

Lloyd's Market News Digest â App Summary

What it is

Local-first pipeline that discovers, extracts, scores, and renders a Lloyd's market news digest from configured sources. Produces HTML outputs (and optional email delivery) for London Lloyd's Market coverage.

Who it's for

Not found in repo. Repo only states it targets the London Lloyd's Market.

What it does

- Ingests sources from sources.csv (RSS/Atom feeds and listing pages).
- Discovers candidates, canonicalizes URLs, and de-duplicates by canonical hash.
- Fetches pages over HTTP with retries/backoff and optional Mongo cache.
- Extracts article text using a multi-method chain (trafilatura, readability, bs4, crawl4ai stub).
- Runs local-first AI stages for relevance, classification, and summarization with caching.
- Renders HTML digests by source type/topic; optional SMTP delivery.
- Logs run metrics and method health (JSONL + stats).

How it works

sources.csv â Discovery (RSS/listing) â Candidates â HTTP Fetcher + Cache â Extractors â AI stages â Digest rendering
Outputs: HTML digest (internal/public), optional email; Storage: Postgres (processed) + Mongo (raw/cache); Observability

How to run

- conda activate 314 (Python 3.14)
- Copy .env.example to .env and set DB/SMTP/Ollama values
- Ensure config.yaml exists in repo root (sample provided)
- Run: python -m lloyds_digest run --now