Refine answers based on {all_answers_previous_round}

Human and LLM designed agents → Question

Round 1: CoT, Code, Search, ⋮, Guided Agent

Round 2: Code, Code+, Dual-Tool, ⋮, Guided Agent

Termination Round T: Code+, Search, Search+, ⋮, Guided Agent+

Selector → Answer

Termination and selection methods

Humanity's Last Exam (HLE) — Pro

Method	Score
w/o TTS	21.6
Self-MoA	29.3
Symbolic-MoE	29.5
DEI	29.1
SciMaster	26.9
GSA	28.7
TUMIX (Ours)	32.3
TUMIX+ (Ours)	34.1

GPQA, Diamond — Pro

Method	Score
w/o TTS	84.6
Self-MoA	85.5
Symbolic-MoE	86.7
DEI	86.0
SciMaster	86.9
GSA	85.8
TUMIX (Ours)	87.9
TUMIX+ (Ours)	88.1

AIME 24&25 — Pro

Method	Score
w/o TTS	87.3
Self-MoA	94.7
Symbolic-MoE	94.7
DEI	95.0
SciMaster	94.1
GSA	93.7
TUMIX (Ours)	96.7
TUMIX+ (Ours)	96.7

Humanity's Last Exam (HLE) — Flash

Method	Score
w/o TTS	9.7
Self-MoA	18.2
Symbolic-MoE	18.5
DEI	19.3
SciMaster	18.0
GSA	17.4
TUMIX (Ours)	21.2
TUMIX+ (Ours)	23.1

GPQA, Diamond — Flash

Method	Score
w/o TTS	50.0
Self-MoA	65.4
Symbolic-MoE	64.1
DEI	64.9
SciMaster	67.9
GSA	62.6
TUMIX (Ours)	77.3
TUMIX+ (Ours)	82.1

AIME 24&25 — Flash

Method	Score
w/o TTS	70.0
Self-MoA	80.3
Symbolic-MoE	80.7
DEI	82.3
SciMaster	79.1
GSA	79.7
TUMIX (Ours)	83.3
TUMIX+ (Ours)	86.7