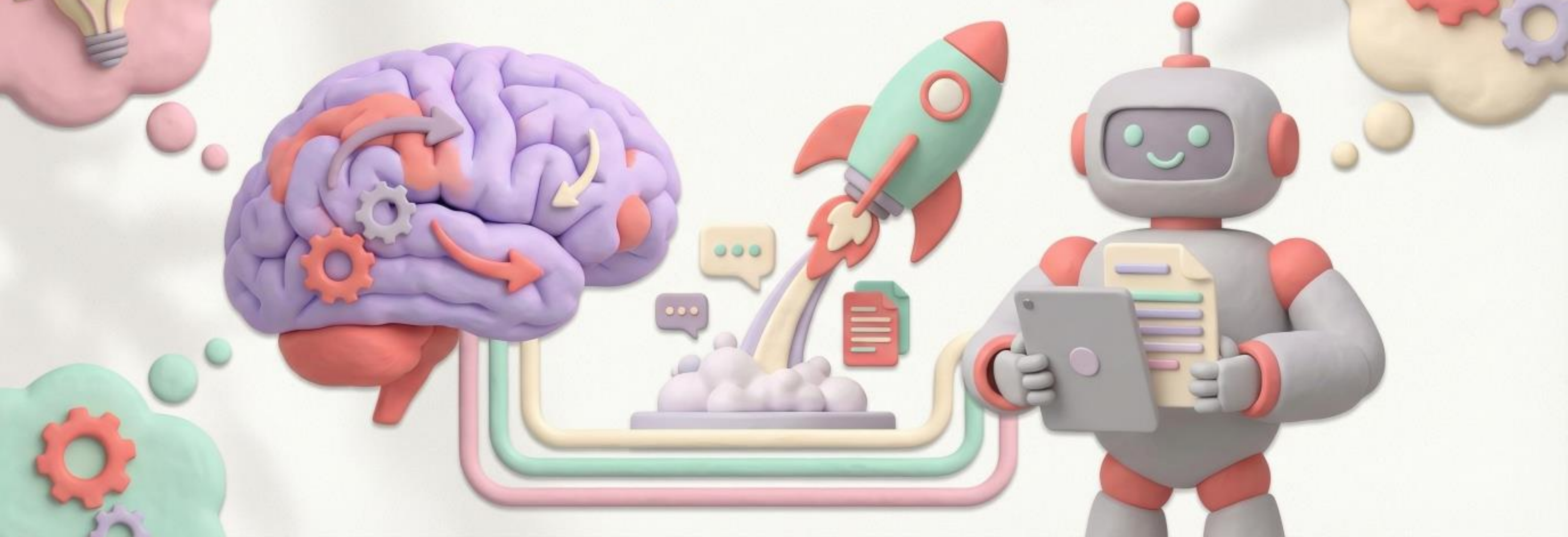
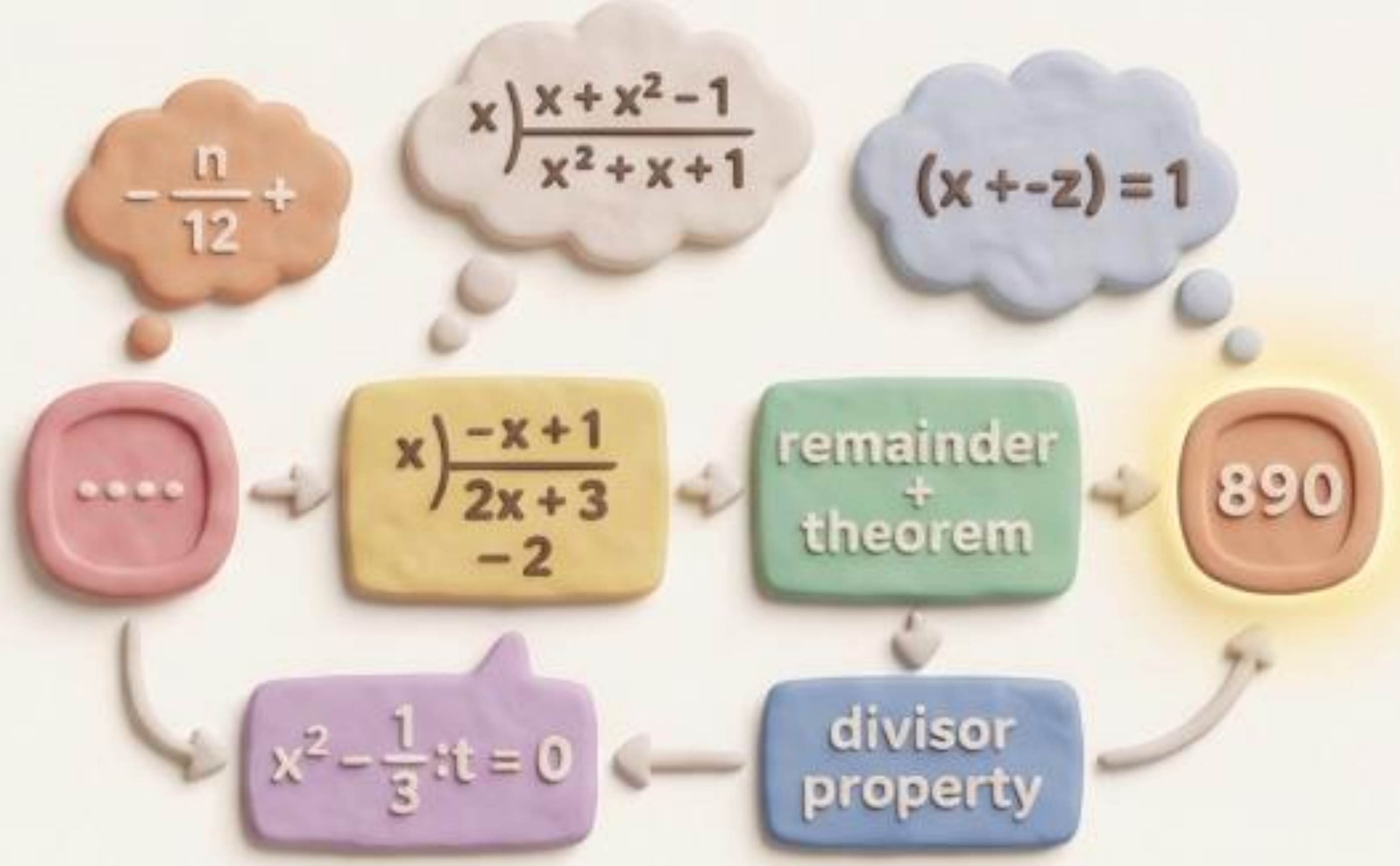
THINKING AUGMENTED PRE-TRAINING

Enhancing LLM Data Efficiency



The Challenge: Data Scarcity in LLM Scaling

Compute is growing exponentially, but **high-quality, human-authored data is finite** and largely **exhausted.**



Certain high-quality tokens (like ‘890’) are outputs of intricate multi-step reasoning, exceptionally **difficult to learn through simple single next-token prediction.**

TPT Solution: Universal Thinking Augmentation

SCALABILITY
Automated process, no human annotation, document-level operation.

DYNAMIC COMPUTE
Allocates more training compute to challenging samples.

LLM-FRIENDLY FORMAT
Transforms noisy web data into a learnable format.

Experimental Setup: Comprehensive Training Configurations

Abundant Data Pre-training

100B Tokens

8B Parameter Models

Constrained Data Pre-training

10B Raw Tokens

40B Training Budget

Mid-training from Checkpoints

Checkpoint

1.5B to 7B Models

Qwen2.5 & LLaMA-3 Families

Datasets Used

- MegaMath-Web-Pro-Max
- FineWeb-Edu
- Mixture-of-Thoughts (350k)

Evaluation Benchmarks

- Math
- Coding
- General Reasoning

Abundant Data Results: 3× Data Efficiency Achievement

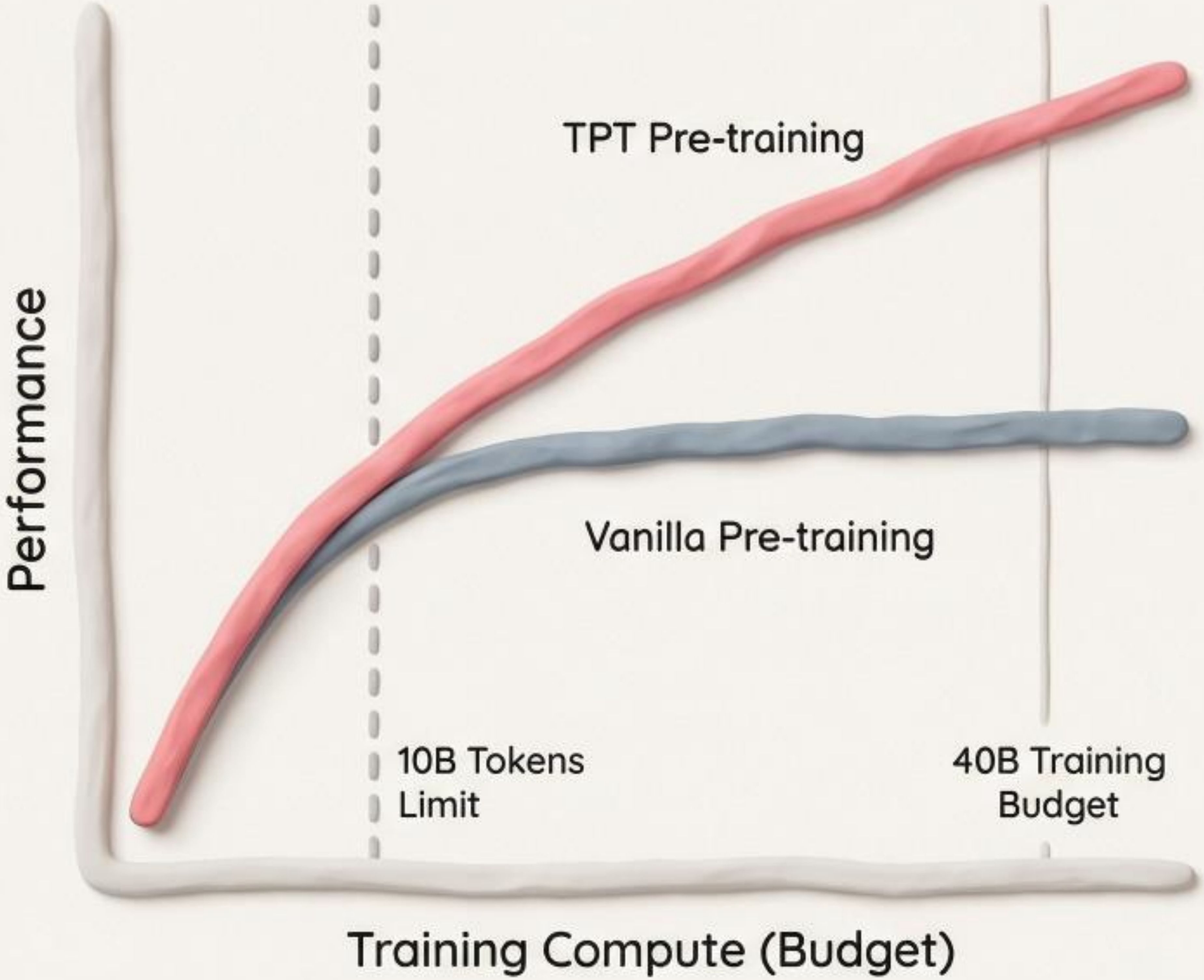
Key Quantitative Results from Abundant Data Pre-Training

- **GSM8k** 19.2% → 50.1%
- **MATH** 9.1% → 21.8% (>2× Increase)

Lower Loss,
Less Noisy,
More Learnable Data

After SFT on 2B Tokens, TPT-8B Outperforms LLaMA-3.1-8B-Instruct on EVERY Benchmark!

Constrained Data Results: Sustained Improvement

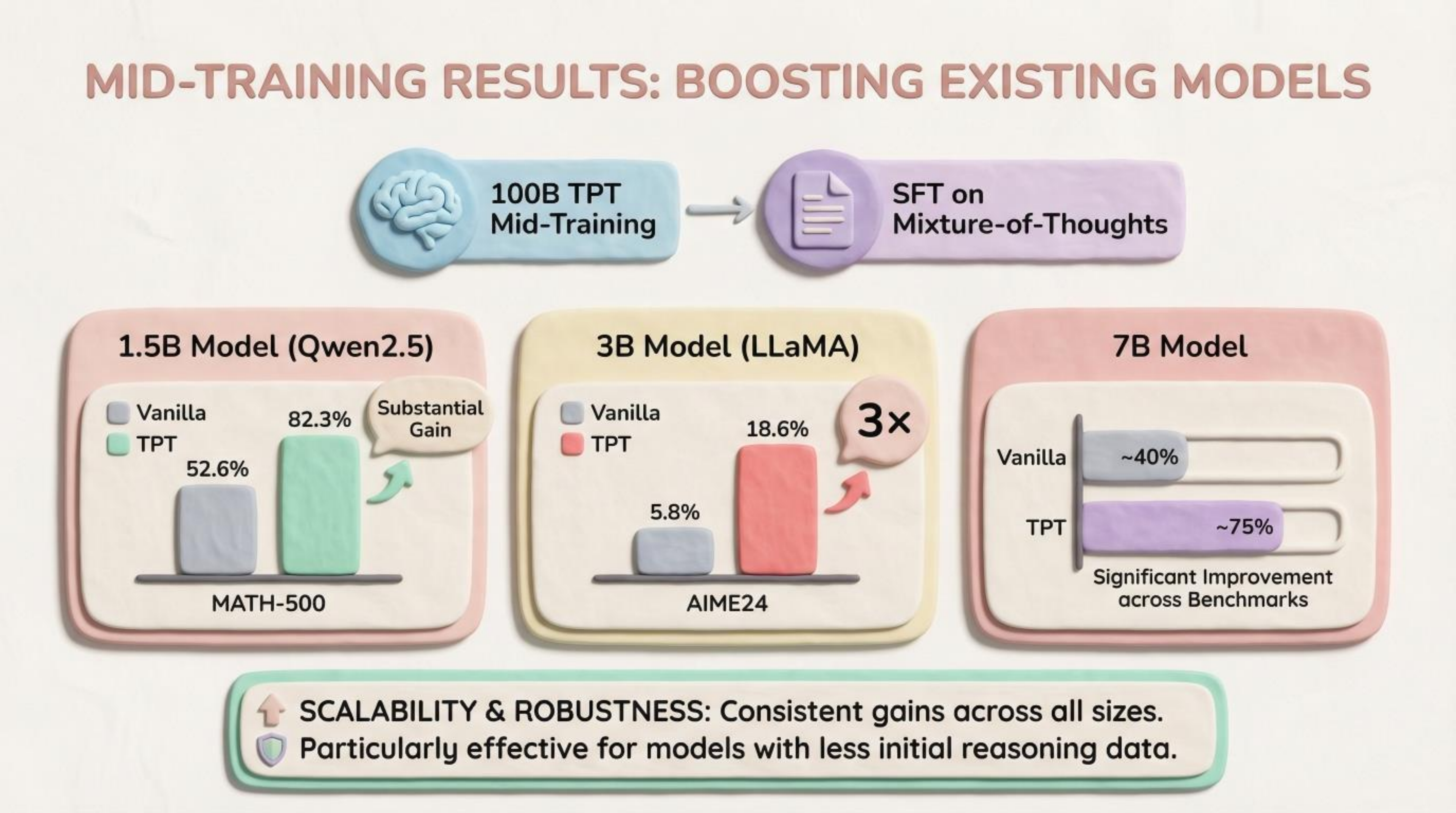


- TPT continues steady improvement while vanilla plateaus after exhausting unique tokens.

- Significant divergence on mathematical reasoning tasks (GSM8k & MATH).

- GSM8k: TPT 53.6% (Vanilla 16.6%)
- MATH: TPT 33.7% (Vanilla 12.8%)

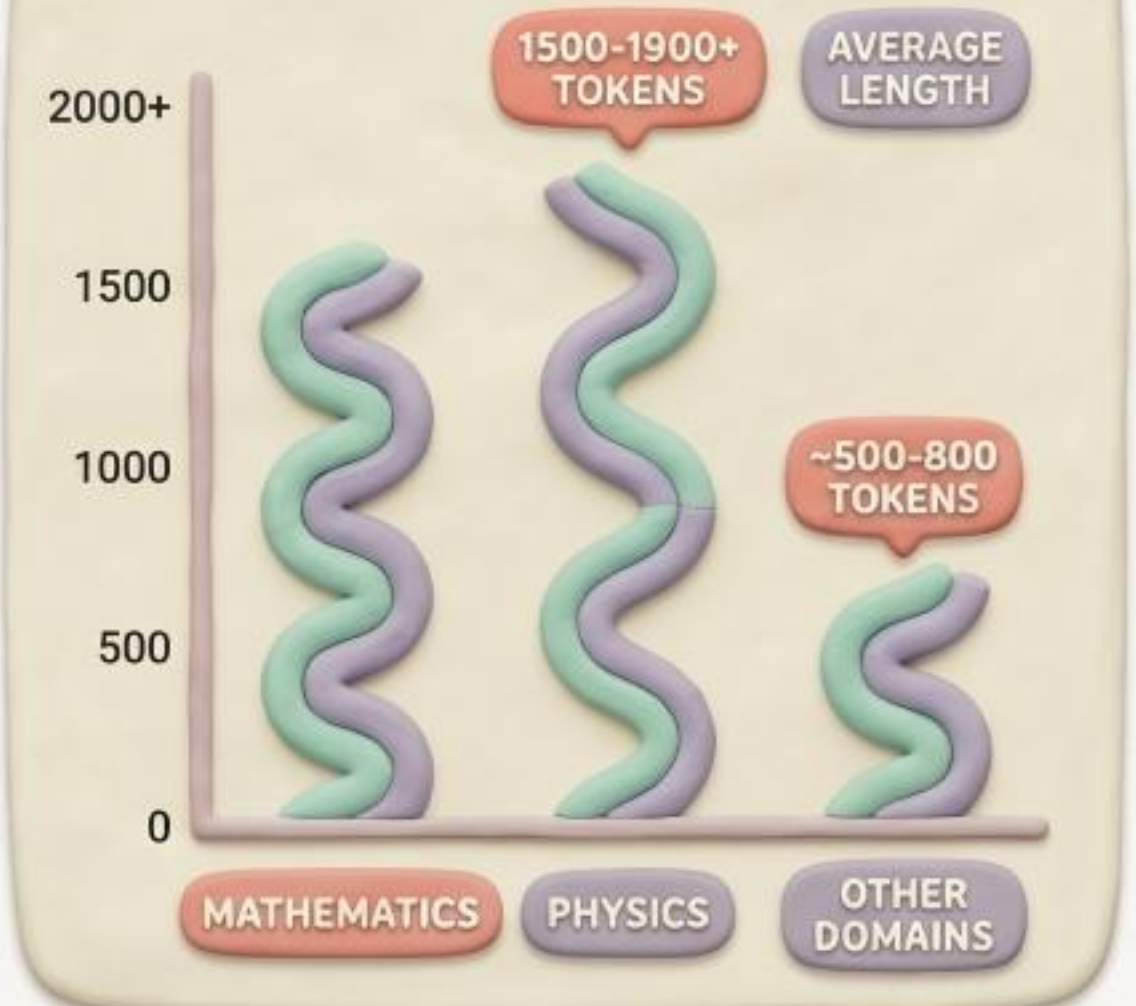
TPT enables models to extract more value from the same underlying data.

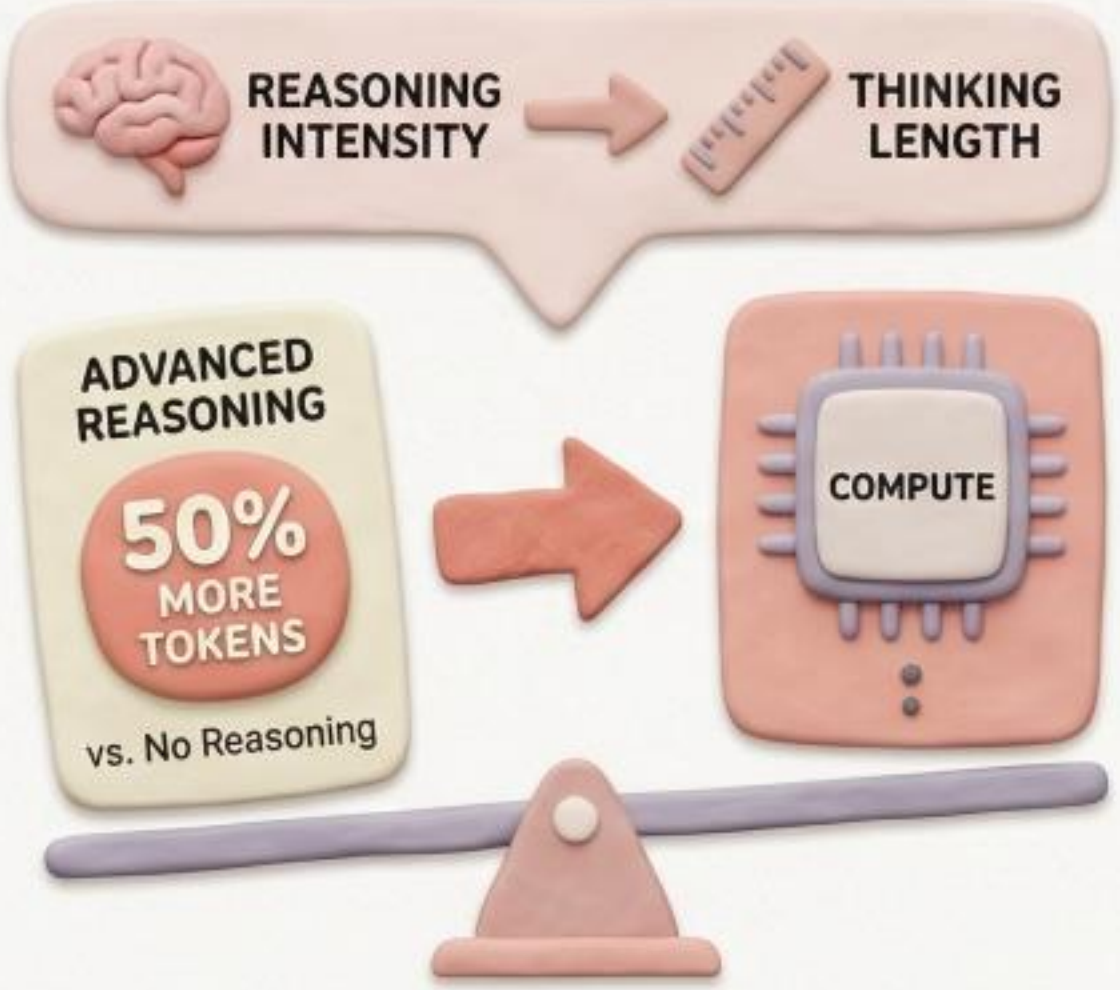

MID-TRAINING RESULTS: BOOSTING EXISTING MODELS
100B TPT
Mid-Training
SFT on
Mixture-of-Thoughts
1.5B Model (Qwen2.5)
Vanilla
TPT
82.3%
Substantial Gain
52.6%
MATH-500
3B Model (LLaMA)
Vanilla
TPT
18.6%
3×
5.8%
AIME24
7B Model
Vanilla
~40%
TPT
~75%
Significant Improvement across Benchmarks
SCALABILITY & ROBUSTNESS: Consistent gains across all sizes.
Particularly effective for models with less initial reasoning data.

THINKING PATTERN ANALYSIS: NATURAL QUALITY UP-SAMPLING

Analyzing Thinking Trajectories & Automatic Compute Allocation

THINKING TRAJECTORY LENGTH BY DOMAIN





This natural up-sampling automatically allocates more training compute to high-value, intense reasoning content, without manual heuristics.

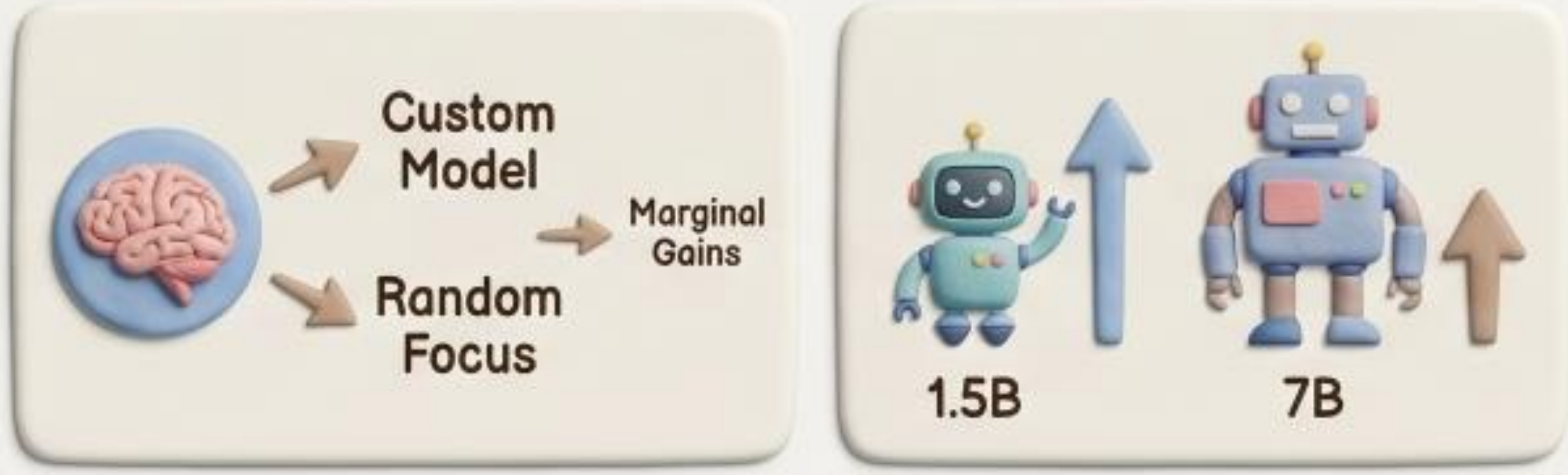
AUTOMATIC QUALITY FILTER

Thinking length positively correlates with document difficulty, functioning as an effective filter that emphasizes valuable tokens for training.

Ablation Studies: Methodological Insights

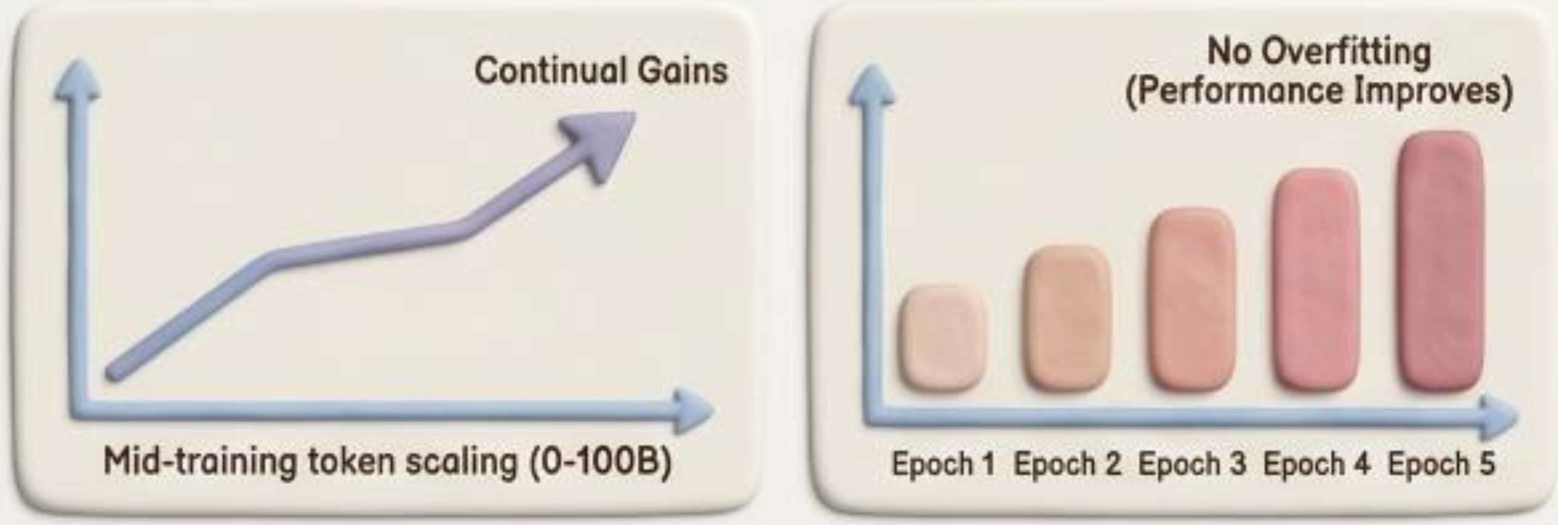
Evaluating thinking generation strategies and training parameters

Thinking Generation Strategy



Smaller models (1.5B) outperform larger ones (7B), suggesting better trajectories for downstream learning

Scaling and Data Size



100B tokens provide substantial improvements over direct SFT.
No serious overfitting even at 5 epochs.

Key Insights

- Simpler, smaller thinking models are surprisingly effective.
- Increased pre-training tokens yield continual performance gains.
- SFT data size does not cause overfitting, allowing for extended training.

Optimizing model size and training data is crucial for efficient LLM scaling.