

仓库地图

- **根目录** (`README.md`, `Makefile`, 等): `README.md` 提供项目概述和运行说明, `Makefile` 定义了常用任务 (如 `make smoke` 运行快速测试、`make train_retriever` 等) ¹。
- **configs/**: 存放流水线各阶段的 YAML 配置文件 (如 `smoke.yaml`、`run_baseline.yaml` 等), 所有脚本通过 `--config configs/...` 加载配置 ²。
- **scripts/**: 包含执行各步骤的 Python 脚本 (如 `smoke.py`、`prepare_data.py`、`build_corpus.py`、`eval_retrieval.py`、`run_baseline.py`、`run_multistep_retrieval.py`、`eval_multistep_retrieval.py`、`run_with_calculator.py`、`eval_numeric.py` 等)。这些脚本对应 `README/Makefile` 中列出的命令 (参见项目说明示例命令) ² ¹。
- **src/**: 核心代码库, 进一步分为子模块:
- **src/data/**: FinDER 数据集加载与处理模块 (例如 `finder.py` 从 HuggingFace FinDER 数据集加载 QA 样本并做标准化) ³。
- **src/indexing/**: 文本分块工具 (如 `chunking.py` 提供 `chunk_text()` 以带元信息地拆分证据段落)。
- **src/retrieval/**: 检索器实现 (如 `retriever.py` 定义 `TfidfRetriever`、`BM25kapi`、`HybridRetriever` 等; 其中 `HybridRetriever` 结合了 BM25 与句向量 (SBERT, 使用 `all-MiniLM-L6-v2` 模型) 进行检索) ⁴。
- **src/multistep/**: 多步检索逻辑模块 (包括 `engine.py`、`planner.py`、`refiner.py`、`gap.py`、`stop.py` 等, 用于迭代检索和推理)。
- **src/calculator/**: 数值问答模块 (如 `extract.py` 提取数值信息, `compute.py` 实现同比增长、差值等运算)。
- **src/finder_rag/**: 共享工具和配置加载 (如 `config.py`), 以及评估指标实现 (`metrics.py` 实现了 Exact Match、Recall@K、MRR 等检索/QA 评测函数) ⁵。
- **outputs/**: 实验结果输出目录, 每次运行都会在 `outputs/<run_id>/` 下生成结果文件, 包括 `config.yaml` (运行使用的配置快照)、`metrics.json` (评测结果)、`logs.txt` (日志)、`retrieval_results.jsonl` (最终检索到的证据块)、`predictions.jsonl` (QA 预测)、`predictions_calc.jsonl` (数值 QA 预测)、`facts.jsonl` (提取的数值事实)、`calc_traces.jsonl` (计算过程跟踪) 和 `multistep_traces.jsonl` (多步检索轨迹) 等 ⁶。

事实表

主张/结论	数值/参数	证据文件路径	证据定位说明
FinDER数据集包含5,703个查询-证据-答案三元组	5703	金融AI选题评价.pdf	第1页 ⁷
项目研究周期约为两个月用于模型训练与调优	约2个月	金融AI选题评价.pdf	文档段落 ⁸
Smoke测试默认子集大小 (<code>subset_size</code>) 为20	20	scripts/ smoke.py	行 142– 145 ⁹

主张/结论	数值/参数	证据文件路径	证据定位说明
Smoke测试默认随机种子为42	42	scripts/ smoke.py	行 121– 124 10
Smoke测试默认召回值k (Recall@5) 为5	5	scripts/ smoke.py	行 149– 151 11
Tfidf检索器默认参数： max_features=20000, ngram_range=(1,2)	max_features=20000, ngram_range=(1,2)	scripts/ smoke.py	行 153– 157 12
HybridRetriever结合BM25和SBERT(all-MiniLM-L6-v2)进行检索	SBERT模型: all-MiniLM-L6-v2	CLAUDE.md	段落 说明 4
实现的检索评估指标包括Exact Match、Recall@K、MRR	指标名称如Exact Match、Recall@K、MRR	CLAUDE.md	段落 说明 5
多步检索的 top_k_final 缺省值为10	10	README.md	行 113– 115 13
Smoke测试使用Recall@5和Exact Match作为评估指标	Recall@5, Exact Match	README.md	行33– 36 14

数据来源： 所有信息均来自提供的项目说明文档和 [WCIS24/ohyeah](#) 仓库代码。如上表所示，第一列为研究主张或结论，第二列列出相关的具体数值或参数，后两列给出对应的证据文件及位置。所有证据均可在对应文件中查阅。以上内容构成项目技术档案的基础数据，不涉及任何外部资料或主观推测。

[1](#) Makefile

<https://github.com/WCIS24/ohyeah/blob/3e69a8ca6f5be75741fe4ff96e72a2129839786c/Makefile>

[2](#) [3](#) [4](#) [5](#) [6](#) CLAUDE.md

<https://github.com/WCIS24/ohyeah/blob/3e69a8ca6f5be75741fe4ff96e72a2129839786c/CLAUDE.md>

[7](#) [8](#) 金融AI选题评价.pdf

file:///file_00000000969c722f803f1b9658241c79

[9](#) [10](#) [11](#) [12](#) smoke.py

<https://github.com/WCIS24/ohyeah/blob/3e69a8ca6f5be75741fe4ff96e72a2129839786c/scripts/smoke.py>

[13](#) [14](#) README.md

<https://github.com/WCIS24/ohyeah/blob/3e69a8ca6f5be75741fe4ff96e72a2129839786c/README.md>