Data Sources for Live News Feed Page

Implementation Plan

Phase Overview

Phase	Focus	Sources	Effort	Dependencies
Phase 1	Core News Feed UI	None (UI only)	1 hour	None
Phase 2	No-Key Sources	SEC EDGAR, USPTO, pygooglenews	2 hours	Phase 1
Phase 3	Free API Sources	GNews, MediaStack, NewsData.io	1.5 hours	Phase 1
Phase 4	ML Enhancement	Hugging Face Sentiment	1 hour	Phase 1
Phase 5	AI Subscription Integration	ChatGPT, Gemini, Firecrawl	2 hours	Phases 1-4

Total Estimated Effort: ~7.5 hours

Complete Data Sources Table (3 Existing + 10 New)

EXISTING SOURCES (Currently in Codebase)

#	Source	API Key	Limit	Status
1	Google News RSS	No	Unlimited	Active
2	NewsAPI.org	Yes	100/day	Inactive
3	Bing News API	Yes	1000/mo	Inactive

NEW SOURCES: Category A - Free News APIs

#	Source	API Key	Account	Free Limit	Data Provided
4	GNews API	Yes	Yes (free)	100/day	60K+ sources, 6yr history
5	MediaStack	Yes	Yes (free)	500/mo	7,500+ sources, 50 countries
6	TheNewsAPI	Yes	Yes (no CC)	Free tier	40K+ sources, 1M+/week
7	NewsData.io	Yes	Yes (free)	200/day	89 languages, tech/health

NEW SOURCES: Category B - Government Open Data (100% Free)

#	Source	API Key	Limit	Data Provided
8	SEC EDGAR API	No	Unlimited (10/sec)	8-K filings, M&A, financials
9	USPTO PatentsView	No	Unlimited	Patents, R&D, innovation

NEW SOURCES: Category C - Open Source (GitHub)

#	Source	API Key	Limit	Data Provided
10	pygooglenews	No	Unlimited	Enhanced Google News filters
11	Universal-News-Scraper	No	Unlimited	Multi-source, anti-blocking

NEW SOURCES: Category D - AI Subscriptions (Included in Plans)

#	Source	Subscription	Limit	Data Provided
12	ChatGPT Deep Research	$20/mo Plus	25 queries/mo	AI competitor reports
13	Gemini Deep Research	$20/mo Pro	Included	Google-grounded research
14	Firecrawl MCP	$200/mo Max	500 free credits	Website scraping

NEW SOURCES: Category E - Enhancement Tools

#	Source	API Key	Limit	Data Provided
15	Hugging Face Sentiment	No	Unlimited	ML sentiment analysis
16	Growjo	No	Unlimited	Company growth, hiring data

Phase 1: Core News Feed UI

Goal: Build the News Feed page structure using existing Google News RSS

#	Task	File	Description
1.1	Add sidebar menu item	frontend/index.html	Add "News Feed" link after Change Log
1.2	Create page section	frontend/index.html	Add filter bar + results table HTML
1.3	Add JavaScript functions	frontend/app_v2.js	loadNewsFeed(), populateNewsDropdown()
1.4	Add backend endpoint	backend/main.py	GET /api/news-feed with filters
1.5	Add CSS styles	frontend/styles.css	Sentiment badges, table styling

Phase 2: Government APIs (No API Key Required)

Source 2A: SEC EDGAR API

Endpoint: https://data.sec.gov/submissions/CIK{cik}.json

Rate Limit: 10 requests/second | Data: 8-K filings, M&A;, leadership changes

#	Task	File
2A.1	Create SEC scraper	backend/sec_edgar_scraper.py
2A.2	Add CIK lookup	backend/sec_edgar_scraper.py
2A.3	Integrate with news feed	backend/news_monitor.py

Source 2B: USPTO PatentsView API

Endpoint: https://api.patentsview.org/patents/query

Rate Limit: Unlimited | Data: Patent filings, innovation trends

Source 2C: pygooglenews Library

Install: pip install pygooglenews

Data: Enhanced Google News with geo, topic, date filters

Phase 3: Free News APIs (API Key Required)

Source	Endpoint	Free Limit	Env Variable
GNews API	https://gnews.io/api/v4/search	100/day	GNEWS_API_KEY
MediaStack	http://api.mediastack.com/v1/news	500/mo	MEDIASTACK_API_KEY

NewsData.io	https://newsdata.io/api/1/news	200/day	NEWSDATA_API_KEY

Phase 4: ML Sentiment Enhancement

Goal: Replace keyword-based sentiment with Hugging Face ML model

Install: pip install transformers torch

Model: cardiffnlp/twitter-roberta-base-sentiment-latest

Note: First run downloads ~500MB model (cached thereafter)

#	Task	File
4.1	Add dependencies	backend/requirements.txt
4.2	Create sentiment module	backend/ml_sentiment.py
4.3	Replace keyword analysis	backend/news_monitor.py

Phase 5: AI Subscription Integration

Source	Access	Limit	New File
ChatGPT Deep Research	OpenAI Responses API	25/mo (Plus)	backend/ai_research.py
Gemini Deep Research	Google AI SDK	Included	backend/gemini_research.py
Firecrawl MCP	Firecrawl API	500 free	backend/firecrawl_scraper.py

Environment Variables Summary

New Variables to Add to .env.example:

Variable	Phase	Required
GNEWS_API_KEY	Phase 3	Optional
MEDIASTACK_API_KEY	Phase 3	Optional
NEWSDATA_API_KEY	Phase 3	Optional
GOOGLE_AI_API_KEY	Phase 5	Optional
FIRECRAWL_API_KEY	Phase 5	Optional

New Files to Create

Phase	File	Purpose
2A	backend/sec_edgar_scraper.py	SEC EDGAR API integration
4A	backend/ml_sentiment.py	Hugging Face sentiment model
5A	backend/ai_research.py	ChatGPT Deep Research
5B	backend/gemini_research.py	Gemini Deep Research
5C	backend/firecrawl_scraper.py	Firecrawl web scraping

Dependencies to Add to requirements.txt

Package	Phase	Purpose
pygooglenews	Phase 2	Enhanced Google News
transformers	Phase 4	ML Sentiment
torch	Phase 4	ML Sentiment
google-generativeai	Phase 5	Gemini API
firecrawl-py	Phase 5	Firecrawl API

Success Metrics

Metric	Current	Target
News sources available	3	10+

Articles per search	~10	50+
Sentiment accuracy	~60%	85%+
Data freshness	Real-time	Real-time to 6yr
Geographic coverage	US only	50+ countries

Implementation Order (Recommended)

Week 1: Phase 1 (UI) + Phase 2 (No-Key APIs)

→ Deliverable: Working News Feed with SEC + USPTO + enhanced Google News

Week 2: Phase 3 (Free APIs) + Phase 4 (ML Sentiment)

→ Deliverable: Multi-source news with ML sentiment analysis

Week 3: Phase 5 (AI Subscriptions)

→ Deliverable: AI-generated reports + website scraping