

---

# News Discovery

AI-Curated Market Intelligence

SCHEDULE	SOURCES	COST / RUN	MONTHLY
Mon/Thu 7:00 UTC	6	~\$2	~\$16

01

---

## Overview

News Discovery is the largest agent in the system — 17 modules, ~3,100 lines of Python. It runs a 4-stage pipeline: discover articles via Perplexity Sonar, Serper, and 19 RSS feeds; filter and deduplicate using Groq Llama 3.3 70B for cheap semantic comparison; scrape full text via Firecrawl; then score with Claude Sonnet using the IPA/FPA taxonomy. A two-LLM strategy (Groq for filtering, Claude for scoring) keeps per-run costs at ~\$2 while processing 200+ articles per cycle.



# How It Works

---

A 4-stage pipeline using a two-LLM strategy to balance cost and quality. Groq handles high-volume filtering at pennies per article; Claude Sonnet scores only the articles that survive deduplication.

## 1 Discover

Perplexity Sonar searches for tracked company news, Serper for Google results, 19 RSS feeds for publisher monitoring. ~200 raw articles per run.

## 2 Filter & Dedup

3-layer deduplication: URL normalization, Airtable check, Groq Llama 3.3 semantic dedup. Publisher quality filter removes low-value sources.

## 3 Scrape & Score

Firecrawl extracts full article text. Claude Sonnet scores each article using IPA (Industry/Portfolio Analysis) and FPA (Firm/Practice Analysis) taxonomy.

## 4 Push

Scored articles written to Airtable News Agents table with source category tagging (bcg\_tl, competitor\_pub, industry/client news).

---

### EXTERNAL APIS

- Perplexity Sonar Pro
- Firecrawl (full-text extraction)
- Groq Llama 3.3 70B (llama-3.3-70b-versatile)
- Serper (Google Search API)
- Claude Sonnet 4 (claude-sonnet-4-6)
- Airtable

# Economics

The two-LLM strategy is the key cost lever. Groq Llama 3.3 70B handles the high-volume filtering stage at ~\$0.10 per run (\$0.0005/article), while Claude Sonnet 4 is reserved for deep scoring of only the articles that survive deduplication. Scoring uses an 18-point rubric across 5 core dimensions with the IPA/FPA taxonomy (Industry/Portfolio Analysis and Firm/Practice Analysis), producing numerical relevance thresholds: 0.75+ rare/exceptional, 0.60-0.74 strong, below 0.45 filtered out.

## Cost Breakdown per Run

SERVICE	PURPOSE	COST
Perplexity Sonar	Company news search queries	~\$0.80
Firecrawl	Full article text scraping	~\$0.60
Claude Sonnet 4	18-point IPA/FPA scoring (post-dedup only)	~\$0.50
Groq Llama 3.3 70B	Semantic dedup + publisher filter (75x cheaper than Claude)	~\$0.10
<b>Total per run</b>		<b>~\$2.00</b>

MONTHLY COST	ARTICLES DISCOVERED	SCORED & PUSHED
~\$16 8 runs / month	200+ Per run	40-60 Per run