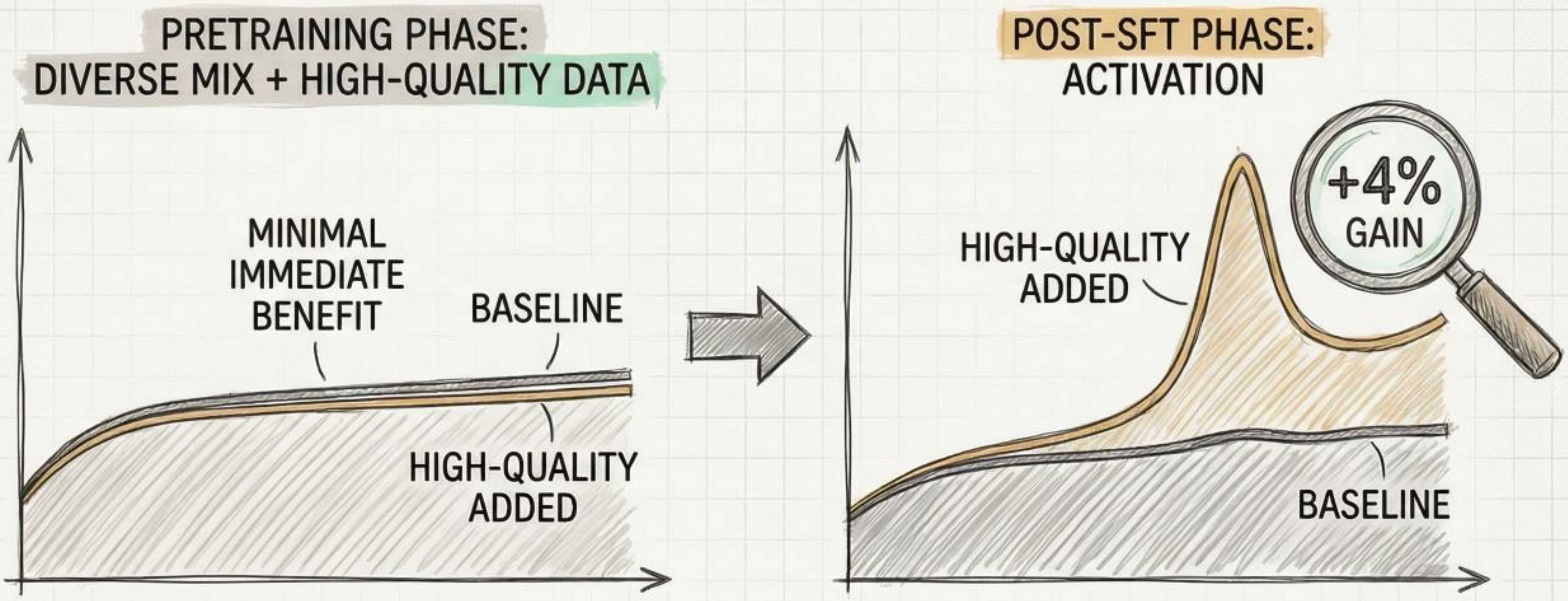

KEY FINDING 2: ASYMMETRIC DATA ALLOCATION PRINCIPLE
A Systematic Study of Data Allocation Strategies for LLM Reasoning Development
PRETRAINING
DIVERSE CORPUS
A+
{</>}
+11% GAIN
BENEFITS MOST FROM DIVERSITY & SCALE
Massive Data Volume
Varied Sources
Broad Knowledge Base
ASYMMETRIC PRINCIPLE: ALLOCATION FLOW
SUPERVISED FINE-TUNING (SFT)
HIGH-QUALITY DATA
+15% GAIN
DOMINATED BY DATA QUALITY
Curated Samples
Expert-Level Answers
High-Precision Labels
ACTIONABLE HEURISTIC
OPTIMIZE: P(DIVERSE) + S(QUALITY) = MAX PERFORMANCE
STRATEGIC DATA ALLOCATION IS KEY TO MAXIMIZE REASONING GAINS.

KEY FINDING 3: NAIVE SFT SCALING IS HARMFUL



- BLIND SCALING ACTIVELY HARMS MATHEMATICAL REASONING.
- SMALL ADDITION OF HIGH-QUALITY DATA YIELDS CONSISTENT GAINS.

KEY FINDING 4: HIGH-QUALITY DATA HAS LATENT EFFECTS



LATENT EFFECT DISCOVERY: High-quality data added to pretraining shows little immediate gain, but unlocks a +4% advantage after SFT. This reveals a deeper synergy where pretraining instills latent potential activated during alignment.

BASE MODEL RESULTS: IMMEDIATE FOUNDATIONAL GAINS

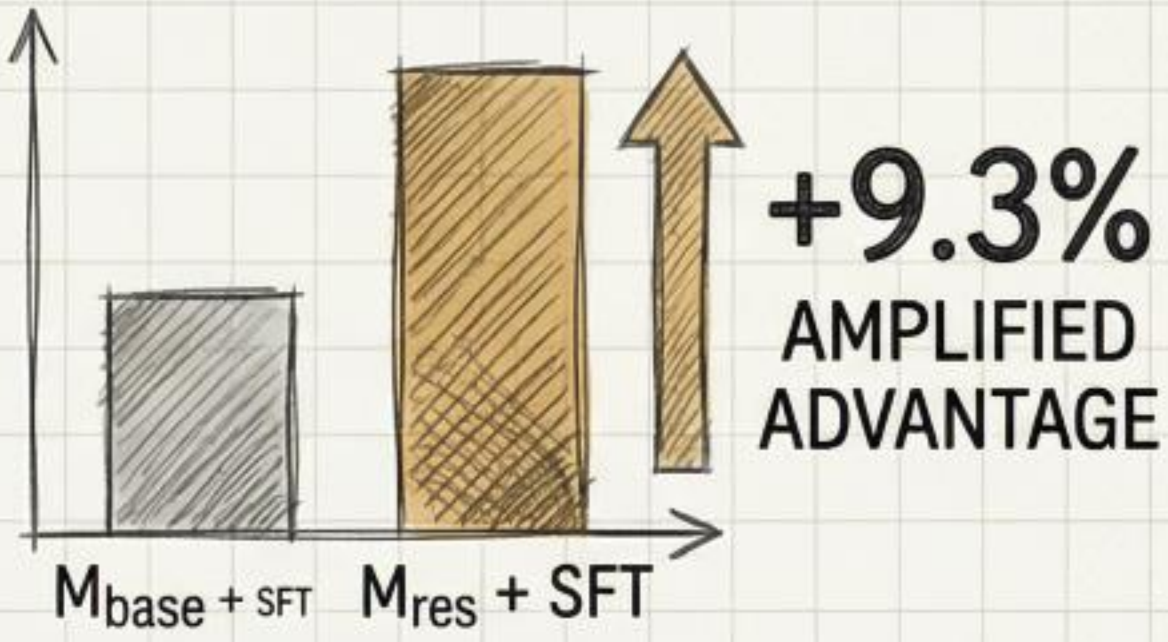
MODEL	MATH (ACC)	CODE (ACC)	AVERAGE
$M_{baseline}$	38.2%	21.5%	29.8%
MLDQ	66.6% (+28.4% GAIN	30.5% (+9%	48.5%
M...[Other]	60.1%	28.2%	44.1%
M...[Other]	55.8%	25.9%	40.8%
M...[Other]	51.3%	23.7%	37.5%

- Every model exposed to reasoning data surpasses $M_{baseline}$.
- MLDQ achieves highest average (+28.4% Math, +9% Code).
- Scale and diversity are more critical than curated quality at this early stage.

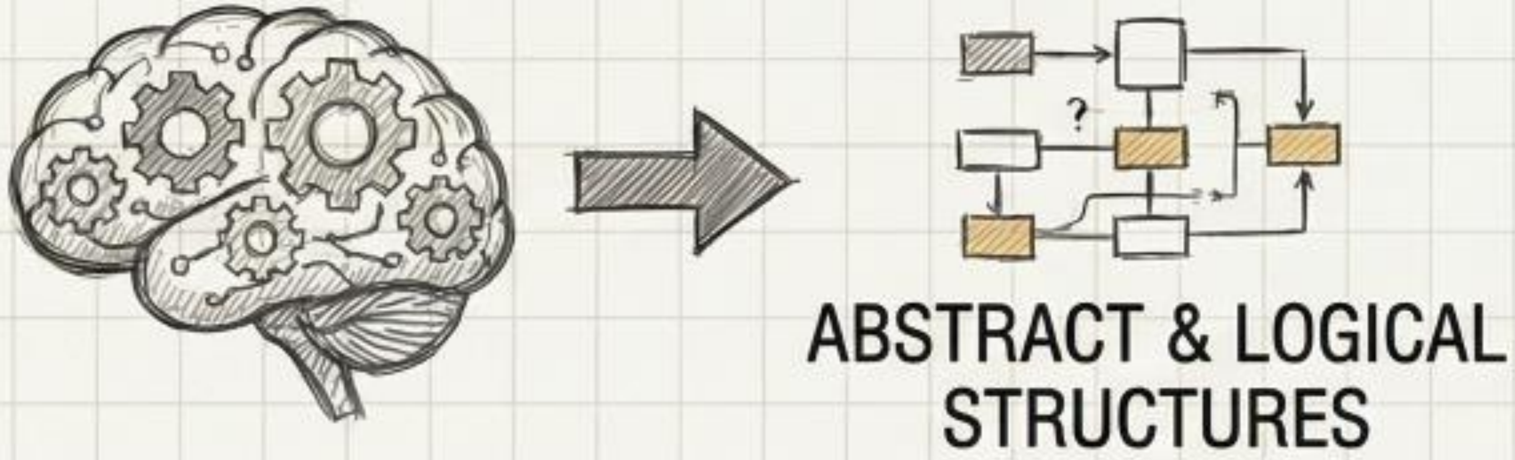
SFT RESULTS: ADVANTAGE AMPLIFIED

TABLE 2: PERFORMANCE COMPARISON (SFT MODELS)

	M_{base} + SFT	M_{res} + SFT
M_{base} + SFT	87.7	86.6
M_{res} + SFT	87.4	82.7



DOMAIN-SPECIFIC IMPACT: SCIENCE



- Largest Disparity in Science.
- Pretraining develops effective internal representations.

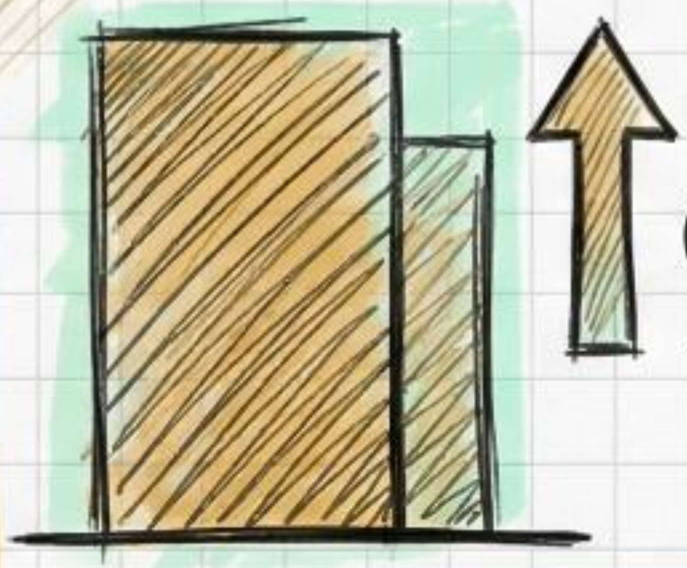
CATCH-UP HYPOTHESIS REFUTED: Advantage is Durable & Amplified.

RL Results: Pretraining Dictates Final Accuracy

BASELINE
(Mbase + SFT)

Standard performance.

FRONT-LOADED
(MLMQ + SFT + RL)



+19% AVG. GAIN
(EXPERT BENCHMARKS)

+18.57% LEAD OVER MBASE ON AVERAGE

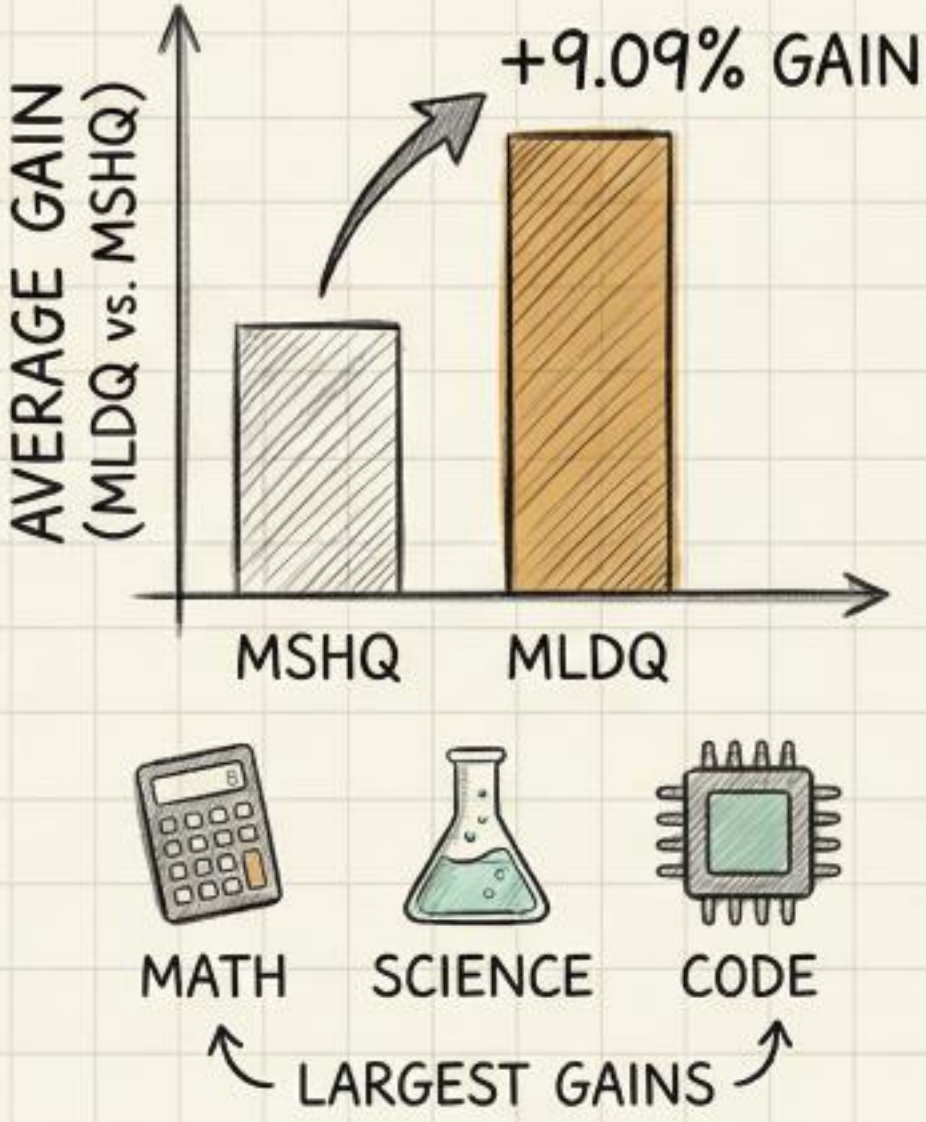
KEY INSIGHTS: COMPOUNDING RETURNS

- ☑ **+39.32%** IMPROVEMENT ON AIME COMPETITION MATH PROBLEMS (vs Baseline)
- ☑ CONCLUSIVE EVIDENCE OF COMPOUNDING RETURNS

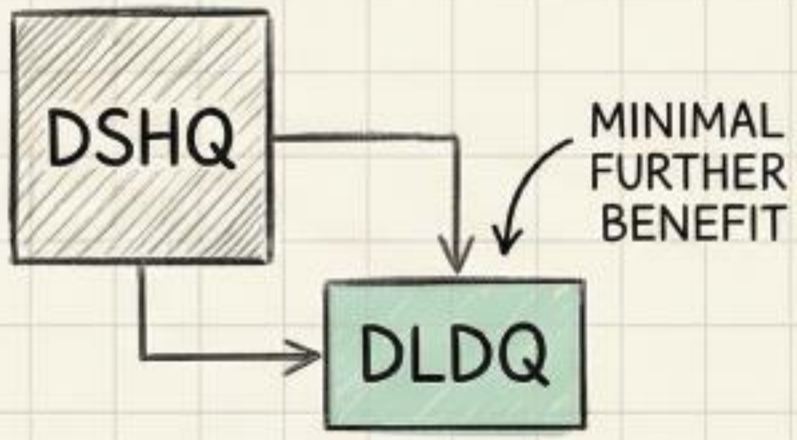
ABLATIONS: DOES SCALE & DIVERSITY MATTER IN PRETRAINING?

INCREASING **SIZE** & **DIVERSITY** of Dres in PRETRAINING shows **INCREAMING EKTKINED FROM THE** interraator & OOJECTS SIGNIFICANT IMPROVEMENT.

1.



2.

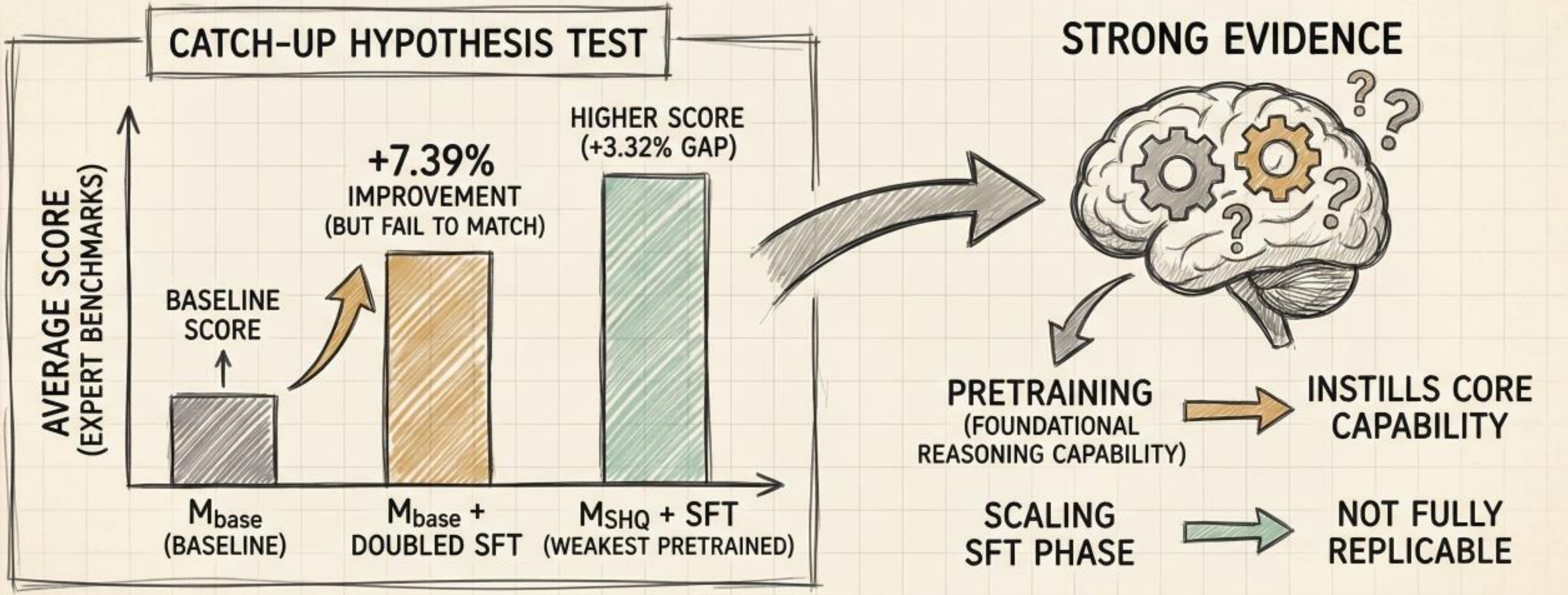


- SCALING DLDQ with DSHQ provides MINIMAL FURTHER BENEFIT.

3. **BROAD EXPOSURE** to **DIVERSE REASONING PATTERNS** is IMPACTFUL for BUILDING STRONG FOUNDATION.

ABLATIONS: PRETRAINING ADVANTAGE RESISTS CATCH-UP

A Systematic Study of Data Allocation Strategies for LLM Reasoning Development

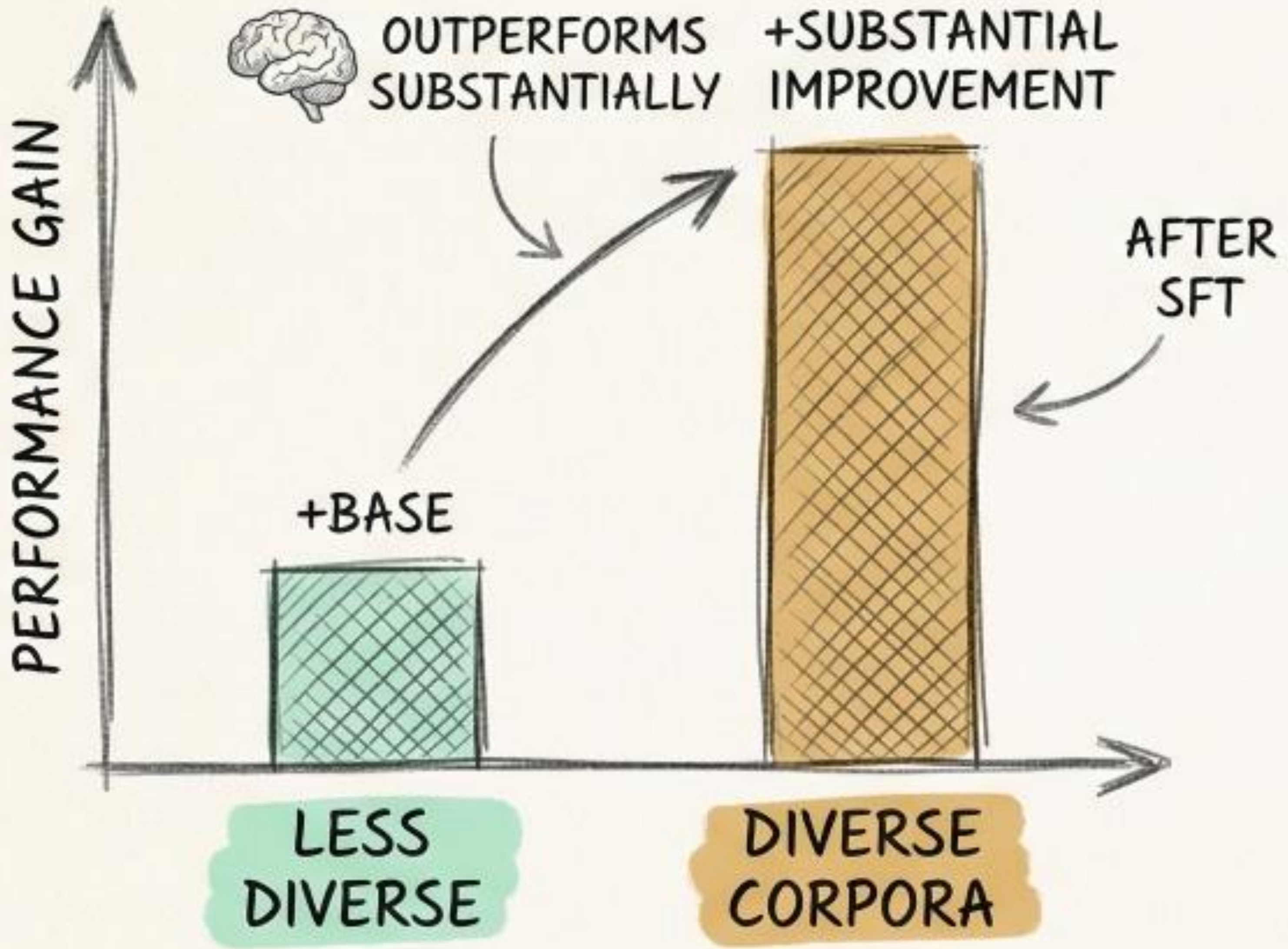


Doubling SFT for M_{base} improves score by 7.39% but fails to match even the weakest reasoning-pretrained model (+3.32%).

Pretraining instills foundational reasoning capability that cannot be fully replicated by simply scaling SFT phase.

ABLATIONS: POST-SFT HIGH-QUALITY DATA REVEALS LATENT VALUE

POST-SFT WITH SAME HIGH-QUALITY RECIPE (DSHQ)



SURPRISING MLMQ ADVANTAGE

HIGH-QUALITY BUT LESS DIVERSE DATA ACTS AS COMPLEMENTARY AMPLIFIER WHOSE BENEFITS EMERGE AFTER ALIGNMENT.

ABLATIONS: SFT IS DOMINATED BY DATA QUALITY

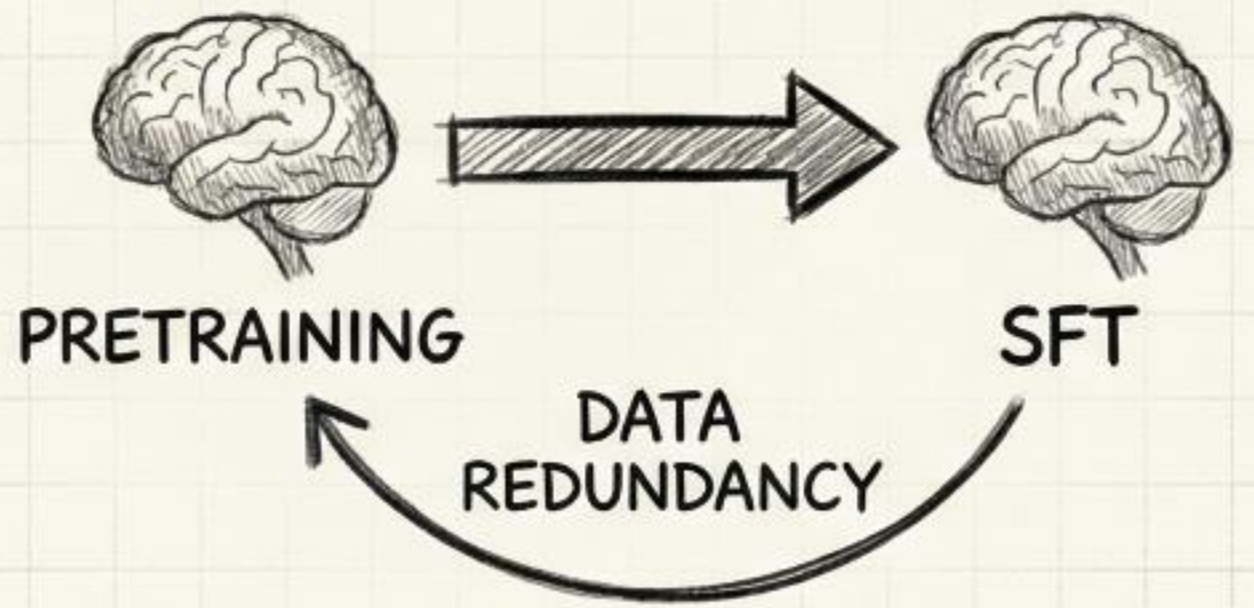
TABLE 5 RESULTS

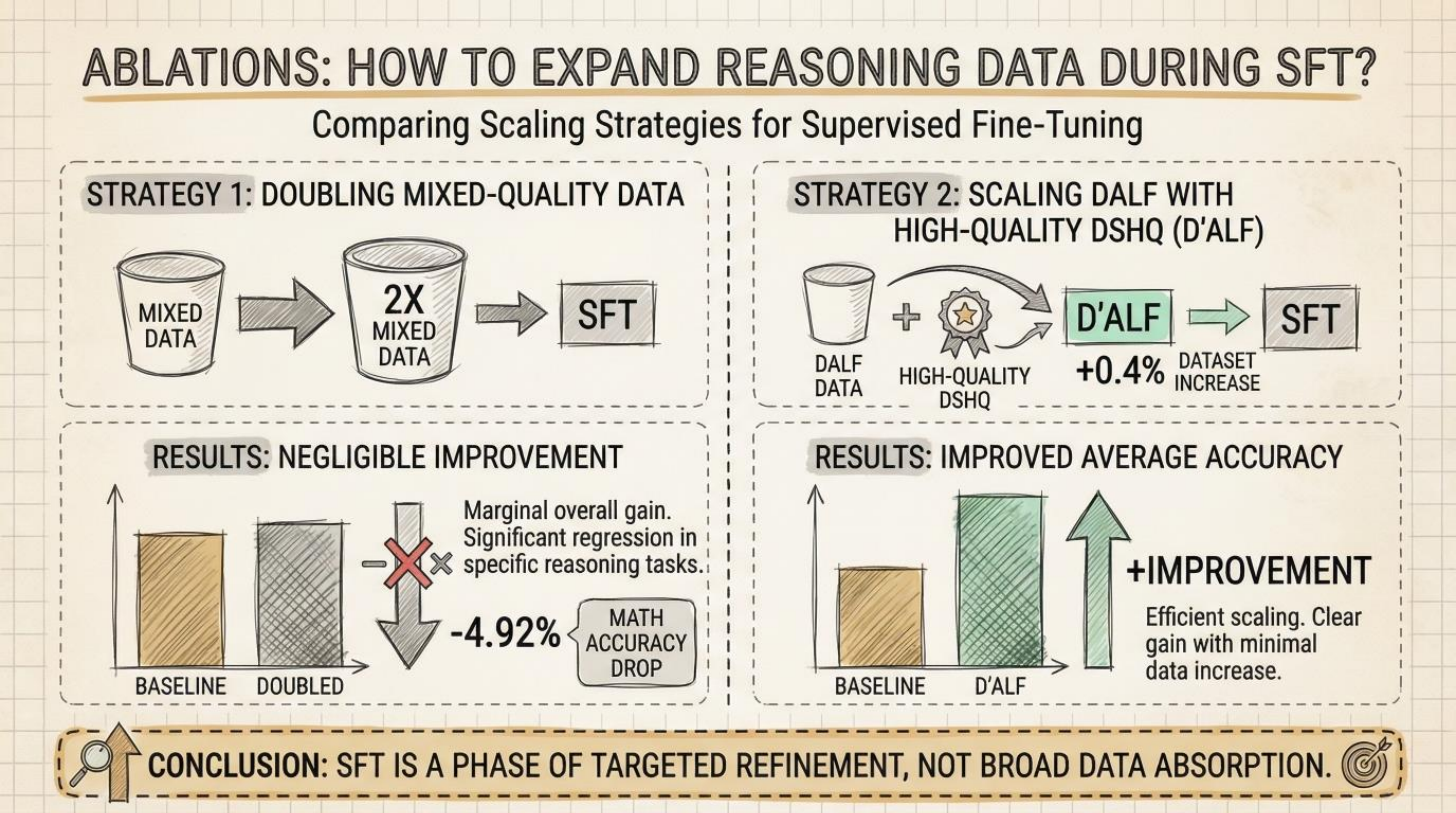
MODEL (SFT DATA)	SFT PERFORMANCE (RELATIVE)
DLDQ (DIVERSE, LOW QUALITY)	DEGRADES PERFORMANCE
DLMQ (DIVERSE, MED. QUALITY)	UNDERPERFORMS
DSHQ (SMALL, HIGH QUALITY, LONG-CoT)	+HIGHEST PERFORMANCE

→ BLINDLY SCALING DIVERSE REASONING DATA DURING SFT DEGRADES PERFORMANCE.

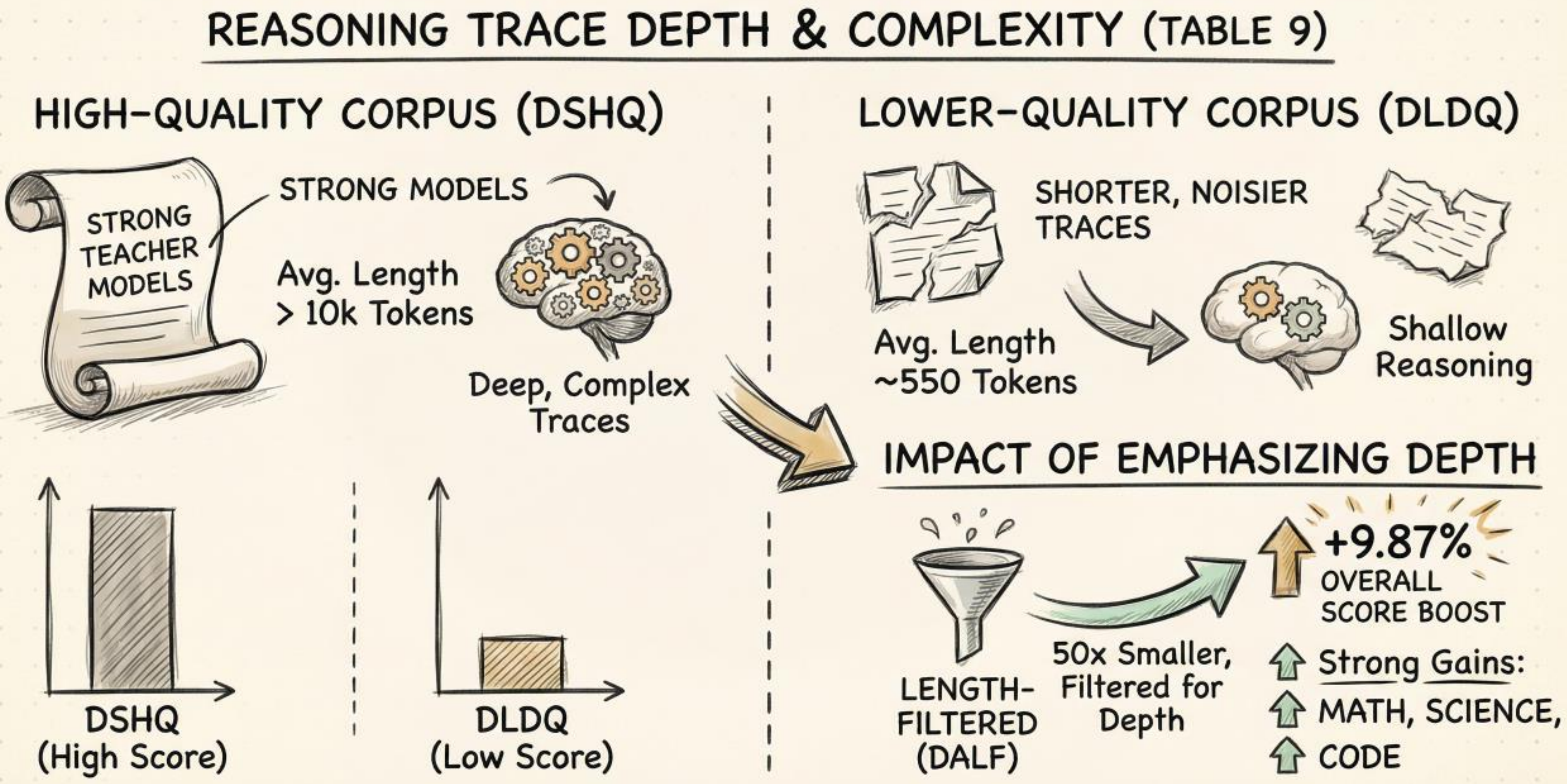
SKILL CONSOLIDATION MECHANISM

DATA REDUNDANCY BETWEEN PRETRAINING & SFT = POWERFUL MECHANISM FOR SKILL CONSOLIDATION.



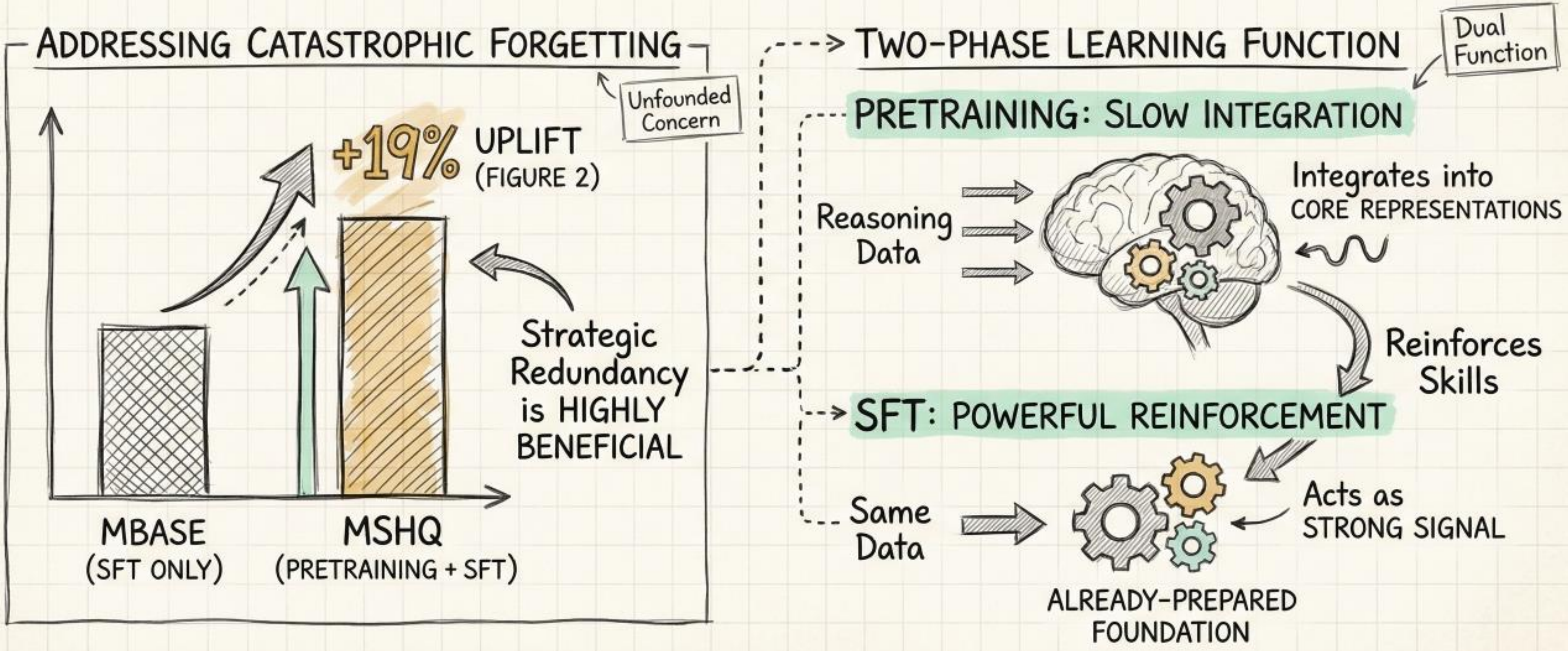

ABLATIONS: HOW TO EXPAND REASONING DATA DURING SFT?
Comparing Scaling Strategies for Supervised Fine-Tuning
STRATEGY 1: DOUBLING MIXED-QUALITY DATA
MIXED DATA
2X MIXED DATA
SFT
STRATEGY 2: SCALING DALF WITH HIGH-QUALITY DSHQ (D'ALF)
DALF DATA
HIGH-QUALITY DSHQ
D'ALF
SFT
+0.4%
DATASET INCREASE
RESULTS: NEGLIGIBLE IMPROVEMENT
Marginal overall gain. Significant regression in specific reasoning tasks.
-4.92%
MATH ACCURACY DROP
BASELINE
DOUBLED
RESULTS: IMPROVED AVERAGE ACCURACY
+IMPROVEMENT
Efficient scaling. Clear gain with minimal data increase.
BASELINE
D'ALF
CONCLUSION: SFT IS A PHASE OF TARGETED REFINEMENT, NOT BROAD DATA ABSORPTION.

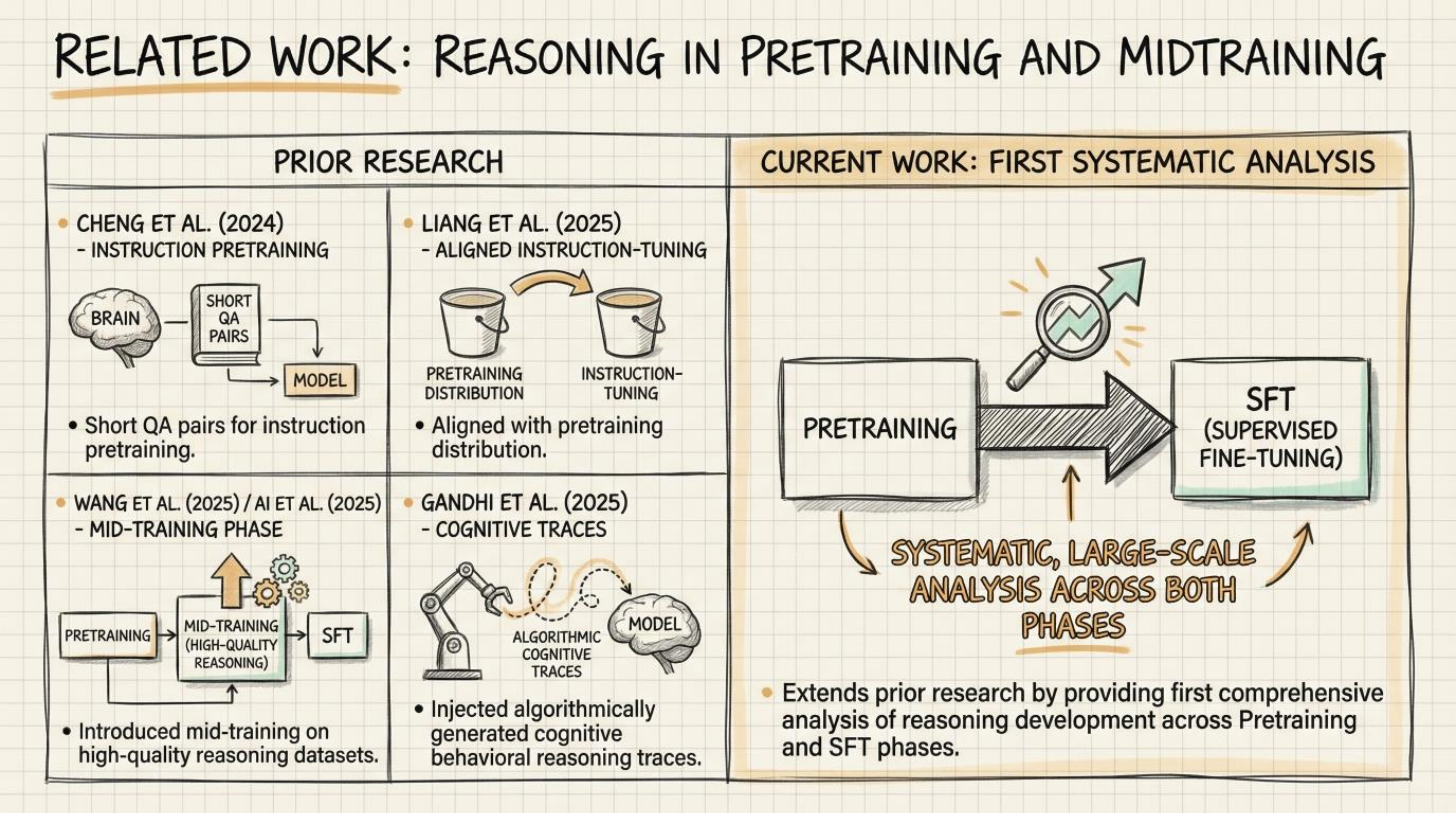
ANATOMY OF HIGH-QUALITY REASONING DATA IN SFT



DATA REDUNDANCY REINFORCES FOUNDATIONAL SKILLS

Strategic Redundancy > Catastrophic Forgetting Concerns




RELATED WORK: REASONING IN PRETRAINING AND MIDTRAINING
PRIOR RESEARCH
CHENG ET AL. (2024)
- INSTRUCTION PRETRAINING
BRAIN
SHORT QA PAIRS
MODEL
• Short QA pairs for instruction pretraining.
LIANG ET AL. (2025)
- ALIGNED INSTRUCTION-TUNING
PRETRAINING DISTRIBUTION
INSTRUCTION-TUNING
• Aligned with pretraining distribution.
WANG ET AL. (2025) / AI ET AL. (2025)
- MID-TRAINING PHASE
PRETRAINING
MID-TRAINING (HIGH-QUALITY REASONING)
SFT
• Introduced mid-training on high-quality reasoning datasets.
GANDHI ET AL. (2025)
- COGNITIVE TRACES
ALGORITHMIC COGNITIVE TRACES
MODEL
• Injected algorithmically generated cognitive behavioral reasoning traces.
CURRENT WORK: FIRST SYSTEMATIC ANALYSIS
PRETRAINING
SFT (SUPERVISED FINE-TUNING)
SYSTEMATIC, LARGE-SCALE ANALYSIS ACROSS BOTH PHASES
Extends prior research by providing first comprehensive analysis of reasoning development across Pretraining and SFT phases.

CONCLUSION: STRATEGIC DATA ALLOCATION FRAMEWORK