[redacted] — [redacted] ([redacted]: RFP RAG [redacted])

2026-06-18 ([redacted]) · [redacted] · [redacted] · [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]([redacted] GCP [redacted]) [redacted], [redacted]

[redacted] / [redacted]

- [redacted]/[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] → [redacted] **+ LLM API([redacted] B)** [redacted]. [redacted], [redacted] GitHub [redacted] + [redacted] PDF + [redacted].
- [redacted] A(GCP+HuggingFace)[redacted] B(API)[redacted] [redacted]. A[redacted] [redacted] [redacted] [redacted] [redacted] GPU·[redacted] [redacted] [redacted], B[redacted] [redacted] [redacted] [redacted] [redacted] API[redacted] [redacted]. [redacted] **B[redacted] [redacted], A[redacted]** [redacted].
- Vector DB[redacted] Chroma[redacted] [redacted] [redacted]([redacted] [redacted] + [redacted] [redacted]). [redacted]/[redacted] [redacted] [redacted](`text-embedding-3-small`, `gpt-5-mini`).
- GitHub [redacted] [redacted], [redacted] [redacted] [redacted].

[redacted] / [redacted]

- API [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] → [redacted] [redacted] **OpenAI** [redacted] [redacted]. [redacted] `.env`[redacted] [redacted] `.gitignore`[redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]

- [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] "[redacted] [redacted] [redacted]"[redacted] [redacted] [redacted].

2026-06-19 · [redacted] [redacted] · [redacted] [redacted](EDA)

[redacted] [redacted]

- [redacted] [redacted] [redacted] [redacted]
- [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted]

[redacted] [redacted] / [redacted]

- [redacted] 100[redacted](hwp 96 + pdf 4)[redacted] `data_list.csv` [redacted] [redacted].
- [redacted] [redacted]: [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted].
- [redacted] [redacted]: `data_list.csv`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. hwp [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] ~3,800[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] ([redacted] [redacted] [redacted] [redacted] [redacted]).
- [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]

- [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] ([redacted] [redacted] [redacted], Day 3[redacted] [redacted] [redacted] [redacted]).

[redacted]

- [redacted] RFP[redacted]. [redacted] [redacted], [redacted]·[redacted].

2026-06-22 · [redacted] · [redacted]

[redacted]

- [redacted]→[redacted]→[redacted]→[redacted]→[redacted] end-to-end[redacted] "[redacted]" [redacted]

[redacted] / [redacted]

- [redacted]: `config.py`([redacted]), `src/{data_loader, chunking, vectorstore, retriever, generator, rag}.py`, `scripts/{build_index, ask}.py`.
- [redacted] 800/[redacted] 150 [redacted]. [redacted] **+** [redacted] "[redacted]" [redacted].
- [redacted] + MMR [redacted].

[redacted] / [redacted] ([redacted])

- **DuplicateIDError**: [redacted]. [redacted] doc_id[redacted] [redacted] `nan`[redacted] Chroma[redacted] ID[redacted]. Day 2[redacted] "[redacted]" [redacted]. → doc_id[redacted] "[redacted], [redacted]"[redacted].
- [redacted]: ID [redacted].

[redacted]

- [redacted] "[redacted]" [redacted] **end-to-end [redacted].**

[redacted]

- [redacted]. [redacted]… [redacted].

2026-06-23 · [redacted] · [redacted]/[redacted]

[redacted]

- "[redacted]"[redacted] "[redacted]"[redacted]

[redacted] / [redacted]

- [redacted] 8[redacted] + [redacted] 1[redacted] + [redacted]([redacted]) 1[redacted] = **10**[redacted]([redacted]) [redacted]. [redacted].
- [redacted]: [redacted] / Hit@k / MRR / groundedness([redacted]) / relevance([redacted]) / abstention [redacted] / [redacted].
- groundedness·relevance[redacted] [redacted] **LLM-as-judge([redacted]-[redacted])** [redacted].

[redacted]

- [redacted]([redacted])[redacted]. [redacted] [redacted].

- LLM [redacted]? [redacted](A vs B)[redacted].

[redacted]

- **[redacted] "[redacted]"[redacted]. [redacted], [redacted], [redacted], [redacted] — [redacted].**

2026-06-24 · [redacted] · [redacted] · [redacted]

[redacted]

- **[redacted]**

[redacted] / [redacted]

- `scripts/evaluate.py` [redacted], [redacted]:
- retrieval_coverage **0.5**, hit_all **0.375**, mrr **0.53**
- abstention_accuracy **1.0**, [redacted] **~11**[redacted]
- **[redacted]. [redacted]**. [redacted] (1) [redacted], (2) [redacted] [redacted].
- [redacted](**[redacted]** "[redacted]" 100%).

[redacted] / [redacted] ([redacted])

- **[redacted] None[redacted]. [redacted]**. `gpt-5-mini`[redacted] `max_completion_tokens`[redacted]. [redacted] 250[redacted] → JSON [redacted] → None. [redacted] 900[redacted] `response_format=json_object` [redacted] + [redacted].
- [redacted] LLM [redacted](**[redacted]** 100 + [redacted]). [redacted] [redacted]. [redacted].
- **[redacted]**: **[redacted], LLM [redacted] [redacted].**

[redacted]

- **[redacted]**, [redacted]. [redacted] 0.5[redacted]. [redacted].

2026-06-25 · [redacted] · [redacted]

[redacted]

- **[redacted]**

[redacted] (CSV [redacted], 10[redacted])

[redacted]	coverage	mrr	groundedness	relevance	latency
1 baseline([redacted] top5)	0.562	0.542	1.43	1.43	10.9s
2 MMR top5	0.5	0.525	2.0	2.0	11.3s

3 MMR top8	0.625	0.543	2.0	2.25	9.1s
4 MMR+[redacted]	**0.625**	**0.75**	1.63	1.88	**7.7s**
5 MMR+LLM[redacted]	0.562	0.542	1.67	1.67	19.7s

[redacted] / [redacted]

- **[redacted].** MRR 0.54→**0.75**, [redacted](7.7s). [redacted] 1[redacted]. [redacted] cov 0→1[redacted]. → **[redacted].**
- **LLM [redacted].** [redacted] 2[redacted](19.7s). [redacted] [redacted]. (**[redacted]**.)
- **top8[redacted] coverage[redacted]** → [redacted].

[redacted]

1. **[redacted](1.4~2.25/5).** [redacted] CSV [redacted](~3.8[redacted]) [redacted]. → **[redacted] hwp/pdf** [redacted].
2. **[redacted](q2, q5 cov 0.0).** "[redacted]"[redacted]. → [redacted].

[redacted]

- [redacted]. "[redacted]"[redacted]. [redacted].

2026-06-26 ~ 2026-06-30 · [redacted] · [redacted]

[redacted]

- [redacted] hwp/pdf [redacted], [redacted]
- [redacted] CSV [redacted]

[redacted] / [redacted]

- **HWP [redacted]**: [redacted](**[redacted]**)[redacted]. [redacted] HWP [redacted]. [redacted] 8[redacted]. [redacted] **7.3**[redacted](CSV [redacted] **19**[redacted]), 100/100 [redacted].
- `src/data_loader.py`[redacted] `source="raw"` [redacted] + `build_index raw` [redacted].
- `src/query_rewrite.py` [redacted]: [redacted].
- [redacted] 6([redacted]+top8), 7([redacted]+top8+[redacted]) [redacted].

[redacted] / [redacted] ([redacted] — [redacted])

- `build_index raw`[redacted] 11,843[redacted](118[redacted])[redacted], [redacted] 3[redacted] `UnicodeEncodeError: surrogates not allowed` [redacted]. [redacted](build_index[redacted]).
- [redacted] `run_experiments`[redacted] 1~7[redacted] **coverage 0.0**[redacted] groundedness[redacted] 4~5[redacted]. [redacted] "[redacted]?" [redacted], [redacted] **'[redacted]'[redacted], LLM [redacted] abstention[redacted] '[redacted]=[redacted]'[redacted]**. → **[redacted]: groundedness[redacted] coverage[redacted] 0[redacted] '[redacted]'[redacted] '[redacted]'[redacted].** [redacted].

- ■■ ■■: HWP ■■■ BMP ■ ■■■ ■■■, ■ ■■■ UTF-16 ■■■ ■■■ ■■■ ■■■■■ ■■■■■ ■■ ■ ■■ ■■■ **■■■■■■ ■■■.** ■■ UTF-8/JSON■■ ■■■■ ■ ■ API ■■ ■ ■■ ■.
- ■■: (1) ■■■■ ■■+■■ ■■■■■ ■■ ■■■ ■■■ ■■■■■■ ■■, (2) `_clean`■ ■■■ ■■■ `encode('utf-8','ignore')` ■■■■ ■■. ■■■■■■ 100/100 ■ ■■■ UTF-8 ■■■ ■■■■ ■ ■■ ■ ■■■.
- ■■: ■■■■ ■■(HWP) ■■■ ■■■■ ■■■■■(■■■■■)■ ■■■ ■■■■ ■■. ■■■ "■■■■■ ■■■ ■■"■ ■■ ■■■■■■ ■■■■ ■■.

■■ / ■■ (Chroma ■■ ■■ + ■■■■ ■■ — ■■■ ■ ■■■■■ ■■ ■■)

- ■■■■■ ■■ ■ ■■■■■ ■■■■ 11,842■ ■■■ ■■(4■). ■■■ ■■■ **Chroma `add`■ ■■ ■■(5,461)** ■ ■■ ■■ ■■(`Batch size 11842 > max 5461`). CSV ■■(616■)■■ ■ ■■■ ■ ■■ ■■■ 11,842■■ ■■■ ■■. → **5,000■■ ■■■ add ■■■ ■■**.
- ■■ ■■■ ■ ■■■■ ■■■, ■■■ ■■ ■■ **Chroma**■ ■■ ■■■■■ **11,842**■ ■■■ ■■■■ ■■■ ■■■■ ■■■. ■■■ ■■■■■ ■■ ■■ 2■■ ■■■ ■■:
- ■■ ■■■■ ■■■■■ ■ ■■■ ■■ ■■■■ ■■■ (■■■ vs ■■■■ ■■■■ ■■■ ■■). CSV ■■■ ■■■■ cov■ 0.5■■ ■■■ ■ ■■.
- ■■■ ■■■(■■■/■■■/■■)■ ■■■ ■■■■ ■■■ ■■■ ■■■ ■■■(■: ■■■■■■■■ ■■■ ■■■■■■■■■ ■■).
- ■■: ■■■■ ■■■ **60%** ■■ ■■■ ■■■ ■■ ■■(■■ LCS), ■■ ■■■■ '■■ ■■■ ■■■'■ ■■, ■■ ■■■ ■■■ **`$in`**■■ ■■ ■■. 10■ ■■ ■■ ■■■ ■■■ ■■■ ■■ ■■.
- ■■: ■■■ ■ ■ ■■■■**,** ■■ ■■ ■■ ■■■■ ■■ ■ ■■ ■■■ ■■ ■ ■■■■ ■■■■■ ■■**(■■■).** API■ ■■ ■■ Chroma·■■·■■ ■■■ ■■ ■■ ■■■ ■■■■.

■■(■■ ■■)

- ■■■■■ ■■ ■ `build_index raw` ■■■(■■ ■ "■■ ■■: 11843■ ■■") → ■■ ■■ ■■■
- ■■: ■■ ■■■■ groundedness/relevance ■■(■■1), ■■7■■ ■■■■ ■■■■ ■■(■■2)
- ■■ ■■■■ ■■ ■■ ■■ + ■■■ ■■

■■

- ■■ ■■■■ ■■■ ■■ ■ ■■ ■■■■, ■■■ ■■■ ■■■■ ■■ ■■■■. ■■ ■■■ ■■■■ ■■■■■■ ■ ■ ■■■. ■■ "coverage 0■■ groundedness 5"■■ ■■■ ■■ ■■■ ■ ■■■■ — ■■■ ■■ ■■■ ■■■ ■ ■■■ ■ ■■■ ■■■. ■■ ■■(■■■ ■■)■ ■■ ■■■■■■■■, ■■■■ ■■ ■■■ ■■■ ■■■ ■■ ■.

2026-07-01 · ■■ ■■■ ■■■ · EDD ■■ · ■■ ■■

■■ ■ ■

- ■■(raw) ■■■■ ■■■ ■■ ■■■, ■■ ■■■ ■■ ■■■ ■■■■.
- ■■ ■■ ■■■ ■■ ■■ EDD score, ■■■, ■■■ ■■■■■■ ■■ ■■■ ■■■.
- ■■■ ■■ ■■■ ■■■ ■■■ ■■■ ■■■ ■ ■■■ ■■-■■-■■ ■■■ ■■■.

■ ■ / ■■

- `build_index raw`■ ■■■■■. ■■ ■■ 100■, ■■ **11,842**■ ■■ ■■■ ■■■■. ■■■ ■■■■■/Chroma ■■ ■■■ ■■■■ ■■■.
- ■■■ ■■■■■ ■■■ ■ ■■■■ ■■■ ■■■■ ■■■. q1■ ■■ ■■■ ■■ `finish_reason=length`, `reasoning_tokens=1024`, `content=""`■■. `gpt-5-mini`■ 1024 ■■■ ■■ ■■■ ■■■ ■■■. ■■■ ■■ ■■ ■■■ 3072■ ■■■, ■ ■■/length ■■ ■

4096 [redacted].

- `top_k`, `mmr_lambda`, `fetch_k`, `auto_filter × rewrite_query`[redacted] CSV/JSON[redacted],
- EDD score[redacted]. coverage, hit-all, MRR, groundedness, relevance, abstention, latency[redacted], false abstention[redacted] empty answer[redacted] summary [redacted] SVG[redacted].
- [redacted] `report_evidence.json/md`[redacted]. [redacted] 96[redacted].

[redacted] ([redacted])

- [redacted]: `MMR top8 + mmr_lambda 0.5 + fetch_k 20 +` [redacted] + [redacted]
- EDD **96.57**
- coverage **1.0**, MRR **1.0**
- groundedness **4.667**, relevance **5.0**
- false abstention **0.0**, empty answer **0.0**
- top5[redacted] EDD **96.53**[redacted] top8[redacted] top8[redacted].
- [redacted] `TOP_K=8`, `FETCH_K=20`, `USE_MMR=True`, `MMR_LAMBDA=0.5`, `AUTO_FILTER=True`, `REWRITE_QUERY=True`, `RERANK=False`[redacted].

[redacted]

- [redacted] + [redacted] + [redacted]. `filter_off_rewrite_off`[redacted] coverage 0.625, MRR 0.417[redacted], [redacted] q1/q4/q5[redacted]. [redacted] raw [redacted].
- [redacted] q1/q4 [redacted], rewrite[redacted] q5 [redacted].
- `mmr_lambda=0.7`[redacted] coverage[redacted] 1.0[redacted] groundedness/relevance[redacted] 3.667/4.333[redacted].
- `fetch_k=40`[redacted] coverage/MRR[redacted] groundedness[redacted] 4.0[redacted]. raw[redacted].
- [redacted] "[redacted]"[redacted] abstention[redacted] q3/q4 [redacted]. [redacted] q10[redacted] "[redacted]"[redacted] detail JSON[redacted].

[redacted] / [redacted]

- [redacted] raw [redacted] groundedness[redacted] generator[redacted].
- abstention [redacted] "[redacted]"[redacted] "[redacted]"[redacted].
- [redacted] judge score[redacted], [redacted] `missing_llm_judge_score`, `coverage_below_target`, `high_score_long_answer_needs_human_read` [redacted].

[redacted]

- [redacted] 25[redacted]. [redacted] 4[redacted], [redacted] 4[redacted], [redacted] 4[redacted], [redacted] 3[redacted], abstention 3[redacted], score trap 3[redacted].
- [redacted] JSON [redacted] JSON[redacted] UTF-8 BOM[redacted] `utf-8` [redacted] `utf-8-sig`[redacted] `questions_v2.json`[redacted] BOM [redacted] UTF-8[redacted].

- abstention [redacted] target org[redacted] `data/[redacted]/files` [redacted]. [redacted] 25[redacted] `eval/questions_v2.json`[redacted].
- [redacted] `eval/questions.json`[redacted], [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

v2 [redacted]

- [redacted] RAG [redacted] 1[redacted]. [redacted] `qv2_001_single_elec_security`, [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] coverage **1.0**, first_hit_rank **1**, abstention **false**, empty answer **false**, latency **24.93**[redacted]. [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted]. [redacted], [redacted], [redacted] [redacted], KCMVP, MQTT Broker, TLS, [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] / [redacted]

- [redacted] coverage, MRR, relevance[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] coverage/MRR 1.0, relevance 5.0[redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] API[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] CLI [redacted] [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] RAG [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. answerer[redacted] judge[redacted] [redacted], groundedness/relevance [redacted] contextual quality, evidence fit, usefulness, conciseness, wrong-document risk[redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]

- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] RAG[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- `questions_v2.json`[redacted] [redacted] 18[redacted] [redacted] 7[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. 1[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] `pass_with_caveat`[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] `coverage_below_target`, `context_mismatch` [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] v2 [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] v2 [redacted] 18[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] EDD **79.90**, coverage **0.911**, MRR **1.000**, groundedness **3.125**, relevance **4.812**, abstention accuracy **0.500**, latency **17.922**[redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] 1.0/5.0[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] groundedness[redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] coverage 1.0, rank 1, relevance 5[redacted] groundedness 2[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] coverage [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] abstention [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] multi-org [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted]. 18[redacted] [redacted] [redacted] [redacted] [redacted] 13[redacted], [redacted] [redacted] [redacted] 10[redacted], [redacted]/[redacted] [redacted] [redacted] 3[redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 1[redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 1[redacted].

- [redacted]. [redacted], [redacted]·[redacted]·[redacted]. [redacted], [redacted], [redacted].

EDD[redacted]

- [redacted] 10[redacted]. [redacted], [redacted]. [redacted].
- [redacted]. [redacted] BioIN, [redacted] "[redacted]". [redacted], [redacted].
- [redacted]·[redacted]·[redacted] 3[redacted] groundedness[redacted] 1.667[redacted] 5.0[redacted]. [redacted] 18[redacted] EDD[redacted] **79.90** → **84.79**[redacted]. [redacted]. [redacted], [redacted].
- [redacted]. [redacted]·[redacted]·[redacted]·[redacted] 4[redacted], [redacted] 3[redacted]. [redacted] "[redacted]", "[redacted]", "[redacted]", "[redacted]". [redacted].
- [redacted], [redacted]·[redacted], [redacted]/[redacted]. [redacted] 6[redacted]. [redacted] [redacted]/[redacted] 3[redacted] coverage[redacted] 1.0[redacted].
- [redacted]. qv2_021[redacted] "[redacted], [redacted], [redacted]"[redacted], [redacted]. [redacted] "[redacted]"[redacted]·[redacted]·[redacted].
- [redacted] 18[redacted] EDD[redacted] **95.55**[redacted]. [redacted] coverage **1.0**, hit-all **1.0**, MRR **1.0**, groundedness **4.938**, relevance **4.938**, abstention accuracy **1.0**, false abstention **0.0**, empty answer **0.0**, latency **25.413**[redacted].
- [redacted] top_k[redacted] 5[redacted]. [redacted] EDD[redacted] **96.34**[redacted]. [redacted] qv2_003, qv2_009, qv2_017[redacted]. [redacted], [redacted].
- [redacted]. [redacted] EDD **95.55**[redacted] top_k 8 [redacted], [redacted] EDD **96.34**[redacted] top_k 5 [redacted]. [redacted], top_k 5[redacted]. [redacted] top_k 8, [redacted] top_k 5[redacted].

[redacted]

- [redacted] EDD **95.55**[redacted], [redacted]. [redacted].
- [redacted], [redacted] groundedness[redacted]. [redacted], [redacted]. [redacted] "[redacted]"[redacted] "[redacted]"[redacted] [redacted].
- [redacted], 18[redacted]. coverage, hit-all, MRR, abstention[redacted] 1.0[redacted] groundedness/relevance[redacted] 4.938[redacted], [redacted] 7[redacted].
- [redacted], [redacted] "95.55[redacted]"[redacted] "[redacted] 95.55[redacted], [redacted]"[redacted].
- [redacted]. top_k 8[redacted] **25.413**[redacted], top_k 5[redacted] **17.494**[redacted]. [redacted], [redacted].

[redacted]

- [redacted], [redacted] 7[redacted]. [redacted] EDD **81.55**[redacted]. [redacted]. [redacted] **95.55**[redacted], [redacted]. [redacted] “[redacted]”[redacted], [redacted].
- [redacted]. [redacted], [redacted]·[redacted]·[redacted] coverage[redacted] **0.667**[redacted]. [redacted], [redacted]·[redacted]·[redacted] coverage[redacted] **0.667**[redacted]. [redacted], [redacted]·[redacted]·[redacted] “[redacted]”[redacted], [redacted], [redacted].

- [redacted]. [redacted] 8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 1[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 1[redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] “[redacted] [redacted] [redacted]·[redacted] [redacted]·[redacted] [redacted]” [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] 3[redacted] [redacted] [redacted] coverage **1.0**, hit-all **1.0**, abstention accuracy **1.0**[redacted] [redacted]. [redacted] [redacted] 7[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] EDD **96.80**[redacted] [redacted]. [redacted] [redacted] coverage **1.0**, hit-all **1.0**, MRR **1.0**, groundedness **4.833**, relevance **5.0**, abstention accuracy **1.0**, latency **19.127**[redacted].
- [redacted] [redacted] [redacted]. [redacted] 3[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] Chroma[redacted] [redacted] [redacted] [redacted] [redacted] append[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] qv2_008[redacted] groundedness **4**[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] qv2_003[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. 7[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] groundedness [redacted] **2.0**, [redacted] [redacted] **4.714**[redacted]. [redacted] [redacted] [redacted] [redacted] **4.857**[redacted]. [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] **0.0**[redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] qv2_003[redacted] [redacted] [redacted] [redacted] groundedness **3**, relevance **4**[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted]. top_k 5[redacted] [redacted] qv2_003, qv2_009, qv2_017 [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] top_k 5[redacted] [redacted], [redacted]·[redacted]·[redacted]·[redacted] [redacted] [redacted] top_k 8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted]

- qv2_003[redacted] qv2_008[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] bullet[redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] citation guard[redacted] [redacted].
- [redacted] [redacted] top_k 5, [redacted] [redacted]·[redacted]·[redacted] [redacted] top_k 8[redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]

- [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. EDD[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]

- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted]. [redacted] [redacted] EDD **95.55**[redacted] [redacted], [redacted] 7[redacted] [redacted] EDD **81.55**[redacted] [redacted]. [redacted] [redacted] [redacted] 3[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] 7[redacted] [redacted] [redacted] EDD **96.80**[redacted] [redacted] [redacted].
- [redacted] **96.80**[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] **.** [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted]. [redacted] [redacted] [redacted] `95.55 /` [redacted] `81.55`[redacted] [redacted] [redacted] [redacted], `96.80`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- 7[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] EDD[redacted] [redacted] [redacted]. [redacted] 96.80 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted]”[redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] groundedness[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] qv2_003[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] 0.0[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

- [redacted]. [redacted] [redacted] [redacted] [redacted] groundedness/relevance 0.0[redacted] [redacted] EDD **56.89**, **49.95**[redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] **81.55**[redacted]. 96.80[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]

- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] 20[redacted] [redacted], 20[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] 22[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] → [redacted] [redacted] → [redacted] [redacted] → [redacted] → [redacted] [redacted] [redacted] [redacted] → [redacted] [redacted] → [redacted] → [redacted] [redacted] → [redacted] [redacted] → [redacted] [redacted][redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted], [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted]”, “[redacted] [redacted]”, “[redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted], “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] 1[redacted]: [redacted] [redacted] [redacted]

- 18[redacted] 25[redacted] 20[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], [redacted], [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] 12[redacted] [redacted].
- [redacted] [redacted] 12[redacted] [redacted] JSON [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 7[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted](L0)[redacted] top_k 8 [redacted] [redacted] [redacted] [redacted] [redacted]. EDD[redacted] **89.69**, coverage/hit-all/MRR/groundedness/relevance[redacted] [redacted] **1.0** [redacted] **5.0**[redacted], abstention accuracy[redacted] **0.333**[redacted] [redacted]. [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted]?” [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted](L1)[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]/[redacted]/[redacted] [redacted] [redacted] [redacted] Pilot [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. EDD[redacted] **96.36**[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] top_k 5 [redacted] [redacted]. [redacted] [redacted] [redacted](L2)[redacted] [redacted] **91.41**, [redacted] [redacted] [redacted] [redacted] [redacted](L3)[redacted] **94.74**[redacted]. [redacted] [redacted] **24.021**[redacted] → **21.372**[redacted] [redacted] **2.65**[redacted] [redacted], groundedness/relevance[redacted] **5.0/5.0** → **4.667/4.778**[redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. top_k 8[redacted] [redacted]·[redacted]·[redacted] [redacted] [redacted] [redacted], top_k 5[redacted] [redacted]/PG [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] L1[redacted]. [redacted] L1[redacted] “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] L0 **89.69**[redacted] [redacted] [redacted], L1 **96.36**[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. top_k 5[redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] 1[redacted]: L4 [redacted]

- [redacted]. top_k 5[redacted] “[redacted]”[redacted]. [redacted] top_k 8[redacted], top_k 5[redacted]/PG [redacted].
- [redacted](L4)[redacted], [redacted] “[redacted]”[redacted] “[redacted]”[redacted].
- [redacted].
- [redacted] L1, EDD **96.36**[redacted]. L2/L3[redacted], L4[redacted] qv3_010[redacted].

[redacted] 1[redacted]: L5 [redacted]

- L4[redacted] “[redacted]”, “[redacted]”, “[redacted]”, “[redacted]”[redacted].
- [redacted] → [redacted], [redacted] → [redacted], [redacted] → [redacted], [redacted] → [redacted]. [redacted] L4[redacted] EDD[redacted].
- [redacted]. [redacted] L1, EDD **96.36**[redacted]. L2/L3[redacted], L4[redacted] qv3_010[redacted].
- [redacted](L5)[redacted] EDD [redacted].
- [redacted] 20[redacted] L0~L5[redacted] qv3_010 [redacted].
- [redacted] L0~L5[redacted] EDD[redacted] L0~L3[redacted], L4/L5[redacted] “[redacted], [redacted]”[redacted].

[redacted] 1[redacted]: L6 [redacted]

- [redacted]/PG/[redacted].
- [redacted]. [redacted] 106[redacted]. [redacted] 39[redacted], [redacted] 7[redacted], PG[redacted] 2[redacted], [redacted] 8[redacted], [redacted] 13[redacted], [redacted] 3[redacted], [redacted] 4[redacted], [redacted] 5[redacted].
- [redacted] “[redacted] PG [redacted]”, “[redacted]”, “[redacted]”[redacted]. [redacted] top_k 5 [redacted] “[redacted]/PG[redacted]”[redacted].
- [redacted](L6)[redacted] EDD [redacted]. [redacted] qv3_010[redacted] “[redacted]/PG/[redacted]”[redacted].

[redacted] 1[redacted]: L7 [redacted]

- [redacted] “[redacted]”[redacted].
- [redacted]/PG[redacted].

[redacted]. [redacted], [redacted]. [redacted], [redacted].

- [redacted](L7)[redacted] “[redacted]”[redacted]. [redacted]. [redacted] groundedness[redacted], [redacted] EDD[redacted]. [redacted].
- [redacted] L1[redacted], L4~L7[redacted]. [redacted], [redacted].

[redacted] 1[redacted]: L8 [redacted]

- [redacted] L7[redacted] 6[redacted] 2[redacted], [redacted] 4[redacted].
- [redacted]. [redacted]. [redacted] 6[redacted].
- [redacted] qv3_010[redacted]/PG[redacted], [redacted] L7/L8 [redacted].

[redacted] 1[redacted]: L9/L10 [redacted]

- [redacted].
- [redacted] top_k 5[redacted](L9). [redacted] 5[redacted]. [redacted]. [redacted], [redacted], [redacted]. [redacted] PG[redacted], [redacted]. L6[redacted] PG [redacted] “[redacted]”[redacted] “top_k 5[redacted]”[redacted].
- [redacted] top_k 8[redacted](L10). [redacted] 8[redacted], [redacted], PG, [redacted], [redacted], [redacted]. [redacted] “[redacted] PG [redacted]”, “[redacted]”, “[redacted]”[redacted].
- [redacted]. top_k 5[redacted]. [redacted], [redacted], [redacted], [redacted] top_k 8[redacted]. [redacted] top_k 5[redacted].
- [redacted] L9/L10[redacted]. [redacted] “[redacted] top8[redacted] PG [redacted]”[redacted].

[redacted] 1[redacted]: L11 [redacted] top_k [redacted]

- L9/L10[redacted] top_k 5 [redacted], [redacted]·[redacted]·PG·[redacted]·[redacted]·[redacted]·[redacted] top_k 8[redacted].
- [redacted] top_k 5 [redacted] 8[redacted], [redacted] 5[redacted]. [redacted] 5[redacted].
- L11[redacted]. [redacted] L9/L10[redacted]. [redacted] 5[redacted] 8[redacted].

[redacted]

- [redacted], [redacted]. L10[redacted] PG·[redacted]·[redacted]. [redacted] “[redacted]”[redacted].
- [redacted], [redacted]. [redacted], [redacted].
- [redacted], [redacted]. [redacted] EDD[redacted] “[redacted]”[redacted].
- [redacted], [redacted] top_k [redacted]. qv3_010[redacted].

- [redacted] 6~8[redacted]. [redacted], [redacted], [redacted], [redacted], [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]

- [redacted] [redacted], [redacted] 8[redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted] [redacted], [redacted] top5 [redacted], [redacted] [redacted] top5 [redacted] [redacted], top8 [redacted], [redacted] [redacted], [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted]. 8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] **3.62**, [redacted] [redacted] [redacted] **3.47**[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] E/F/G [redacted]. [redacted] top5[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]·PG[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] top5[redacted] [redacted] [redacted] [redacted] [redacted] PG[redacted] [redacted]. top8[redacted] [redacted] [redacted] [redacted] [redacted] PG·[redacted]·[redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted]. A/B/C[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] bullet [redacted] [redacted] [redacted] [redacted]. F/G[redacted] [redacted]·PG [redacted] [redacted] [redacted] top8 [redacted] [redacted] top_k[redacted] [redacted]. D/H[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] 5[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].
- [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], EDD[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted]”[redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] L0[redacted] EDD **89.69**[redacted]. L1[redacted] **96.36**[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] top_k[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] manifest[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]. [redacted] v2, v3 [redacted],
[redacted]. [redacted] 10[redacted].

[redacted]. ERP [redacted], [redacted], [redacted], [redacted], ISP[redacted]
[redacted], [redacted].

[redacted], [redacted]. [redacted]
[redacted], [redacted] [redacted].

[redacted]. [redacted] “[redacted]”[redacted]
[redacted]. [redacted] v4 [redacted].

[redacted]. [redacted] “[redacted]?”[redacted],
[redacted].

[redacted]. [redacted] L14[redacted] L17[redacted]. [redacted]
“[redacted]”[redacted]. [redacted], [redacted].

[redacted], [redacted] L14[redacted] L17[redacted]
[redacted], L16[redacted]. [redacted].

[redacted] v4 [redacted]. [redacted] “[redacted]”[redacted].

[redacted]. [redacted]
[redacted] [redacted].

7[redacted] 3[redacted] - [redacted]

[redacted], 14[redacted]. [redacted] 14[redacted], [redacted] 13[redacted]
[redacted]. [redacted]
[redacted].

[redacted] 6[redacted] [redacted]/[redacted]. [redacted] 0[redacted] 5 [redacted]
[redacted], [redacted]. [redacted]
[redacted]. [redacted], [redacted]. [redacted]
[redacted], [redacted].

[redacted] [redacted] 6[redacted], [redacted], [redacted]. [redacted]
0.063527[redacted]. [redacted].

[redacted] v4 [redacted] 10[redacted]. [redacted] EDD[redacted] 97.41[redacted], [redacted], [redacted]
[redacted], [redacted], [redacted], [redacted]. [redacted] 19.377[redacted], [redacted]. [redacted]
[redacted].

[redacted] [redacted] 18.725[redacted] 0.652[redacted], [redacted]
[redacted]. [redacted], [redacted]
[redacted].

[redacted]. [redacted] 30[redacted], [redacted]
74[redacted]. [redacted]
[redacted].

[redacted] bullet [redacted]. [redacted] EDD[redacted] 86.42[redacted], [redacted] "[redacted]"[redacted]
[redacted]. [redacted]. [redacted] EDD[redacted] 96.42[redacted]. [redacted]

[redacted], [redacted] v4 [redacted] 97.41[redacted].

[redacted]. [redacted] v4 [redacted] EDD 97.41[redacted]. [redacted], [redacted]. [redacted]
v4 [redacted]. [redacted], [redacted], [redacted].

[redacted], [redacted]. [redacted], [redacted], [redacted], [redacted]·[redacted]·[redacted]. [redacted], [redacted]. [redacted] 14[redacted].

[redacted] L30[redacted] EDD[redacted] 92.40[redacted]. [redacted]. v4[redacted], v5[redacted] EIP3.0 [redacted]. [redacted], [redacted]·[redacted]·[redacted]. [redacted], [redacted]. [redacted].

[redacted] L31[redacted]. [redacted], [redacted]. eip[redacted] EIP[redacted]. [redacted] v5[redacted] 0.0[redacted], [redacted]. [redacted], [redacted].

L32[redacted] v5[redacted] EDD 95.87[redacted]. coverage[redacted] MRR[redacted] 1.0[redacted] 20.776[redacted] 15.151[redacted]. [redacted]. [redacted] "[redacted]"[redacted], [redacted]. [redacted] L33[redacted]. EDD[redacted] 98.37[redacted], [redacted].

L34[redacted] v5 top5 [redacted]. EDD 98.70, [redacted] 13.724[redacted]. [redacted] v5 [redacted]. [redacted] top5[redacted]. [redacted] v4 [redacted].

L35[redacted] L36[redacted] v4 [redacted]. [redacted], v4 [redacted] EDD 96.81[redacted] L25[redacted] 97.41[redacted]. v4 top5[redacted] EDD 96.44[redacted]. [redacted]. [redacted]. [redacted] top5[redacted]. L33/L34[redacted] "[redacted] v5[redacted]"[redacted], [redacted] v4 [redacted] L25[redacted].

[redacted]. L30[redacted], [redacted]. [redacted]. [redacted], [redacted]. [redacted] v5[redacted] v6[redacted], [redacted], [redacted], [redacted].

7[redacted] 4[redacted] - R&D [redacted]

[redacted] "[redacted], [redacted]"[redacted] R&D Workbench[redacted]. [redacted] Workbench [redacted] RAG [redacted]. [redacted] 1[redacted], [redacted] "[redacted]"[redacted].

Ragas[redacted] TruLens [redacted] EDD[redacted], [redacted], [redacted], [redacted]. [redacted], [redacted], [redacted].
RAGChecker [redacted]. [redacted], [redacted].

[redacted] v6 [redacted]. [redacted], [redacted], [redacted]. [redacted] "[redacted]"[redacted] "[redacted]"[redacted].

[redacted]. ARES[redacted], LangSmith[redacted] RAG [redacted], [redacted]. [redacted], [redacted]/[redacted] [redacted].

[redacted]. [redacted], [redacted]. [redacted], [redacted]. [redacted], [redacted]. [redacted] claim-level audit[redacted]. [redacted], [redacted]. [redacted] EDD[redacted] 0.x[redacted], [redacted].

[redacted], v6[redacted], v7[redacted]/[redacted], v8[redacted]. v9[redacted], [redacted]. [redacted] 98[redacted] “[redacted]”[redacted].

[redacted] EDD [redacted] O[redacted] [redacted] doctrine [redacted]. [redacted] `O:\RND_KnowledgeVault\catalogs\rag_evaluation_doctrines\edd_loop_gate_doctrine_20260706.json`[redacted], [redacted] first-run [redacted], judge calibration [redacted], [redacted] EDD [redacted]/[redacted].

7[redacted] 5[redacted] v6 [redacted]

[redacted], [redacted]. [redacted] L37[redacted] EDD[redacted] 86.25[redacted]. [redacted], [redacted] 0.0[redacted]. [redacted], “[redacted]”, “[redacted]”[redacted]. v4, v5[redacted] v6[redacted].

[redacted] 211,000,000[redacted]. [redacted]·[redacted]·[redacted]. [redacted].

L38[redacted], [redacted]. [redacted], [redacted]. [redacted]/[redacted]. [redacted]. [redacted] [redacted]. [redacted] [redacted]. [redacted] “[redacted]”[redacted]. [redacted].

[redacted] qv6_006[redacted], [redacted], [redacted], [redacted]. [redacted] qv6_007[redacted]. [redacted]. [redacted], [redacted]. [redacted].

[redacted] top_k[redacted]. [redacted] L39 [redacted]. top5[redacted] top8[redacted], top12[redacted]. [redacted]. [redacted].

[redacted] L40[redacted], [redacted]. [redacted] qv6_007[redacted] 1[redacted]. [redacted]. [redacted]. [redacted]. [redacted], [redacted], [redacted], 2SET [redacted]. [redacted], [redacted].

[redacted]. [redacted], v6[redacted] EDD 86.25[redacted]. [redacted] L38, L39, L40[redacted]. [redacted], [redacted]. [redacted]. [redacted], [redacted], [redacted], [redacted] claim-level [redacted].

[redacted]. [redacted], [redacted]/[redacted]. [redacted], [redacted]. [redacted] 10[redacted], [redacted]. [redacted].

[redacted]. qv6_006[redacted] qv6_007[redacted], [redacted]. [redacted]. [redacted] v7, v8, v9 [redacted], [redacted]/[redacted]. [redacted] “[redacted]”[redacted] “[redacted]”[redacted].

[redacted]. qv6_006[redacted], [redacted], [redacted], [redacted], qv6_007[redacted]. [redacted] L40[redacted] qv6_007[redacted]. [redacted] 1[redacted], [redacted]. [redacted]. [redacted] H/W [redacted], [redacted]·[redacted]. [redacted], [redacted].

[redacted] L40[redacted] “[redacted]”[redacted] “[redacted]”[redacted]. [redacted]. [redacted] EDD[redacted] MRR[redacted], [redacted] “[redacted]?”[redacted]. [redacted], [redacted].

[redacted]. L38 [redacted]. [redacted]·[redacted]·[redacted]. [redacted], [redacted], [redacted], [redacted]. [redacted], [redacted]. [redacted] `ambiguous_identifier_refusal_with_excessive_candidate_summary`[redacted]. [redacted], [redacted].

[redacted]. [redacted], [redacted], [redacted], [redacted] `cost_summary.json`[redacted] `budget_ledger.jsonl`[redacted]. [redacted] dry-run[redacted]. [redacted], [redacted].

[redacted] qv6_007 [redacted]. [redacted]. [redacted] 2[redacted]. [redacted] 0.000002[redacted], [redacted] 0.008310[redacted], [redacted] 0.008312[redacted]. [redacted]. [redacted] 21.46[redacted], [redacted]. [redacted].

[redacted]. [redacted]. [redacted], [redacted], [redacted]. [redacted]. [redacted] 1[redacted] trace[redacted], [redacted] qv6_007[redacted] “[redacted]”[redacted].

[redacted]. [redacted]. [redacted]. [redacted] 3[redacted]. [redacted] 0.000002[redacted], [redacted] 0.007432[redacted], [redacted] 0.005753[redacted], [redacted] 0.013187[redacted]. [redacted].

[redacted]. [redacted] 5[redacted], [redacted] 5[redacted]. [redacted]. [redacted], [redacted]. [redacted] “[redacted]”[redacted]. [redacted]. [redacted], [redacted]

[redacted].

[redacted] hard-stop [redacted]. [redacted], [redacted] qv6_007[redacted], qv6_010[redacted] EDD[redacted], [redacted].

7[redacted] 6[redacted] L45 [redacted] / L46 [redacted] L45 [redacted]

[redacted] trace[redacted], trace[redacted] worker [redacted] `--hard-stop-usd`, `--budget-cap-usd`, `--starting-spent-usd`, `--preflight-case-estimate-usd`[redacted], [redacted] `skipped_preflight_budget`[redacted].

[redacted] 0.001[redacted]. [redacted] 0.01[redacted]. [redacted]. `wrote_rows=0`, `budget_events=1`, `skipped_preflight_budget=1`[redacted], API [redacted]. [redacted] rows[redacted] 10, scoreboard rows[redacted] 1[redacted]. [redacted] L45[redacted], [redacted].

[redacted]. qv6_010[redacted] “[redacted]”[redacted], [redacted], [redacted]. [redacted] abstention accuracy[redacted]. [redacted] `ambiguous_identifier_refusal_with_excessive_candidate_summary`[redacted] evaluate [redacted].

[redacted] L38 [redacted] qv6_004[redacted], qv6_010[redacted].

[redacted]. qv6_007[redacted], [redacted] claim-preservation [redacted]. [redacted] L45[redacted] “[redacted]”[redacted].

L45[redacted]. [redacted] L45[redacted] “dry-run[redacted]”[redacted]. dry-run[redacted] smoke [redacted], [redacted] dry-run[redacted]. [redacted], [redacted] non-dry preflight skip[redacted].

[redacted] L46[redacted]. dry-run[redacted] 0[redacted] zero-cost [redacted], non-dry [redacted]. [redacted] non-dry skip[redacted] `status=blocked`, `budget_gate_all_skipped=true`, `wrote_rows=0`, `observed_cost_usd=0`[redacted]. [redacted] dry-run[redacted] `wrote_rows=1`, `skipped_preflight_budget=0`, `observed_cost_usd=0`[redacted].

[redacted]. [redacted] budget event[redacted] `budget_ledger.jsonl`[redacted] append[redacted]. [redacted] run id, worker id, [redacted], dry-run [redacted], no-judge [redacted], [redacted], hard stop, skip reason, record type, cost basis[redacted].

[redacted] bullet [redacted], [redacted]·[redacted]·[redacted]·[redacted]·[redacted] [redacted]. [redacted] post-refusal tail chars, bullet count, detail keyword hits[redacted]. [redacted] L38 [redacted] qv6_004[redacted], qv6_010[redacted] `ambiguous_identifier_refusal_with_excessive_candidate_summary`[redacted].

L46 [redacted] rows 11, scoreboard rows 1, diagnostic-only rows 10[redacted]. smoke[redacted] worker output 18[redacted]. [redacted] EDD[redacted] “[redacted] EDD[redacted]”[redacted]. [redacted] qv6_007[redacted] claim-preservation, [redacted].

7■ 7■ L47 ■■ ■■ ■■■■ ■■ ■■■ ■■ ■■

■■■■ ■■ ■ ■ ■■ ■■■ ■■■■■ ■■ ■■■. ■■■ ■■■■ ■■ ■■ ■■ ■■■■■, ■ ■■■ ■ ■■■ ■■■ ■■ ■■■. ■■■ qv6_007■ ■■ ■■■ ■■■ ■■■■■ ■■■ ■■■■ ■■■■, ■■ ■■■ ■ ■■■■ ■■ ■■■■ ■■■ ■■■ ■■■ ■■■ ■■■ ■■ ■■■■ ■■■. ■ ■ protected ■■■ ■■ ■■■■ ■■■, ■■ ■■■ ■■ ■■.

■■ ■■■ ■■■ EDD■ ■■■ ■■ ■■■■. ■■ L44■■ ■■ ■■■■ qv6_007 ■■■ groundedness 5■, relevance 5■■ ■■. ■■■ ■■■ ■■■ ■■ ■■■ ■■■ ■■■■■■■ ■■·■■ ■■■ “■■ ■■”■■ ■■■ ■■■. ■ ■■■ ■■ ■■■■ ■■■ ■■■ ■■■ ■■ ■■■■. ■■■ L47■■■ ■■■ ■ ■■■■■, ■■ ■■■ ■■ no-API ■■■■ ■■ ■■■■.

■■■■ `eval/claim_preservation_expectations.json`■ ■■ ■■■ ■■, `scripts/check_claim_preservation.py`■ ■■■ ■■■ ■■ ■■ ■■ ■■■ ■■ ■■ ■■■ ■■■■ ■■. ■■■ ■■■ ■■ ■■■ ■■ ■■■. ■■ CSV ■■ ■■■ ■■■ ■■■, ■■■■■■■■ ■■■ ■■ ■■ ■■ ■■■■■, ■■■■, ■■■■ ■■ ■■■ ■■ ■■■ ■■■ ■ ■■. ■■■ ■■■ ■■ ■■■ ■■■ ■■ ■■■ ■■ ■■■■ ■■ ■■■ CSV ■ ■■■■ ■■■ ■■■. ■ ■■ ■■■ ■■ ■■■■ “■■■■ ■■ ■■■”■ ■■■.

■■ ■■■ ■■■■. L37 ■ v6 ■■■ qv6_007■ ■■ 2■ ■ 0■■ ■■■■. ■■ ■■ ■■■■■ ■■■ ■■■■ ■ ■ ■■■■ ■■■ ■■■■ ■■■. L40■ ■■ ■■ ■■ ■■■ 2■ ■ 1■■ ■■■■. ■■ ■■■ H/W ■■■ ■■■■■, ■■■ ■■■■■ ■■■ ■■ ■■■■ ■■■. L44■ ■■■■■ 2■ ■ 1■■ ■■■■. ■■ ■■■■ 5/5■ ■■■, claim gate■ ■■■ ■■■ ■■■■.

■ ■■■ ■■ ■■■ ■■■■. L40■■ ■■ ■■■ ■■■, ■■■■■■■■ ■■■ ■ ■■■ ■■. ■■■ ■■ ■■ ■■■ ■■■ ■■ “■■■■■ ■ ■■■■■■■ ■■ ■ ■■”■ ■■■ ■■■■ ■■■. ■ ■■■■ L40 ■■■■■ ■■■ ■■■ ■■■■, ■■■■, ■■■■ ■■■■ ■■ ■■ ■■■■■ ■■■ ■■■■■ ■■■■ ■■■■■, ■■ ■■ ■■■ ■■■■■■■■ ■■■■ false friend■ ■ ■ ■■. ■■■ ■■■■ ■■ ■■■ ■■ ■■■■■ ■■.

■■■ ■■■ “■ ■■■”■ ■■■ “■■■ ■■ ■■■ ■ ■■■ ■■■”■. L47■ EDD■ ■■■ ■■■■, ■■■ EDD■ ■■■ ■ ■ ■■■ ■■■■ ■■ ■■ ■ ■■■■ ■■■■. ■■ ■■ ■■ ■■■ ■■ ■■ qv6_007 ■■ ■■■ ■■ ■■ ■■ ■■ ■■ ■ ■■. ■■ ■■■ ■■ ■■ ■■■ evidence-use guard■ ■■, ■■■ ■■■ ■■■■■■■ ■■■ ■■■ ■■ ■■■■■ ■ ■■■ ■■■ ■■ ■■■ ■■■.

7■ 7■ ■■ - L48■■ L76■■, ■■■■ ■■■ ■■ ■ ■■ ■■■

■■■ ■■ ■■ ■■ ■■■ ■■■ ■■, ■■ ■■■ ■■ ■■■. ■■ ■■ ■■■ ■■■ EDD■ ■ ■■■ ■■ ■■■■. ■■ L37 ■■ v6 ■■■ ■■ ■■■ ■ ■■■ ■■■, ■■ ■■■■ ■■■ ■■ ■■■ ■ ■■ ■■■ ■■■. ■■■ ■■ ■■■ ■■ ■■ ■■ ■■■■■■ ■■■, ■ ■■■■■■, ■■ ■■■ ■■■■, ■■ ■■■ ■■■ ■ ■■■■ ■■ ■■ ■■■■ ■■.

■ ■■ ■ ■■■ qv6_007■■■. ■■ claim gate■ ■■ ■■ ■■■■■ ■■■ ■■■■■■■■■■■ ■ ■■ ■ ■■■■ ■■ ■■ ■■■■ ■■■. ■■■■ ■■ ■■ ■■ “■■■■ ■■ ■■” ■■ ■■■ ■■■ ■ ■ ■■■. ■■■ L48■ ■■ ■■ ■■■■. ■■■ ■■ ■■■ ■■■ ■■■■ ■■■■. ■■■ raw HWP ■■■■ “■■■■■ ■ ■■■■■■■ ■■ ■ ■■” ■■■ ■ ■■■■, ■ ■■■ CSV■ ■■ ■■■ ■■■. ■■■ target org■ ■■■■ ■■ ■■ CSV ■■ ■■■ ■■■■ ■■ ■■■. ■■■ ■■ CSV■ ■■■ ■■■ ■■■ ■■ ■■■■■■, ■■■■■ ■■■ ■■■ ■■ ■■ ■■■■■ ■■■.

■ ■■ qv6_007■ L49■■ claim 2/2■ ■■■■, L59■■■ judge■■ 5/5■ ■■■. ■■■ ■■■ ■■ ■■■■ ■■■. L60 ■■ v6 ■■ ■■■ ■■■■ ■■ ■■■ EDD 97.59■■ ■■■■■, qv6_010■ ■■■ ■■■■. ■■■ “■■ ■■■■■■ ■■■ ■■■ ■ ■■”■ ■■■■ ■■■■, ■ ■■■ ■■ ■■■ ■■■■■ ■■ ■■■■. ■■■■ ■■■■ ■■■ ■■■■ ■■■ ■■ ■■■ ■■ ■ ■■■■. ■■■ ■■■ ■■ ■■■■ ■■ ■■/■■■ ■■■ ■■ ■■■, ■■■■■■■■ ■■■■ ■■ ■■ ■■■■■ ■■■■■■ ■■■. L62■■ ■■ ■■■ 181■■ ■■■, ■■ ■■ ■■■ ■■■■.

■■ ■■■ qv6_001■■ ■■■. L64■■ qv6_010■ ■■■■■, UICC ■■■ groundedness■ 4■ ■■■■. ■■■ ■■ ■■■ ■■■■. ■■ ■■■■■■■ ■■■■ rank■ 1■■■. ■■■ ■■■ ■■ ■■■■■■ ■■ ■■■ ■■■■■■■■■ ■ ■■·■■ ■■■ UICC ■■■■■■■ ■■ ■■. ■■ “■■■ ■■■ ■■”■ ■■■ ■■■ ■■■■■■. ■■■ ■■ ■■ ■■ ■■ ■■■■ ■■ ■■, ■■ ■■ UICC■ ■■■■■■■ ■■ ■■■ ■■■■ ■■■ ■■■ ■■ ■ ■■ ■■■ ■■■ ■■■ ■■ ■■■ ■■■. ■■■ ■■■ ■■■■ ■■■ ■■■ ■■■ ■■■ fixture■ ■■ ■■■■. L67■■ UICC■ ■■ 5/5■ ■■, ■■■ ■■ ■■ ■■ 4■■ ■■■ ■■ ■■■■.

[redacted]. qv6_007[redacted] claim gate[redacted] [redacted] judge[redacted] 5/5[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted]·[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted], [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted]·[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] fixture[redacted] 14[redacted] [redacted] [redacted] [redacted].

[redacted] L76 [redacted] v6 [redacted] [redacted] EDD 97.13, coverage 1.0, MRR 1.0, groundedness 5.0, relevance 5.0, abstention accuracy 1.0[redacted] [redacted]. qv6_007 claim gate[redacted] 2/2[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] v6 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] L37[redacted] EDD 86.25[redacted] [redacted] [redacted] v6 [redacted] [redacted], L76[redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] coverage[redacted] MRR[redacted] 1.0[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted], judge 5/5[redacted] claim gate pass[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted]. [redacted] run [redacted] `cost_summary.json`[redacted] [redacted] [redacted] [redacted] [redacted] 0.943855[redacted]. [redacted] 10[redacted] [redacted] 5[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L76 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L76 [redacted] latency[redacted] 20.642[redacted], qv6_006[redacted] qv6_010[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L76[redacted] [redacted] [redacted] [redacted] [redacted] 20.642[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] `prompt_concise_verified_only`[redacted] [redacted] v6 [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted]. L77[redacted] [redacted] latency[redacted] 20.112[redacted] [redacted] 0.53[redacted] [redacted] [redacted]. [redacted] EDD[redacted] 97.13[redacted] 92.25[redacted] [redacted], abstention accuracy[redacted] 1.0[redacted] 0.5[redacted] [redacted]. [redacted] [redacted] qv6_010[redacted]. [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L77 [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L77[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] top_k[redacted] context [redacted] [redacted] [redacted] [redacted] [redacted], L76 [redacted] [redacted] [redacted] [redacted]. qv6_007[redacted] claim 2/2[redacted] [redacted] [redacted], qv6_010[redacted] [redacted] [redacted] [redacted]. groundedness[redacted] relevance[redacted] 5/5[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] 1[redacted] [redacted] 10% [redacted] [redacted] [redacted] [redacted]. 0.53[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] v6[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] v7/v8[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted].

2026-07-07 [redacted] L79: [redacted] [redacted] context [redacted] [redacted] [redacted] [redacted]

L77[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] context [redacted] [redacted]. [redacted] [redacted] [redacted]. qv6_007[redacted] claim 2/2[redacted] [redacted] [redacted], qv6_010[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], groundedness[redacted] relevance[redacted] 5/5[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] v6 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. topk5[redacted] EDD 98.13, [redacted] latency 16.209[redacted], [redacted] 5/5, abstention 1.0, qv6_007 claim 2/2[redacted] [redacted]. [redacted] 0.116122[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. qv6_006[redacted] 31.87[redacted], qv6_010[redacted] 33.42[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

topk8 control[redacted] EDD 98.15, [redacted] latency 16.130[redacted] [redacted] [redacted]. qv6_007 claim[redacted] 2/2[redacted] qv6_010[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] control [redacted] [redacted] [redacted] [redacted]. [redacted] “topk8[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]

[redacted], [redacted] L76[redacted]”[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted].

topk12[redacted] [redacted]. [redacted] claim gate[redacted] 2/2[redacted] [redacted] groundedness[redacted] 4.5, relevance[redacted] 4.875[redacted] [redacted]. [redacted] [redacted] qv6_007[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] judge[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. claim gate[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] claim [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L79[redacted] [redacted] [redacted]. topk12[redacted] [redacted], topk5[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] topk8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] v7/v8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L80: [redacted] [redacted] registry [redacted] [redacted] [redacted]

L79 [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] v5 adversarial frozen [redacted] registry[redacted] `frozen_unscored_candidate`[redacted] [redacted] [redacted], [redacted] [redacted] scored run[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] run [redacted] L30-L36 v5 [redacted] [redacted] [redacted]. L30[redacted] EDD 92.40[redacted] v5 [redacted] [redacted], L31-L34[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] v5[redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. registry[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] validation[redacted] [redacted] [redacted] [redacted]. [redacted] `scripts/build_exposure_registry.py`[redacted] [redacted] v5 draft[redacted] frozen [redacted] [redacted] [redacted] [redacted]. [redacted] qv6_l61, qv6_l65[redacted] [redacted] probe [redacted] `unknown_needs_review`[redacted] [redacted] [redacted] [redacted] [redacted] aggregator[redacted] [redacted].

[redacted] [redacted] aggregate[redacted] [redacted] [redacted] scoreboard[redacted] [redacted] 1[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] L37 v6 [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] registry [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] v5[redacted] [redacted] [redacted] [redacted] [redacted] v7/v8[redacted] [redacted] [redacted].

L80[redacted] registry[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] corpus[redacted] [redacted] [redacted] [redacted] [redacted] source project[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L81[redacted] [redacted] [redacted] source-exposed prompt diagnostic[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L81 [redacted] [redacted] [redacted] [redacted] [redacted]. EDD 97.41, coverage[redacted] MRR[redacted] 1.0, groundedness[redacted] relevance[redacted] 5.0, abstention[redacted] 1.0[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. qv7_002[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], qv7_009[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] qv7_006[redacted]. [redacted] [redacted] “[redacted] [redacted], [redacted] [redacted], [redacted]/[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted].

[redacted] [redacted] qv7_006[redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. abstention accuracy 1.0[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L82[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. `sensitive_or_forbidden_refusal_with_detail_tail`[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L81 [redacted] [redacted] [redacted] EDD[redacted] 97.41 [redacted] qv7_006[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L83[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] qv7_006 [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]: [redacted] [redacted]/[redacted], [redacted] [redacted], [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L83[redacted] EDD[redacted] 57.16[redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted]. L81[redacted] 0.173302[redacted], L83[redacted] 0.011772[redacted], [redacted] run [redacted] [redacted] [redacted] [redacted] [redacted] 1.690291[redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. qv7 [redacted] [redacted] [redacted] 19[redacted], [redacted] [redacted]/[redacted] [redacted] 25~33[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted]. L81[redacted] L83[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 19[redacted]. [redacted] v7 source-exposed [redacted] topk5, topk8, topk12[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. 12[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] worker [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. `topk5_only`, `topk8_only`, `topk12_only`[redacted] [redacted] [redacted] [redacted] shard suite[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] topk5[redacted] topk8[redacted] [redacted] [redacted] [redacted]. topk5[redacted] [redacted] 16.865[redacted], topk8[redacted] 16.794[redacted] L81[redacted] 19.407[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] DR [redacted] [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted].

L86[redacted] `is_abstention`[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. no-API fixture[redacted] 21[redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. topk5[redacted] EDD 97.99, [redacted] 16.865[redacted], [redacted] 25.85[redacted]. topk8 control[redacted] EDD 98.00, [redacted] 16.794[redacted], [redacted] 23.46[redacted]. topk12[redacted] EDD 95.51, [redacted] 19.922[redacted], [redacted] 36.95[redacted] groundedness/relevance[redacted] 4.667/4.889[redacted] [redacted] [redacted] [redacted] [redacted]. topk5[redacted] [redacted] [redacted] topk8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. topk8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] control[redacted] source-exposed [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L85 shard [redacted] [redacted] [redacted] [redacted] 0.540602[redacted], [redacted] run [redacted] [redacted] [redacted] [redacted] 2.230893[redacted]. [redacted] L84 [redacted] sweep[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] local summary[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] shard[redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L86[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. qv7_009[redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] judge[redacted] [redacted] groundedness 5[redacted], relevance 5[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. `plain_language_answer_over_structured`[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. fixture[redacted] 22[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], 22[redacted] [redacted] [redacted] [redacted].

[redacted] L81[redacted] L85 [redacted] [redacted] [redacted] [redacted] qv7_009[redacted] [redacted] [redacted]. L81 baseline [redacted] 1142[redacted], 25[redacted], [redacted] 20[redacted] judge[redacted] [redacted] 5/5[redacted]. L85 topk8 control[redacted] 1294[redacted], 42[redacted], [redacted] 16[redacted] [redacted] 5/5[redacted]. topk12[redacted] [redacted] [redacted] 1448[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. EDD[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. qv7_006[redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted], qv7_009[redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] qv7_009[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] L77[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] format probe[redacted] [redacted] [redacted]. [redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L87[redacted] [redacted] [redacted] qv7_009[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted]” [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] `questions_v7_l88_plain_language_probe.json`[redacted] [redacted] [redacted] [redacted], [redacted] diagnostic-only[redacted] [redacted]. [redacted] [redacted] source[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L88[redacted] prompt [redacted] [redacted]. [redacted] [redacted]. default[redacted] judge 5/5[redacted] 1228[redacted], 47[redacted], [redacted] 29[redacted] [redacted] [redacted] [redacted] [redacted]. `strict_evidence`[redacted] `concise_verified`[redacted] [redacted] [redacted] [redacted]. `report_ready`[redacted] [redacted] [redacted] [redacted]. [redacted] 644[redacted], 8[redacted], [redacted] 5[redacted] [redacted] [redacted] judge[redacted] 5/5[redacted]. [redacted] [redacted] 50.94[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]” [redacted] [redacted] [redacted].

[redacted] L89[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] “[redacted] [redacted]”, “[redacted] [redacted]” [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L88[redacted] [redacted] [redacted]. [redacted] 605[redacted], 9[redacted], [redacted] 5[redacted] [redacted] `plain_language_answer_over_structured`[redacted] [redacted]. judge[redacted] [redacted] groundedness 5[redacted], relevance 5[redacted]. [redacted] 16.7[redacted] L88 default 18.8[redacted] [redacted], L88 report_ready 50.94[redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] L89[redacted] false abstention[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L90[redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] bullet[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. fixture[redacted] 23[redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. qv7_009 [redacted] [redacted] [redacted] plain-language hint[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] v7[redacted] [redacted] “[redacted]”[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] source-exposed regression[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted]. L88 prompt sweep[redacted] 0.111066[redacted], L89 [redacted] [redacted] 0.017385[redacted]. [redacted] run [redacted] cost summary[redacted] [redacted] [redacted] 2.359344[redacted]. L84 timeout[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L89[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] v7 source-exposed [redacted] 12[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] source[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted]. qv7_009[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted]. L91[redacted] EDD 97.82, coverage/MRR 1.0/1.0, groundedness/relevance 5.0/5.0, abstention accuracy 1.0[redacted]. 12[redacted] [redacted] `answer_quality_issues=[]`[redacted]. qv7_009[redacted] 718[redacted], 9[redacted], [redacted] 5[redacted] [redacted] judge[redacted] 5/5[redacted]. L87[redacted] [redacted] `plain_language_answer_over_structured`[redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L91 [redacted] [redacted] 17.613[redacted] L81[redacted] [redacted], L85 topk8 control[redacted] 16.794[redacted] [redacted] [redacted]. [redacted] qv7_006[redacted] 28.81[redacted] [redacted]. plain-language hint[redacted] qv7_006[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. qv7_009 [redacted] [redacted] [redacted] hint[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] v7 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] source-exposed[redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] hint trigger[redacted] [redacted] [redacted] [redacted] [redacted] false-positive [redacted], qv7_006 [redacted] [redacted] tail [redacted] [redacted] [redacted] [redacted].

L91 [redacted] [redacted] 0.171454[redacted], [redacted] run [redacted] [redacted] [redacted] [redacted] 2.530798[redacted]. [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] claim[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L91[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. plain-language hint[redacted] “[redacted] [redacted]” [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] RFP [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted]? [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L92[redacted] [redacted] [redacted] [redacted] trigger [redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] plain-language hint[redacted] [redacted] [redacted] [redacted]. fixture[redacted] [redacted] [redacted] [redacted]. [redacted] “[redacted] [redacted]. [redacted] [redacted]”[redacted] [redacted] hint[redacted] [redacted], [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] hint[redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] 25/25 [redacted]. [redacted] EDD[redacted] [redacted] [redacted] [redacted]. [redacted] L89-L91[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L91[redacted] [redacted] [redacted] qv7_006[redacted] [redacted] [redacted]. [redacted] [redacted] 179[redacted] [redacted] [redacted]. “[redacted] [redacted]/[redacted], [redacted] [redacted], [redacted] [redacted]/[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] trace[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 28.49[redacted] [redacted] [redacted] [redacted] [redacted],

[redacted] guard[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L93[redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] preempt guard[redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted], [redacted] [redacted], [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] qv7_009[redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] preempt[redacted] [redacted] [redacted]. fixture[redacted] 27[redacted] [redacted] [redacted] [redacted].

L94[redacted] qv7_006[redacted] qv7_012 [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. qv7_006[redacted] 3.27[redacted], qv7_012[redacted] 0.86[redacted] generation trace[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], generation cost[redacted] 0[redacted]. [redacted] [redacted] embedding [redacted] [redacted] 0.000004[redacted]. EDD[redacted] 60.0[redacted] [redacted], [redacted] [redacted] abstention [redacted] judge[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] EDD[redacted] [redacted] latency, cost, [redacted] [redacted] [redacted] [redacted].

[redacted] L95[redacted] v7 source-exposed 12[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. EDD 98.54, [redacted] latency 14.44[redacted], coverage/MRR 1.0/1.0, groundedness/relevance 5.0/5.0, abstention accuracy 1.0[redacted]. 12[redacted] [redacted] [redacted] [redacted] [redacted]. qv7_006[redacted] 0.35[redacted], qv7_012[redacted] 0.31[redacted] [redacted], qv7_009[redacted] 651[redacted], 9[redacted], [redacted] 5[redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L95[redacted] [redacted] source-exposed [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted]. [redacted] [redacted] v7 source [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] guard[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L95 [redacted] [redacted] 0.156359[redacted], [redacted] run [redacted] [redacted] [redacted] [redacted] 2.687161[redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] exposed v7[redacted] [redacted] [redacted] [redacted] preempt trigger[redacted] false-positive[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L95[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. preempt [redacted] [redacted] [redacted] [redacted] [redacted], “[redacted] [redacted] [redacted] [redacted]” [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted], [redacted] [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L96[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted].[redacted] no-API fixture[redacted] [redacted]. [redacted] [redacted] preempt[redacted] [redacted] [redacted].

[redacted] 28/28 [redacted]. qv7_006[redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], L95[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] guard[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. L95[redacted] EDD 98.54[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] source-exposed [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. L95[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]·[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L97[redacted] [redacted] [redacted] [redacted] fixture[redacted] [redacted]. [redacted] 33[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] marker[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] 33/33[redacted] [redacted].

[redacted] [redacted] qv8[redacted] [redacted]. [redacted] Codex[redacted] [redacted], “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] holdout[redacted] [redacted]. [redacted] `questions_v8_deadline_semi_fresh_diagnostic_frozen.json`[redacted] [redacted], [redacted] diagnostic-only[redacted] [redacted]. [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L98 qv8 [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted]. coverage[redacted] MRR[redacted] 1.0[redacted], judge[redacted] 5/5[redacted]. [redacted] [redacted] [redacted] 17.645[redacted], 31[redacted]·27[redacted]·25[redacted] tail[redacted] [redacted] [redacted]. raw EDD[redacted] 95.81, [redacted] [redacted] 0.126294[redacted]. [redacted] abstention accuracy[redacted] 0.8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] qv8_a12[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L99[redacted] [redacted]/[redacted] [redacted]/[redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] story refusal[redacted] [redacted] [redacted] [redacted] [redacted], fixture[redacted] 34/34[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] qv8 EDD[redacted] 97.81, abstention

accuracy■ 1.0■ ■■.

■■■ ■■■ ■■■■. L99■ ■■ ■■ ■■■ ■■■ ■■ ■■■■. ■■■■■ L98 raw■ L99 recompute■ ■■■ ■■ ■■. ■■■ strict scoreboard■ ■■■ L37■ 86.25■. L95■ ■■ ■■ source-exposed ■■■■■, qv8■ ■■ ■■■■ ■■ ■■■ deadline diagnostic■■. ■■ ■ ■ ■■ ■■■ ■■ ■■■ ■ ■■■ ■■ ■■■, qv8 tail latency■ ■■■ mixed question■ ■■■■ ■■■ ■ ■■ ■■■ ■■■.

L99■■ ■■■■ ■■ ■■ ■■ ■■■ ■■■■ ■■■■■. qv8■ ■■ ■■■■ judge ■■■ ■■ ■■■ ■■■■, ■■ ■■■ ■■■ 20■■■ 30■■■■ ■■. ■■■ ■■■■ ■■■ ■ ■■■■ ■■■ ■■■ ■■ ■■■, ■■ ■■■ ■■■ ■■ ■■■ ■■ ■■■. ■■■ trace■ ■■ ■■ ■■ ■■■ ■■ 25.026■■■, ■■ ■■■■ ■■■ ■■■ ■■ ■■ ■■■■■. ■■ ■■■ ■■ ■■ ■■ ■ ■■■■■ ■ ■■ ■■■ ■■■ ■■ ■■ ■■ ■■■■ ■■■.

■■■ qv8■■ ■■ ■■ ■■■ ■■ ■■ ■■ ■■■■ ■■■■. ■ ■■■ ■■■■ ■ ■■■■ ■■■ ■■■■■■ ■■■■. ■■ ■■ ■■■■ top8 ■■, top5, ■■ ■■■■■ ■■■■. ■■■ ■■■ ■■■■. top8 ■■■ EDD 97.11, ■■ 20.736■■■, top5■ EDD 98.00, ■■ 16.798■■ ■■■■. ■■■ ■■ ■■■■■ abstention accuracy■ 0.0■■ ■■■■ ■■■ 23.280■■ ■■ ■■■. ■■ ■■ ■■■ ■■■ ■■■ ■■■ ■■ ■■■■ ■■■ ■■ ■ ■■■ ■ ■■■■.

■■■ ■■■ ■■ top5■ ■■■■ ■ ■■■ ■■. ■■ ■■ ■■■ ■■ ■■ ■■■■, ■ ■■■ ■■■ ■■ ■■■■■ ■ ■■. ■■■ qv8 ■■ 10■■■ top5■ ■■ ■■■. ■■ qv8 top5■ ■■ ■■■ 14.284■■■ ■■■ coverage, MRR, groundedness, relevance■ ■■ ■■■■. ■■■ 0.103174■■ ■■■ ■■■ ■■■ ■ ■■■. ■■■■■ ■■ ■■ ■■■ ■■■.

■■■ ■■■ ■■■ ■■ ■■■. qv8_a12■ ■■■ ■■■ ■■■ ■■ ■■■ ■■ ■ ■■■■■, qv8_a11■■ ■■ ■■■■ ■■■ ■■■■■■ ■■ ■■■ ■■■■ ■■■■. ■■■ "■■■ ■■■■ ■■■ ■ ■■■■"■■ ■■■ ■■■, ■ ■■ ■■■■ ■■■■ ■■■ ■■ ■■■■. ■■ abstention ■■ ■■■■■ ■■■ ■■■ ■■■ ■■ ■■■. ■ top5■ ■■■ ■■■■, unsupported award-result ■■■■ ■■ ■■■ ■■■■.

■■■ ■■■ ■■■■■. top5■ ■■ ■■■ ■■■. ■■■ ■■ ■■■■■ ■■■■ ■■■. qv8 ■■■■ ■■ ■■■ 17.645■■■ 14.284■■ ■■■■, L99■ recompute ■■■■ abstention accuracy 1.0■ ■■■■ ■■■ ■■■■. ■■■ ■■ ■ ■■ top5■ ■ ■■■■■ ■■ ■■■ qv8_a11 ■■ "■■■ ■■ ■■ ■■■ ■■■■ ■■"■■ ■■■■ ■■■ ■■■ ■■■ ■■■ ■■ ■■ ■■■.

■■ ■■■■ ■■ ■■■■■ ■■■■. ■■■ ■■■ ■■ ■■■ ■■■■ ■■■ ■■. ■■■ ■■■ ■■■■ ■■ ■■■■■ ■■ ■■ ■■■. ■■ ■■■■ ■■■ ■■ ■■■ ■■ ■ ■■, top_k ■■■ ■■■■ ■■■ ■■■ ■■ ■■ ■■ ■■ ■■■ ■■■ ■■ ■ ■■. ■■■ ■■ ■■■ qv8_a11 ■■■ unsupported-result ■■ ■■ probe■ ■■■, ■ ■■■ ■ ■■■ adaptive top_k■ ■■ top5■ ■■ ■■■■ ■■■.

L101■■ ■■■■ ■■■ qv8_a11 ■■ ■■■■■. ■■■ ■■■. ■■■■■■■ ■■■ ■■■■, coverage■ MRR■ ■■■ ■■■. ■■■ ■■ ■■■ ■■■■. top8 ■■■■■ "■■■ ■■■■ ■■■ ■ ■■■■"■ ■■ ■■ ■, ■ ■■■ ■ ■■■ ■■ ■■■■. ■■ top5■■■ ■■■■ ■■■■■ ■■ ■■, ■■■■ ■■ ■■ ■■■■ ■■ ■■■■. ■■ ■■■■ ■■■ ■■■■■■ ■■■ ■■■ ■■ ■■ ■■■■, ■■■ ■■ ■■■ ■■■■■■■ "■■■■ ■■ ■■■■ ■"■ ■■■■.

■■■ ■■■■ ■ ■■■ ■■ ■■■. ■ ■■■ ■■■■■ ■■ ■■ ■■■ ■■■. ■■■■ "RFP■ ■■■■■ ■■■ ■■■ ■■ ■■■■ ■■■■ ■■■■?"■■ ■■■■■. ■■■ ■■■■ ■■■ ■■ ■■■ ■■■■■■ ■■. ■■■ ■■ ■■ ■■■ ■■. ■■■ ■■■■ ■■■ ■■■■■■ ■■ ■■■ ■■■ ■■■ ■■. ■ ■■■ no-API fixture■ ■■ ■■■■.

■■ fixture■ 39■ ■ 1■■ ■■■■. ■■■■ "■■■■ ■■■■ ■■ ■■■ ■■■■ 90%, ■■■■ 10%■■■. ■■ ■■ ■■■■■ ■■ ■■■ ■■■ ■■■■ ■■■ ■ ■■■■." ■■ ■■■ ■■ ■■■ ■■■■. ■■ ■■ ■■■ ■■■ ■■■ granularity ■■■■. ■■■ ■■■■ ■■■■ ■■ ■■■■ ■■■■ ■■■ ■■ ■■■■ ■■■■■ ■■■■. ■■ ■■■ 39/39■ ■■■■.

■ ■■ qv8_a11■ ■■ L102 ■■ probe■ ■■■. ■■■ ■■■■■■. top5 ■■■■ abstention■ true■ ■■■, latency■ 2.44■■■. ■■■ ■■ embedding ■■■■ 0.000002■■■■. ■■ abstention ■■■■ EDD 60.0■ ■■■■ ■ ■■ ■■■■, ■ probe■ ■■■ ■■■ ■■■ ■■ ■■ ■■■ ■■■■.

■■■■■ L103■■ qv8 ■■ 10■■■ ■■ ■■■. ■■■ qv8 ■■ ■■■■■ ■■ ■■■. EDD 98.71, coverage/MRR 1.0/1.0, groundedness/relevance 5.0/5.0, abstention 1.0, ■■ ■■ 13.656■■■. L101■ top5■ ■■■■ a11■ ■■ EDD 96.57,

abstention 0.8[redacted]. [redacted] a11[redacted] 0.32[redacted], [redacted]. [redacted] 0.095118[redacted].

[redacted]. L103[redacted] qv8 [redacted]. [redacted] qv8[redacted] strict validation[redacted]. [redacted] top5 [redacted]. [redacted] qv8[redacted]/[redacted] guard[redacted]. [redacted] L103[redacted] qv8_a12[redacted] 22.06[redacted], [redacted] preempt[redacted].

L103 [redacted] qv8_a12[redacted]. L99[redacted] "[redacted]"[redacted]. [redacted]. [redacted]. L103[redacted] qv8_a12[redacted] 22.06[redacted]. [redacted]. [redacted], [redacted], [redacted] "[redacted]"[redacted].

[redacted], [redacted], [redacted]. [redacted]. "[redacted]", "[redacted]" [redacted]. [redacted] no-API fixture[redacted]. [redacted], [redacted]. fixture[redacted] 41/41[redacted].

L104[redacted] qv8_a12 [redacted] targeted probe[redacted]. latency[redacted] 2.12[redacted], generation[redacted]. [redacted] 0.000002[redacted]. [redacted] EDD 60.0[redacted]. [redacted] "[redacted]"[redacted]. [redacted].

[redacted] L105[redacted] qv8 [redacted] 10[redacted]. [redacted] qv8 [redacted]. EDD 99.13, coverage/MRR 1.0/1.0, groundedness/relevance 5.0/5.0, abstention 1.0, [redacted] 11.813[redacted]. L103[redacted] 13.656[redacted], a11[redacted] a12[redacted] 0.29[redacted], 0.26[redacted]. [redacted] 0.081801[redacted], [redacted] 3.305127[redacted].

[redacted]. L105[redacted] qv8 [redacted]. [redacted], [redacted], [redacted]. [redacted] qv8[redacted]. [redacted] 99.13[redacted] strict validation[redacted]. [redacted]. [redacted] qv8 [redacted]. [redacted] qv8[redacted]/[redacted], [redacted]. [redacted] EDD [redacted] field-level scoring[redacted].

L105[redacted] qv8 [redacted]. EDD 99.13, abstention 1.0, [redacted] 11.813[redacted]. [redacted], [redacted]. qv8[redacted]. [redacted] 99[redacted]. [redacted], guard[redacted].

[redacted]. [redacted] "[redacted]"[redacted]. [redacted], [redacted]. [redacted], [redacted]. [redacted] marker[redacted], [redacted].

[redacted] no-API fixture[redacted] 49/49[redacted]. [redacted]. [redacted]. [redacted]. [redacted]. [redacted], [redacted], [redacted].

[redacted] field-level scoring[redacted]. [redacted], [redacted] EDD[redacted] abstention [redacted]. [redacted]. [redacted]. [redacted] scorer[redacted] fixture 4[redacted]. [redacted], [redacted] 4/4[redacted]. [redacted] `over_refusal`[redacted] `unsafe_exposure`[redacted].

[redacted] EDD [redacted]. [redacted]. L106[redacted] guard [redacted], L107[redacted] qv8[redacted] draft [redacted] 7[redacted]. [redacted] draft[redacted] strict validation[redacted]. [redacted] corpus[redacted], [redacted]. [redacted]. qv8[redacted], L108[redacted] non-qv8 [redacted], field-level [redacted] EDD [redacted].

L106-L107[redacted] non-qv8 [redacted]. L108[redacted] 8[redacted]. [redacted] [redacted] [redacted]. [redacted], [redacted] [redacted] guard[redacted] field scorer[redacted] [redacted] [redacted].

[redacted] L108 [redacted]. false abstention rate[redacted] 0.6[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. q001[redacted] 130,000,000[redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. q002[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. q004[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted].

[redacted] L109[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], field expectation[redacted] [redacted] [redacted] [redacted] scored diagnostic copy[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] marker[redacted] ????[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], marker[redacted] [redacted] UTF-8 [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], marker[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] post-trim guard[redacted]. q002[redacted] “[redacted] [redacted], [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted]/[redacted] guard[redacted] “[redacted] [redacted], [redacted]” [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L110[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. “[redacted] [redacted] [redacted] [redacted]” [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] fixture[redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] guard fixture[redacted] 51/51[redacted] [redacted].

L110[redacted] false abstention 0.0[redacted] [redacted]. q002[redacted] [redacted] [redacted]/[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] q004[redacted] field_score[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] scorer[redacted] [redacted] [redacted] “[redacted] [redacted]/[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] q004[redacted] private side[redacted] `refuse`[redacted] [redacted] `withhold`[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted]”[redacted], [redacted] [redacted] [redacted] [redacted] [redacted].

L111[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] “[redacted] [redacted]: [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted], scorer[redacted] [redacted] [redacted]:[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] 010 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] generic [redacted] leak marker[redacted] [redacted], `010-`, `@gmail`, `@naver`, `@daum` [redacted] [redacted] [redacted] [redacted].

[redacted] L112[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. coverage/MRR[redacted] 1.0/1.0, abstention accuracy[redacted] 1.0, false abstention[redacted] 0.0, field_score issue[redacted] [redacted]. [redacted] [redacted] 7.184[redacted], [redacted] [redacted] 0.040586[redacted]. [redacted] [redacted] [redacted] 3.511722[redacted] [redacted] hard stop 4.00[redacted] [redacted] [redacted]. [redacted], L112[redacted] EDD 60.0[redacted] no-judge[redacted] strict EDD[redacted] [redacted] [redacted] [redacted]. groundedness[redacted] relevance[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] EDD[redacted] [redacted] false abstention, field_score issue, latency, cost[redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. RAG [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted]”[redacted] “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] abstention[redacted] [redacted] [redacted] [redacted]. [redacted] field-level scoring[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] 8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] profiling [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

L112[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] 8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] L112[redacted] [redacted] 8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] v9 [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] L113[redacted] [redacted] [redacted] [redacted]. L112[redacted] [redacted] [redacted] 7.184[redacted] [redacted] top5, top8, top12[redacted] [redacted], top5[redacted] 8.598[redacted], top8 [redacted] [redacted] [redacted] 8.369[redacted], top12[redacted] 14.890[redacted]. false abstention[redacted] 0.0[redacted] field_score issue[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] top12[redacted] q002[redacted] 71.44[redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] L112 top8[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], top_k[redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. v9[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] source/corpus[redacted] [redacted] [redacted] [redacted] [redacted] blind holdout[redacted] [redacted]. [redacted] `source_inspected_v9_mini_diagnostic_frozen`, `diagnostic_only`[redacted] [redacted]. L114 [redacted] [redacted] EDD 93.42,

coverage/MRR 1.0/1.0, groundedness/relevance 5.0/5.0, abstention accuracy 0.5, false abstention 0.0, [redacted] 14.934[redacted]. [redacted] abstention[redacted], q005 [redacted].

[redacted]. [redacted] `is_abstention`[redacted]. [redacted] marker[redacted] no-API fixture[redacted]. guard fixture[redacted] 52/52[redacted], [redacted] L114 [redacted] L115[redacted] EDD 98.42, abstention accuracy 1.0[redacted]. [redacted] L114 raw[redacted] L115 measurement correction[redacted].

[redacted] q001[redacted]. [redacted] judge[redacted] 5/5[redacted], [redacted] “20[redacted]”[redacted] “[redacted]”[redacted]. [redacted] CSV[redacted], RAG [redacted] context[redacted] [redacted] metadata[redacted]. [redacted] CSV-only metadata[redacted], RAG context[redacted].

[redacted]. [redacted], L113[redacted]. [redacted], L115[redacted] strict [redacted]. [redacted] v9[redacted] [redacted], source-field [redacted].

[redacted]. [redacted] DB [redacted], [redacted] RFP RAG [redacted].

[redacted] corpus[redacted] RAG [redacted] CSV metadata[redacted]. [redacted].

[redacted] v10 [redacted] 12[redacted]. [redacted]/ISP/[redacted]/[redacted], [redacted]/[redacted]/[redacted], [redacted]. [redacted] RFP [redacted] CSV[redacted]. [redacted] CSV[redacted] context[redacted]. [redacted] q001[redacted] “[redacted]?”[redacted].

[redacted] v10[redacted] EDD [redacted]. `rfp_rag_user_intent_gate_v10.0.0`[redacted] 12~16[redacted].

[redacted]. RFP [redacted] “[redacted]?”, “[redacted]?”, “[redacted]?”, “[redacted]?”, “[redacted]?”[redacted]. [redacted] v10 [redacted] metadata[redacted].

[redacted] `rfp_question_bank.md`[redacted] `rfp_question_bank.csv`[redacted]. [redacted] 95[redacted]. [redacted] L116[redacted] 12~16[redacted].

[redacted] 95[redacted], [redacted], [redacted], [redacted], [redacted]/[redacted], API/[redacted]/UI/QA/DevOps [redacted], top-k [redacted] latency [redacted] RAG [redacted].

[redacted] 95[redacted] [redacted] 68[redacted] [redacted], 27[redacted] [redacted]. [redacted] [redacted] 68[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] RAG [redacted], [redacted] [redacted] CSV[redacted] metadata [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] top-k [redacted] filter on/off [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] EDD[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. 12~16[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] lane[redacted] [redacted], lane[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] `eval/question_bank_gap_review_20260707.md`, `.json`, `.csv`[redacted]. [redacted] [redacted] [redacted] [redacted] lane[redacted] [redacted] [redacted] [redacted], “[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] seed[redacted] [redacted] [redacted].

68[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] 4[redacted] [redacted]. `Q092`[redacted] metadata filter on/off [redacted] [redacted] [redacted] [redacted] [redacted], `Q093`[redacted] latency [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. `Q094`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], `Q095`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], RAG [redacted] [redacted] [redacted] [redacted].

[redacted] 64[redacted] 16[redacted] [redacted] [redacted] [redacted]. A[redacted] [redacted] [redacted] [redacted], B[redacted] [redacted]/[redacted] [redacted], C[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], D[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] 4[redacted] [redacted] [redacted] [redacted] JSON[redacted] [redacted], smoke[redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] A [redacted]. `Q033~Q035`[redacted] [redacted] A[redacted] [redacted] [redacted], [redacted] metadata [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], D [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] 16[redacted] [redacted] [redacted], A [redacted] `Q033~Q035`[redacted] [redacted] subtotal[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

B[redacted] C[redacted] D[redacted] [redacted] [redacted]. B[redacted] `Q038~Q050`[redacted] “[redacted] [redacted]”[redacted] “[redacted] RFP”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] seed[redacted] [redacted] [redacted]. C[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. D[redacted] `Q090~Q091`[redacted] top-k [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] trace[redacted] citation trace[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] “64[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]”[redacted]. [redacted] `eval/question_bank64_shards_20260707.md`, `.json`, `.csv`, `.manifest.json`[redacted], [redacted] [redacted] `eval/parallel_runs/20260707_144929_L118-question-bank-64-shard-preparation-and-team-review`[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] shard[redacted] runnable manifest[redacted] [redacted] [redacted]. seed [redacted], seed turn, [redacted] [redacted] [redacted] metadata[redacted] [redacted] [redacted], trace[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

2026-07-07 [redacted] L119 - 64[redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]

L118[redacted] 64[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] RAG [redacted], [redacted] [redacted] CSV metadata[redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] top-k trace[redacted] [redacted] export[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] runnable manifest[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] 100[redacted] [redacted] [redacted] seed [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] ERP[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] seed catalog[redacted] [redacted].

[redacted] 64[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] EDD [redacted] 49[redacted], metadata[redacted] trace[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 15[redacted]. `Q001~Q012`[redacted] `Q065`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] sidecar[redacted] [redacted] [redacted]. `Q090~Q091`[redacted] [redacted] [redacted]. [redacted] [redacted] “[redacted] chunk[redacted] [redacted]”, “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] trace wrapper[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] `blocked_until_trace_wrapper`[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. A[redacted] keyword sidecar[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] keyword hit[redacted] [redacted] gold[redacted] [redacted] [redacted] [redacted] preflight[redacted] [redacted] [redacted]. B[redacted] `Q038~Q045`[redacted] [redacted] seed[redacted] [redacted] [redacted] [redacted] [redacted]. “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] seed [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. C[redacted] [redacted] [redacted].

D[redacted] `Q090~Q091` trace wrapper[redacted]/[redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] worker [redacted] [redacted] [redacted] [redacted], [redacted] JSON [redacted] [redacted] [redacted] smoke[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] `worker_outputs/raw_reviews`[redacted] [redacted], [redacted] [redacted] `parallel_team_worker_output.v1` [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] smoke[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] `eval/question_bank64_runnable_manifest_20260707.json`, `.csv`, `.md`, `.manifest.json`[redacted]. [redacted] “62[redacted] [redacted], 2[redacted] trace wrapper [redacted] [redacted], [redacted] [redacted] 64[redacted] [redacted] EDD[redacted] [redacted] [redacted] [redacted]”[redacted]. [redacted] [redacted] metadata sidecar runner[redacted] selected-project runner[redacted] [redacted], blocked trace [redacted] wrapper[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

2026-07-07 [redacted] L120 - [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]

L119[redacted] “metadata sidecar[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] API[redacted] [redacted] [redacted], [redacted] `data_list.csv`[redacted] L119 manifest[redacted] [redacted] [redacted] [redacted] runner[redacted] [redacted]. [redacted] `scripts/run_metadata_sidecar.py`[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. `Q070`[redacted] [redacted] [redacted]. [redacted] [redacted] `Q070`[redacted] “[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted]. [redacted] [redacted] sidecar key[redacted] [redacted] [redacted] [redacted] `Q006`[redacted] [redacted]/[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] runner[redacted] case id[redacted] key id[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] runner[redacted] [redacted] `preflight_answer_key_ref`[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] 17[redacted] [redacted] ready[redacted] [redacted]. sidecar readiness score[redacted] 1.000[redacted]. [redacted] [redacted] [redacted] “RAG[redacted] [redacted] [redacted]”[redacted] [redacted] [redacted]. “[redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] 6[redacted] exact sidecar[redacted] [redacted] [redacted] [redacted], [redacted]/[redacted]/[redacted]/[redacted]/[redacted]/ODA/ERP/AI/[redacted] [redacted] keyword discovery[redacted] 11[redacted] [redacted] [redacted]. [redacted] [redacted] row-level evidence[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] runner[redacted] [redacted] [redacted]. [redacted] strictness [redacted] [redacted] [redacted], [redacted] `Q070 -> Q006`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] JSON[redacted] CSV[redacted] MD[redacted] `sidecar_key_id`[redacted] [redacted], MD[redacted] “Intentional Sidecar Aliases” [redacted] [redacted]. [redacted] `Q070`[redacted] [redacted] [redacted] [redacted] Q006[redacted] [redacted] [redacted] [redacted], follow-up domain expansion[redacted] Q006 [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] “metadata [redacted] [redacted] [redacted] [redacted]”[redacted]. [redacted] [redacted] [redacted] EDD[redacted] [redacted] [redacted] [redacted]. L120[redacted] 1.000[redacted] [redacted] [redacted] [redacted] sidecar [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] selected-project runner[redacted]. `Q038~Q053`, `Q067`, `Q069~Q073`, `Q077~Q089`[redacted] “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] seed [redacted] [redacted], [redacted] turn[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

2026-07-07 [redacted] L121 - “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted]

metadata [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted]”[redacted] “[redacted] [redacted] [redacted]”[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] L121[redacted] selected-project batch builder[redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted]/[redacted] [redacted] `Q038~Q045`[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] `Q046~Q053`[redacted] [redacted] ERP[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] LMS, [redacted] ISP/ISMP [redacted] [redacted] `Q073`[redacted] [redacted] [redacted] ISP seed[redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. [redacted] `questions_l121_selected_project_multiturn_primary.json`[redacted]. [redacted] [redacted] [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted] [redacted]”[redacted] [redacted], [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] 71[redacted] [redacted] [redacted]. [redacted] [redacted] `questions_l121_selected_project_resolved_primary.json`[redacted]. [redacted] seed[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] 35[redacted] [redacted] [redacted]. [redacted] smoke[redacted] [redacted] [redacted] resolved[redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] multiturn[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted]

[redacted].

[redacted]. [redacted] seed[redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] “[redacted] [redacted] [redacted] [redacted]”[redacted] [redacted]. [redacted] seed[redacted] ADD [redacted] [redacted] secondary variant[redacted] [redacted], [redacted] [redacted] [redacted] sparse-field [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted]/persona [redacted] Q067~Q072[redacted] Q073 [redacted] [redacted], resolved one-turn[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. [redacted] seed [redacted] [redacted] [redacted] capacity [redacted] [redacted], [redacted] worker[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] JSON [redacted] [redacted] [redacted] [redacted], [redacted] PowerShell[redacted] UTF-8 [redacted] [redacted] [redacted] [redacted] [redacted]. Python [redacted] [redacted] `Get-Content -Encoding UTF8 | ConvertFrom-Json`[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted], [redacted], [redacted] JSON [redacted] [redacted] [redacted].

[redacted] [redacted] “selected-project [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted]”[redacted]. primary 35[redacted], secondary 16[redacted] [redacted], cheap smoke[redacted] 35[redacted], [redacted] [redacted] [redacted] 71[redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] resolved smoke[redacted] [redacted] [redacted], secondary technical variant[redacted] [redacted] [redacted] sparse-field not-found guard[redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

2026-07-07 [redacted] L122 - [redacted] [redacted] [redacted] [redacted] [redacted]

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. `Q090~Q091`[redacted] [redacted] [redacted] [redacted]. [redacted] trace[redacted] chunk [redacted] [redacted] [redacted] [redacted]. `Q001~Q012`, `Q065`, `Q070` [redacted] [redacted] [redacted] [redacted] [redacted] corpus [redacted] [redacted] sidecar [redacted]. `Q070`[redacted] [redacted] key[redacted] [redacted] `Q006` [redacted] [redacted] [redacted] [redacted] [redacted], case id[redacted] [redacted] key[redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], L121[redacted] [redacted] sparse-field guard[redacted] [redacted] [redacted]. [redacted] ADD [redacted] [redacted] seed[redacted] [redacted] [redacted] [redacted] [redacted] `Q046~Q053`[redacted]. [redacted] CSV[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] sparse[redacted] [redacted]. 31[redacted] [redacted] [redacted] [redacted] 4[redacted] [redacted] 27[redacted] [redacted] [redacted] [redacted]. [redacted] “[redacted] API, DB, [redacted], [redacted] [redacted] [redacted]”[redacted] [redacted] [redacted] [redacted].

[redacted] raw HWP[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. 31[redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. API, [redacted] [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] ADD [redacted] raw [redacted] [redacted] [redacted] [redacted] [redacted], CSV[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. “[redacted] [redacted]”[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. secondary variant 16[redacted] [redacted] `ordinary_edd_candidate=true`[redacted] `EDD` route[redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] L122[redacted] EDD [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] `scripts/run_sparse_field_guard.py`[redacted] L122 [redacted] `analysis/sparse_field_guard` [redacted] [redacted]. [redacted] [redacted]. [redacted] secondary row[redacted] scoreboard[redacted] [redacted] [redacted] [redacted] aggregator[redacted] registry [redacted] [redacted] [redacted] [redacted], raw [redacted] [redacted] [redacted] sparse[redacted] [redacted] seed[redacted] [redacted] [redacted] [redacted].

2026-07-07 [redacted] L123 - [redacted] [redacted] [redacted] [redacted]

L122[redacted] [redacted] [redacted] [redacted] secondary variant[redacted] [redacted] [redacted], [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. `scripts/aggregate_parallel_eval.py`[redacted] secondary variant [redacted] [redacted] [redacted]. [redacted] `secondary_variants`[redacted] `secondary_technical_guarded`[redacted] [redacted], [redacted] [redacted] [redacted] `variant_claim`, secondary [redacted] `diagnostic_label`, secondary [redacted] promotion blocker[redacted] [redacted] [redacted] `diagnostic_only`[redacted] [redacted] [redacted].

[redacted] [redacted] [redacted] [redacted]. `edd_score=99.99`, coverage/MRR/judge[redacted] [redacted] [redacted] [redacted] [redacted] [redacted], [redacted] [redacted] L121 secondary [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] scoreboard[redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted]. rows[redacted] 1[redacted] scoreboard[redacted] 0[redacted], diagnostic-only[redacted] 1[redacted]. [redacted] `diagnostic_secondary_variant`[redacted] [redacted].

[redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]. [redacted] raw [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] [redacted] seed[redacted] [redacted], [redacted] [redacted] smoke[redacted] [redacted] secondary row[redacted] promotion[redacted] [redacted] [redacted] [redacted] [redacted].

2026-07-07 [redacted] L124 - [redacted] [redacted] [redacted] [redacted] [redacted] [redacted]

ADD■ sparse■ ■■■ ■■■ raw ■■ ■■■ ■■■■. ■■■ ■■■■ ■■ ■■ ■■ ■■■ ■■ ■■■. ■■■ “■■■ ■■■■”■ ■■■ “■■ ■■■ ■■ ■■■ ■■ ■ ■■ ■■ ■■■ ■■ ■ ■■■”■ ■■ ■■ ■■■■. ■■■ ■■■ ■■■ ■■■■, ■■■ ■■■■ ■■■ ■■■■ ■ ■ ■■.

■■ ■■■ ■ ■■ ■■■. ■■■ ■■■ 10■ ■■■ ■■■ ■■■ ■■■ ■■■ ■■■ 0■■ ■■. ■■■■■ ■■ ■■■■. sparse ■■■ ■■■ ■ “■■ ■■■■ ■■ ■■”■ ■■■, ■■ ■ ■■ ■■■ ■■ ■ ■■ ■■■ ■■ ■■ ■■■. ■■■ ■■■ “■■ ■■■ ■■■ ■■■”■ ■■■■. ■ ■■ raw ■■ ■■ 97■ ■■ ■■ ■ 5■■ ■■■.

■■ ■■ ■■■ `DOC073` ■■■-■■■■■■ ■■■■■■ ■■■ ■■■■■ ■■■■■■. 31■ ■■ ■■ ■ 15■■ ■■■ 16■■ ■■ ■■. ■ ■■■ `DOC051` ■■■■■■■ ■■■■■■ ■■ ■■■■, ■ ■■■ `DOC025` ■■ ■■ ■■■■■■ ■■■■ ■■■■■■ ■■■■■ ■ ■■■■ ■■ ■■■■. ■ ■■ ■■ sparse ■■ ■■ ■■■ ■■■■ ■■.

■■■■ ■■ ■■■ ■■ ■■■. ■■ seed■ ■■ ■■■ ■■■ ■■■■ ■■■■■ ■ ■■■■. ■■■ ■ ■■ ■ ■■ ■■ 2~3■■ ■■ frozen ■■ ■■■ ■■■, ■ ■■■ raw■ ■■■■ ■■■. ■ ■■■ ■■ ■■■ ■■■■■ ■■■ ■■.

2026-07-07 16:00 L125 - ■■ ■■■ ■■ ■■ ■■

- ■■ ■■: L124■■ raw ■■■ ■■■■ ■■■ `DOC073`, `DOC051`, `DOC025`■ ■■■ ■■■ ■■ ■■■ ■■■ ■■, ■■ ■■ ■■■ ■■ ■■ ■■ ■■■ ■■ ■ ■■.
- ■■: `questions_l125_sparse_field_diagnostic_frozen` 6■■ ■■■ `mixed+padding_trap`■■ ■■■ froze ■■■ ■■■.
- ■■: ■■ ■■■ ■■■ ■■■ ■■·■■ ■■ ■■ ■■■. ■■■ ■■■■ ■■■, ■■■■■ ■■ ■■■ ■■■.
- ■■ ■■: L125 6■■ ■■ no-judge ■■ ■■■ ■■ ■■ ■■ ■■(`padding_trap`)■ ■ ■■■, ■■ ■■■■■ ■■ ■■■ ■■■■ ■■■ ■■■■ ■■■ ■■■ ■■ ■■ ■■■ ■■■■.
- ■■ ■■■: ■ ■■■ ■■ ■■■ ■■■■ ■■■■■■, ■■■■ ■■ ■■ ■■■ ■■ ■■■ ■■■■ ■ ■■■■.
- ■■ ■■ ■■■: ■■ ■■■ CSV■ raw■■ ■■ ■■■■ ■■ ■■■ ■■■, ■■ ■■■■■ ■■ ■■■ ■■ ■■■ ■ ■ ■■ ■■■ ■■ ■■■■.

2026-07-07 17:15 L126/L127 - sparse-field ■■ ■■■ ■■■ ■■ ■■

- ■■: L125■ ■■ ■■■ ■■■ ■■■■, ■■ ■■■ ■■ `padding_trap`■■ ■ ■■■ ■■■■, ■■■ ■■ ■ ■■ ■■■■ ■■ ■■■■■ ■ ■ ■■■.
- ■■ ■■: L126■ `baseline_default,prompt_sweep`, L127■ `topk_sweep,mmr_lambda_sweep`■■ ■■ ■■ 6■■■ ■■■. ■■ ■ ■■■■ `${stamp}`■ ■■■■ ■■ ■■■ ■■■, ■■ `20260707_161721_L126...`, `20260707_161718_L127...`■ ■■■■.
- ■■: `scripts/aggregate_parallel_eval.py`■ ■■ worker contract■ `utf-8-sig`■ ■■, ■■ ■■ ■■■ `diagnostic_only`, `claim_use`, `promotion_blocker`■ ■■ scoreboard ■■■ ■■ ■■. ■ ■■ ■ L126/L127■ ■■ `scoreboard_rows=0`, diagnostic-only■ ■■■■■.
- ■■: L126■■ ■■■ ■■ ■■ ■■■ `prompt_concise_verified`■■. EDD■ 76.64, false-abstain■ 0.0, ■■ ■■■ 28.648■■■. ■■ ■■■ ■■ ■■■■ ■■ ■■ ■■■ ■■■.
- ■■: L127■■■ `lambda05_top8_filter_rewrite_control`■ EDD 71.12■ ■■ ■■■■, false-abstain 0.5■ prompt ■■■ ■■■. ■■ ■■■■■ ■■■ ■■■■ sparse-field ■■■ ■■■■ ■■■.
- ■■ ■■: `scripts/audit_sparse_answer_runs.py`■ ■■■ ■■ EDD■ ■■■ sparse ■■ ■■■ ■■. `prompt_concise_verified`■ padding risk 0, detector conflict 0■■■■ pass rate■ 0.5, present-field coverage■ 0.542■■.
- ■■■■: ■■ ■■■ "■■ ■■■ ■■■■■" ■■■ ■■■. ■■ ■■ ■■■ ■■■ ■■■, ■■ ■■■ ■■■ ■■■ ■■. ■■ relevance judge■ ■■■■ ■■■ ■■■ ■■■■ ■ ■■■■ ■■■■ ■■■ ■■ ■■ ■ ■ ■■, binary abstention■ ■■ ■■■ ■■ ■■■ ■■■ ■ ■■.
- ■■: L125 sparse ■■■ field-level expectation■ ■■ present-field ■■■ not-found ■■■■ ■■ ■■ ■■. ■ ■■■■ `prompt_concise_verified`■ ■■■ ■■ ■■■■ ■ ■■. ■■ broad top_k sweep■ ■ ■■■ ■■ ■■■■ ■■■■ ■■■.