TURING

Turing AGI Advancement

Advanced Reasoning for Science, Technology, and Engineering

Elevate your model's reasoning with SOTA-breaking, human-authored Chain-of-Thought (CoT) data

Frontier models aim to deliver high-performance, advanced reasoning in STEM, finance, and coding. High-quality, human-authored CoT data can improve models—climbing various benchmarks such as GPQA, MMLU, AIME, and others.

What we do

Turing AGI Advancement helps leading AI labs improve real-world capability through:

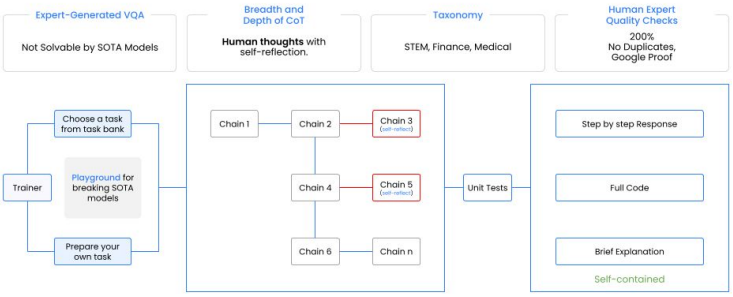
Targeted Model & Trace Evaluation

- Pinpoint model weaknesses across STEM, medical, and finance domains using human-authored, verifiable tasks
- Benchmark against GPQA, AIME, MMLU-Pro, Zerobench, Aider, and more
- Evaluate outputs via SOTA comparisons and your model's own API
- Score reasoning traces step-by-step (+1/0/–1) to identify logical breakdowns—even when answers are correct
- Designed for reward modeling, CoT audits, and performance tuning

Chain-of-Thought Data Packs

- Off-the-shelf (both exclusive and non-exclusive), human-generated reasoning traces with self-reflection
- Designed for SFT, RL, and reward modeling across a wide range of STEM disciplines and taxonomies

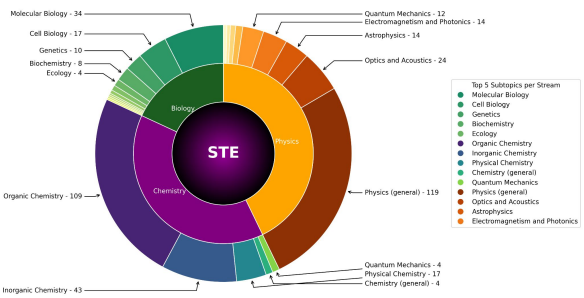
Reasoning Chain-of-Thought (CoT) Workstream



High-Difficulty Verifiable Question and Answer Sets

- Tasks unsolvable by SOTA models, across physics, biology, chemistry, medicine, finance, and competitive programming
- Includes multiple-choice questions, multimodal reasoning (MMR), verifiable questions with ground truth, expert-tier open-ended questions, and code with unit tests

STE Taxonomy



Want to strengthen your model's reasoning where it matters most?

Request premium data packs or explore evaluation pilots at: go.turing.com/llm-reasoning