GOOGLE CLOUD AI RESEARCH

Reverse Thinking Makes LLMs Stronger Reasoners

Introducing REVTHINK: A Framework for Enhancing Reasoning through Bidirectional Learning

University
UNC Chapel Hill

Research
Google Cloud AI

Innovation
Google DeepMind

PRESENTATION STRUCTURE

Research Overview

INTRODUCTION & MOTIVATION

The Power of Reverse Thinking

Carl Jacobi's Principle

"Invert, always, invert."

A fundamental principle in mathematical and logical reasoning that emphasizes working backward from solutions to verify correctness

Two Fundamental Research Questions

Q1 Broader Applicability

Can reverse thinking be applied to **broader, less structured domains** beyond highly structured mathematical problems?

Q2 Training vs. Verification

Instead of using backward reasoning for **verification at test time**, can we **train a model to inherently think backward**?

Human Reasoning Example

F Forward Reasoning

Emma has 2 apples, Jack has 3 → 2 + 3 = 5 apples total

R Reverse Thinking

They have 5 apples, Emma has 2 → How many does Jack have? 5 - 2 = 3 ✓

Key Innovation: REVTHINK Framework

A novel approach consisting of **data augmentation** and **multi-task learning objectives** designed to instill reverse thinking capabilities in language models while maintaining zero-shot test-time efficiency

FORMAL PROBLEM DEFINITION

Problem Setup & Formulation

Original Dataset

$D = \{(Q^{(i)}, A^{(i)})\}$

n samples with questions and answers

- **Q(i)** = Original question
- **A(i)** = Ground truth answer

Teacher Model

T (Black-box)

Large, capable model (Gemini-1.5-Pro)

- Generates reasoning chains
- Creates backward questions
- Validates consistency

Student Model

S (Trainable)

Smaller model to enhance (Mistral/Gemma-7B)

- Learns bidirectional reasoning
- Maintains efficiency at test time

Training Objective

Train student model S using augmented dataset D_{aug} with multi-task objectives

At test time: student performs **zero-shot forwar d reasoning only**

Augmented Dataset Structure

$D_{aug} = \{Q, R_f, Q_b, R_b\}$

- **Q** Original question
- **R_f** Forward reasoning
- **Q_b** Backward question
- **R_b** Backward reasoning

Key Innovation: All components generated by teacher model T through few-shot prom pting with validation

METHODOLOGY: STAGE 1

Data Augmentation: Two-Stage Process

01 Generation Stage

Teacher model T generates three components through few-shot prompting:

① **Forward Reasoning (R_f)**
Chain-of-thought from question to answer

② **Backward Question (Q_b)**
Inversely correlated question using answer

③ **Backward Reasoning (R_b)**
Solution to the backward question

02 Filtering Stage

Quality control through two validation criteria:

✓ **Correct Forward Reasoning**
$g(R_f) = A$ (validated against ground truth)

✓ **Consistent Backward Reasoning**
Teacher validates alignment with original question

Example: Math Word Problem

Q Original
John has 3 apples, Emma has 2. Total?

R_f Forward
3 + 2 = 5 apples total

Q_b Backward Q
John and Emma have 5. If Emma has 2, how man y does John have?

R_b Backward R
5 - 2 = 3 apples ✓

~50%
Average retention rate after filtering

2,000
Typical filtered samples per dataset

100%
Correct forward reasoning in final set

METHODOLOGY: STAGE 2

Three Learning Objectives

a Forward Reasoning

Standard distillation

$Q \to R_f$

Generate forward reasoning from question

Input: Original question Q

Target: Forward reasoning chain R_f

Standard knowledge distillation objective similar to Symbolic Knowledge Distillation (SKD)

b Backward Question

Auxiliary task

$Q \to Q_b$

Generate backward question from original

Input: Original question Q

Target: Backward question Q_b

Teaches model to "think" about how to invert proble ms and determine the right questions to ask

c Backward Reasoning

Auxiliary task

$Q_b \to R_b$

Generate backward reasoning from Q_b

Input: Backward question Q_b

Target: Backward reasoning chain R_b

Reinforces student's ability to reason backward and solve reversed problems

Joint Multi-Task Objective Function

$$L = \tfrac{1}{3}[\ell(S(Q), R_f) + \ell(S(Q), Q_b) + \ell(S(Q_b), R_b)]$$

Forward $\ell(S(Q), R_f)$
Standard reasoning

Backward Q $\ell(S(Q), Q_b)$
Question inversion

Backward R $\ell(S(Q_b), R_b)$
Reverse reasoning

Training Phase

Model learns all three objectives simultaneously

Test Phase

Model performs only forward reasoning (zero-shot)

EVALUATION FRAMEWORK

Experimental Setup & Datasets

Model Configuration

Teacher Model

Gemini-1.5-Pro-001

Black-box access, output only

Student Models

Mistral-7B-Instruct

Gemma-7B-Instruct

Training Parameters

Fine-tuning Method **LoRA (rank=32)**	Decoding Strategy **Greedy (T=0)**
Math Tasks Epochs **3 epochs**	Other Tasks Epochs **10 epochs**
Mistral LR **5e-6**	Gemma LR **2e-4**

12 Evaluation Datasets Across 5 Domains

Commonsense Reasoning

- **StrategyQA:** Multi-hop reasoning
- **CommonsenseQA:** Commonsense knowledge
- **ARC-challenge:** Advanced reasoning

Mathematical Reasoning

- **MATH:** Competition problems
- **GSM8K:** Grade school math
- **TabMWP:** Tabular data

Natural Language Inference

- **ANLI:** Adversarial NLI
- **e-SNLI:** Stanford NLI (held-out)

Logical Reasoning

- **Date Understanding:** Temporal logic
- **BoolQ:** Boolean questions (held-out)

Baseline Categories (3 types)

Zero-shot	Knowledge Distillation	Data Augmentation
Student's raw performance	SKD, Distill Step-by-Step	Rephrase, Q-Aug, A-Aug

EXPERIMENTAL RESULTS

Main Results: Performance Improvements

Key Performance Gains

vs. Zero-shot **+13.53%**

Average improvement across 12 datasets

vs. SKD Baseline **+6.84%**

Improvement over strongest distillation

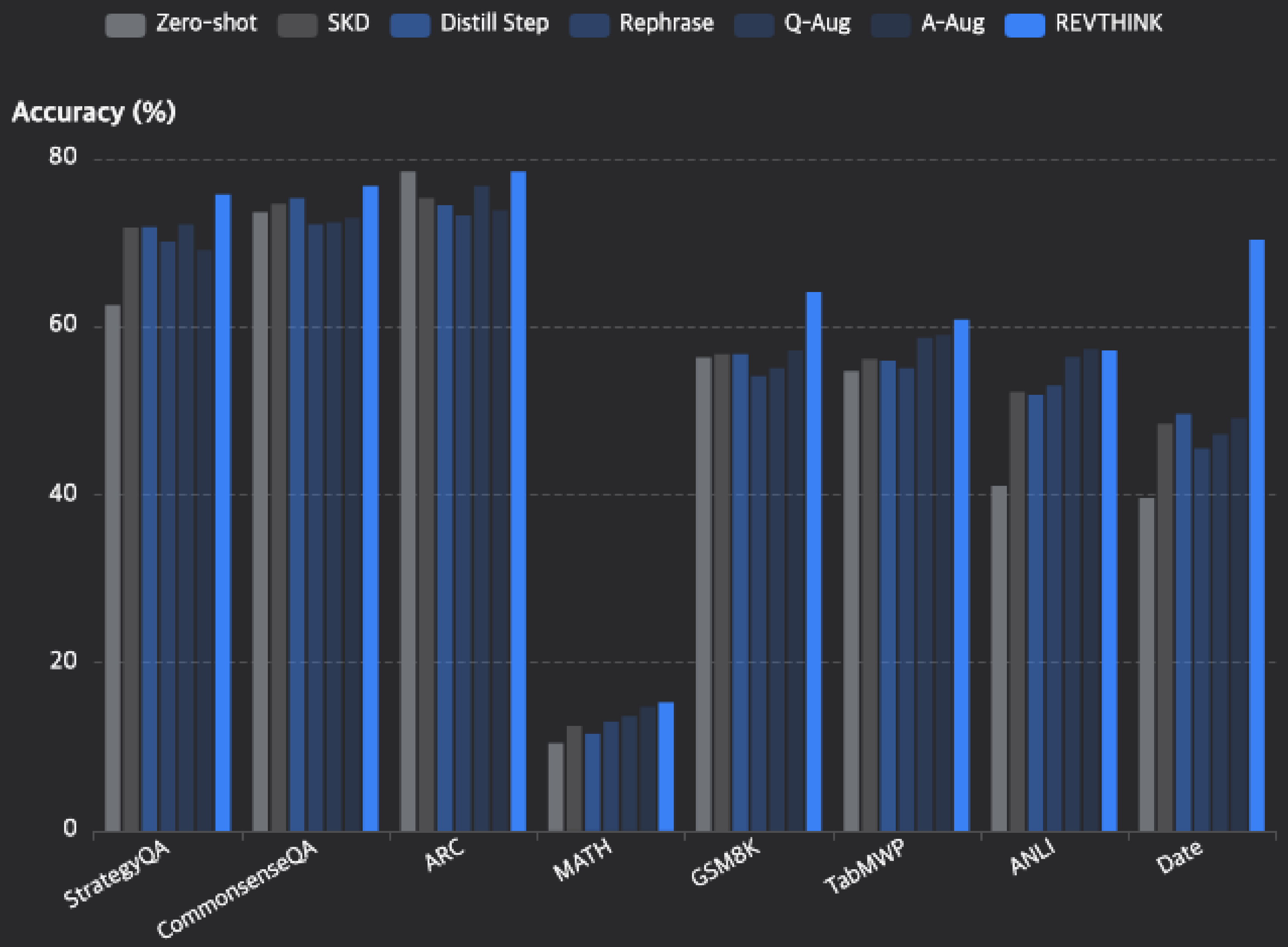
vs. Data Augmentation **+4.52-7.99%**

Range across different methods

Notable Observations

- **Consistent improvements:** Across all domains, not just math
- **Mistral vs. Gemma:** Both models show substantial gains
- **Domain versatility:** Effective for structured and unstructure d tasks

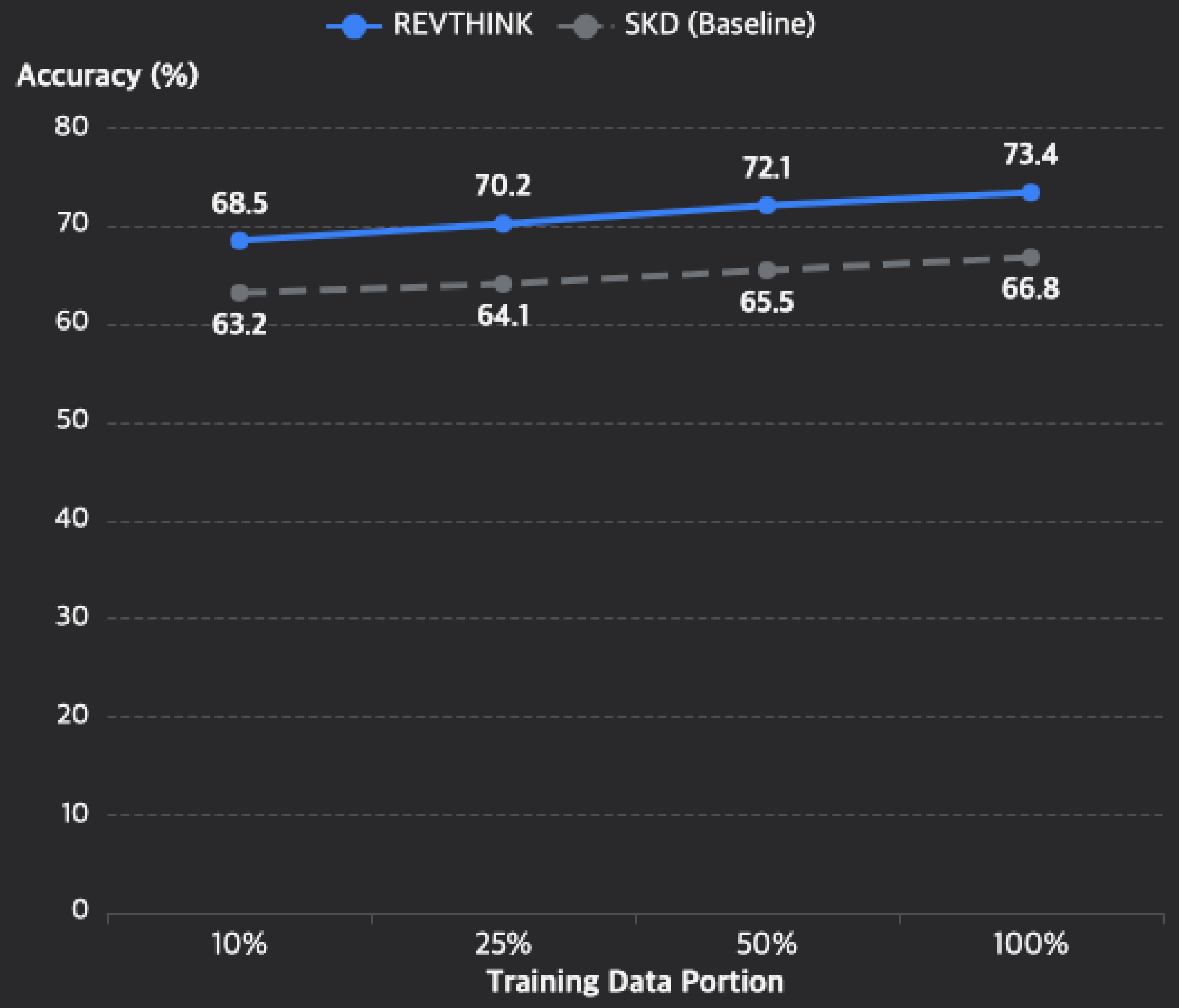
Performance on 8 Held-In Datasets



ADDITIONAL ANALYSIS

Sample Efficiency Analysis

Performance vs. Training Data Size



Key Findings

1 Superior Sample Efficiency

REVTHINK **outperforms SKD at all data portions** (10%, 25%, 50%, 100%)

2 10% Data Performance

Using **only 10% of training data**, REVTHINK surpasses SKD trained on **100% of data** on StrategyQA

3 Clear Scaling Trend

While SKD stagnates on some datasets, REVTHINK shows **clear upward performance t rend** as data increases

Implications

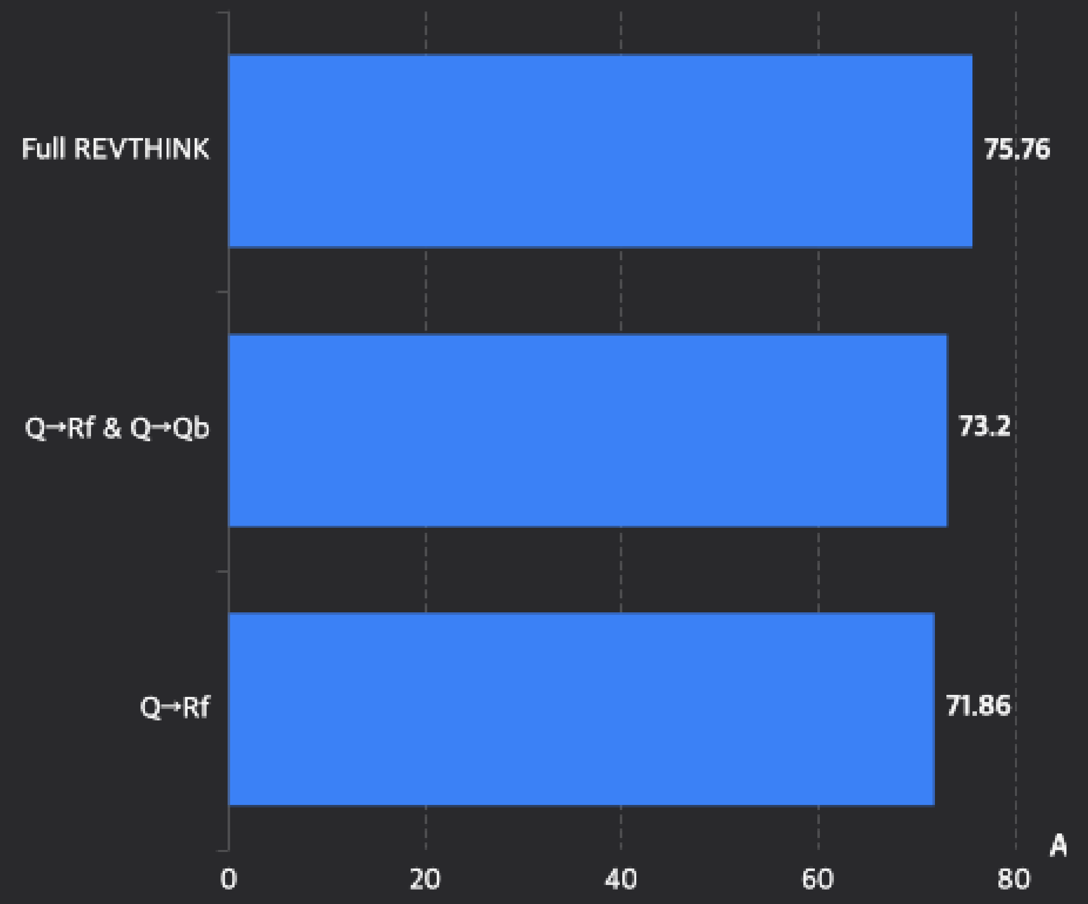
→ **Low-resource advantage:** Effective with limited training data

→ **Data quality over quantity:** Augmented data provides richer learning signal

DESIGN CHOICE VALIDATION

Ablation Studies & Design Choices

Component Ablation Analysis



Key Findings

✓ Full Objectives Work Best

$(Q \rightarrow R_f)$ & $(Q \rightarrow Q_b)$ & $(Q_b \rightarrow R_b)$ achieves highest performance

✗ Backward Reasoning Only Hurts

$Q_b \rightarrow R_b$ alone causes distributional shift

✓ Backward Question Generation Helps

Adding $Q \rightarrow Q_b$ improves over forward only

Learning Variations Comparison

Multi-task (Instruction)	Multi-task (Task-Prefix)	Joint (REVTHINK)
71.19%	69.11%	73.36%
Separated instances	Special tokens	Joint objectives

Verification Impact

StrategyQA	GSM8K	ANLI
70.97%	60.88%	48.58%
vs 68.63%	vs 59.75%	vs 47.63%

✓ Teacher verification improves data quality

MATH Dataset Breakdown

Greatest gains: **Prealgebra, Precalculus, Counting & Probability**

Least gains: **Number theory** (less invertible)

Difficulty: **Level 3 (medium-hard)** benefits most

Token Efficiency

Slightly more training tokens than AnsAug

Largely outperforms with minimal overhead

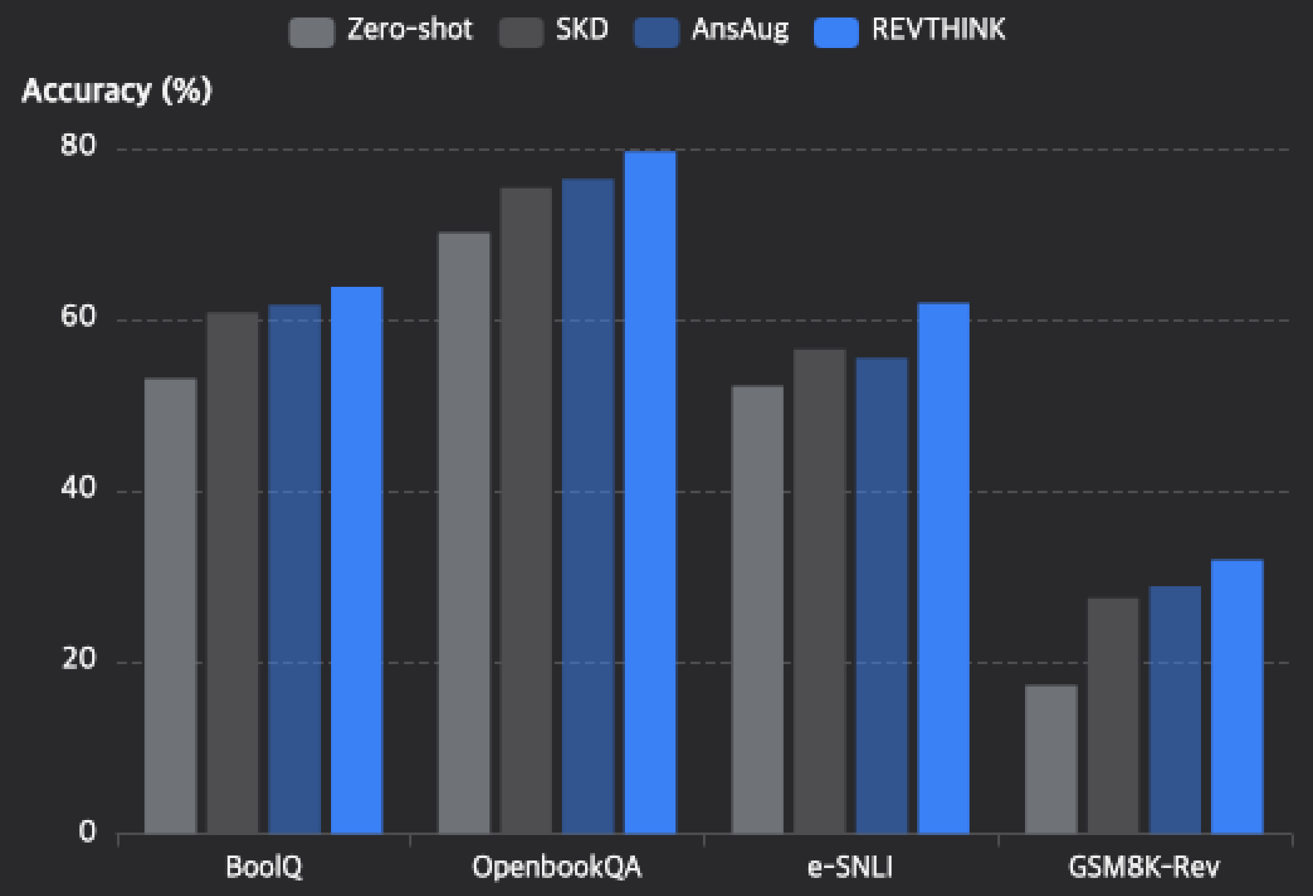
Same test-time efficiency as zero-shot

ROBUSTNESS ANALYSIS

Generalization & Scalability

Out-of-Distribution Generalization

Performance on **held-out datasets** not seen during training



Strong Generalization Performance

BoolQ	OpenbookQA
+2.11% vs best baseline	+3.20% vs best baseline

e-SNLI	GSM8K-Rev
+5.35% vs best baseline	+3.09% vs AnsAug

Model Size Scaling

Positive scaling: Consistent improvements across all model sizes

Mistral-7B + REVTHINK outperforms **Mistral-8x22B zero-shot** by 8.36% despite having **25× fewer parameters**

Complementary Strength

REVTHINK + AnsAug combination achieves **even greater performance** than either method alone

Training Efficiency

Only **slightly more tokens** than baselines during training, **identical cost** at test time

Domain Robustness

Effective across **structured and unstructured domains**, u nlike math-specific methods

CONCLUSION & FUTURE IMPACT

REVTHINK: Transforming LLM Reasoning Through Reverse Thinking

Key Contributions

1. **Effective data augmentation method** generating well-structured forward-backward reasoning chains
2. **Novel multi-task learning objectives** with auxiliary tasks for bidirectional reasoning
3. **Proven effectiveness** across 12 diverse datasets covering multiple reasoning domains
4. **Exceptional sample efficiency** outperforming baselines with 10× less data

Quantitative Achievements

+13.53%
vs Zero-shot
Average improvement

+6.84%
vs SKD
Strongest baseline

10×
Data Efficiency
Less data needed

25×
Parameter Efficiency
Fewer parameters

Impact & Implications

- Demonstrates that **smaller models can outperform much larger ones** through better training methodologies
- Provides a **cost-effective approach** to enhancing reasoning without increasing model size

Future Research Directions

→ Extending REVTHINK to **specialized domains** like scientific reasoning, legal analysis, and medical diagnosis

→ Exploring **multi-modal reverse reasoning** with vision, audio, and text