

Answer-Agnostic

Answer-Aware

Δ (Aware – Agnostic)

↑ Higher is Better

↓ Lower is Better

Relevance

Convergence

Familiarity

Readability

Answer Leakage

LLaMA-3.1-8b

LLaMA-3.1-70b

LLaMA-3.1-405b

GPT-4

0.8

0.9

1.0

0.3

0.4

0.5

0.75

0.80

0.85

1.5

2.0

0.49

0.50

0.51

