Internal factual eval by category

Accuracy

Legend:
- chatgpt-v2
- chatgpt-v3
- chatgpt-v4
- gpt-4

80% —
60% —
40% —
20% —
0% —

Category: learning, technology, writing, history, math, science, recommendation, code, business