AYLIEN

A Quantexa Company

5 ways to access **news** as a data source at scale

How to choose a News API that fits your requirements

Introduction

Unstructured data, and in particular news data, is a fast growing requirement for organizations, as risk and market intelligence teams seek to expand their data sets and improve their intelligence capabilities.

News is a rich and accessible source of information, containing critical business insights that are used as early warning signals for risks, as well as indicators for opportunities. Global news operates at a huge scale, with the majority of it being freely accessible on the open Web. This is a double edged sword, as finding relevant signals amongst the noise is incredibly difficult without the right tools.

If your organization uses news as a data source at scale, you're probably already familiar with News APIs, and may even already have a solution in place, whether it's a licensed third-party or a solution built in-house. However, technology moves on quickly, and it's always a good idea to keep abreast of what options are currently available.

This whitepaper takes a look at the News API landscape in 2023, with the intention of helping you decide which type best suits your organization's needs, whether that's builing your own, settling for a basic provider, or upgrading to a state-of-the-art solution. We hope you find it useful.

Contents

What are the News API options?

Let's get familiar with the options across the News API landscape, from basic to advanced.

	Build in-house	Basic News APIs	Intermediate News APIs	Legacy vendors	AYLIEN News API
Coverage	Each source must be manually integrated, usually via scraping. Hard to scale and license.	Limited coverage. Expect to miss key insights.	Good range of sources. Comprehensive as with AYLIEN.	Good range of sources, although not quite as comprehensive as some alternatives.	Over **80,000 sources**, automatically translated into English from **15 major languages**.
Data enrichment	None.	Limited to simple metadata (title, author, etc).	Metadata, sentiment, basic categorization. Some limited entity recognition.	Metadata, some categorization. Generally limited.	NLP enrichment across **26 datapoints** encompassing metadata, entity and category tags and sentiment analysis.
Search and discovery	As all enrichment must be built from scratch, no intrinsic search and discovery capability.	Simple keyword search only, returning lots of noise and false positives.	Complex boolean searches, plus limited entity search. Still difficult to accurately categorize and thus find news content.	Advanced boolean searches, but hard to maintain and adapt searches to unlock relevant signals.	Easily create sophisticated AI-enhanced queries that pinpoint relevant news using **5.6 million recognised entities** and **4,500+ topical categories**.

	Build in-house	Basic News APIs	Intermediate News APIs	Legacy vendors	AYLIEN News API
Data utilization/delivery	Integrations have to be built internally.	Integration of the basic news API into your own systems requires your own development teams.	Some integration capabilities via APIs.	Some integration capabilities via APIs.	Easily integrate AYLIEN news feeds and archives into internal systems or consumer-facing apps via News API.
Cost	Open-ended and fallible to scope creep. Takes up valuable developer time.	Low monthly costs, adding sophistication will require internal development costs.	Fixed monthly cost.	High cost, complex pricing, can be difficult to buy from.	Simple fixed annual cost covering product, support, future development.
Reliability	Responsibility to maintain systems and uptime lies with the organization.	Low level of support for news APIs themselves, many internal systems built on top of these will require internal support teams.	System generally reliable. Standard commercial support arrangements.	System generally reliable. Some support provided - but slow response times are common.	Reliable, accurate intelligence, supported by the AYLIEN team.
Web UI	None, unless even more developer resources are assigned to the project.	None.	None.	Yes, but similar results issues with noise due to lack of NLP capabilities. Outdated UIs.	The easiest way to build and test queries, using advanced NLP. Instantly view results and gain insights from visualizations. Read more.

1. Building your own in-house

On the face of it, building your own software should be an ideal solution. Not only is there no need to pay license fees to a software provider, but the organization is able to specify exactly what they need and build precisely that.

Unfortunately, things don't tend to work out that way. And in anything other than the very largest organizations, with large internal development teams, building your own solution tends to be something that is done once and then never again: lesson learned. Before you make the same mistake, consider the following questions:

1. Do you know what you want to build?

There are many complexities of the challenge relating to collecting, enriching, and distributing news content, all in real-time. Different parts of the organization may have a different perspective on what is wanted - who will act as Product Manager and resolve these differences of opinion?

2. Who is building it?

Everything has a cost. In this case, that cost is software development resources, which - in case you had not noticed - can be expensive these days. For context, AYLIEN News API is the result of over **20,000 person-hours of development**.

To match those numbers, either you are hiring new engineers, or taking them off other projects - possibly those a lot closer to your core competence. Whichever it is, you are spending money.

3. Who is managing the relationship with source providers?

It's not just building the software - to actually deliver a meaningful amount of content, you will have to negotiate and manage dozens of relationships with providers over time. Do you have those skills in-house?

4. Who is providing and supporting the infrastructure?

You may have noticed that there is a lot of news out there. Are you ready to build and maintain a scalable infrastructure, and associated database, that provides on-demand, archived access to it?

5. Who is continuing to develop your product, and in which direction?

Experience tells us that an initial attempt at a product is rarely the finished article. Are you ready to take responsibility for continuing to evolve the product, respond to user feedback, and ensure that your teams and customers have access to what they need into the future?

At the heart of most of this argument is a simple truth that most people with any experience of software development will recognize: whilst it might feel easy to build a limited solution to a tightly-scoped problem, building a full product, that is easy-to-use and scales to meet your actual requirements, is a whole lot more complicated.

Just to get up-and-running with the most basic service is likely to run into hundreds of thousands of dollars in development spend, and at least 6 months in terms of time. And probably an awful lot more.

Of course, all this presupposes that the problem IS relatively straightforward. But when it comes to delivering meaningful, accurate, and timely news updates, this is not the case.

Specifically, with so much 'news' out there, making sense of this news requires sophisticated NLP technology to layer on meaning to what is otherwise almost useless data. This is not a trivial undertaking and requires specialist skills that come at a premium.

Is that really something your business can or should commit to building?

> **Specifically, with so much 'news' out there, making sense of this news requires sophisticated NLP technology to layer on meaning to what is otherwise almost useless data.**

2. Basic News APIs

For all the reasons noted above, most organizations choose not to build a solution for finding, analyzing, and distributing news from scratch. Instead, they look to the market for a commercial supplier.

Their first port of call is often a simple News API, perhaps better thought of as a raw news feed. This does what it says on the tin: it aggregates news content from a number of sources and provides an API that enables that content to be integrated into any application that needs it.

On the plus side, this approach is both low-cost and flexible. The content delivered can be used within a range of contexts: essentially wherever you might wish to provide a news feed or searchable archive. However, that is about as far as the positives go. This approach continues to suffer from a number of serious limitations:

- **A significant amount of development is still required within the organization.**
  Essentially, a solution like this provides an aggregated news feed to developers looking to integrate this into applications they are building. To use this in any meaningful way, and to present it to either consumers or employees in a way that really empowers decision-making, remains a significant undertaking. All the negatives around building your own solution still apply.

- **Little or no data enrichment.**
  To make sense of any given news item, and to support meaningful categorization, it is necessary to accurately classify it around subject matter, category, and so on. Basic solutions will be lacking in this area, making it difficult or indeed impossible to find or show content relating to a specific subject, and introducing all the difficulties associated with simple keyword searches.

- **News sources are limited.**
  Although it's fair to say that you will still have access to a lot of 'news' via a simple News API, sources are less often updated, may not be comprehensive, and are less likely to be regularly checked for trustworthiness.

In summary, solutions of this type will require significant additional work to do the job: to enable end-users to *find what matters*, *and only see what matters*.

3. Intermediate News APIs

It is no surprise then, that many organizations look beyond the most basic commercial options when looking to bring real intelligence to employees or end-users. When they do so, they will typically move 'up the food chain' to more fully-featured API solutions.

Of course, this involves greater cost, but it does tend to bring more sources into the equation, and a greater level of enrichment of the data delivered. But many of the core issues remain. These solutions tend to compete on the basis of adding additional sources - from blogs, discussion forums, social media - even reviews. And for some specific jobs, such as assessing sentiment in the market towards a product or brand, this approach makes sense.

However, for any organization looking to power intelligence and decision-making with news content, **it really just means more noise and a signal that is harder to find.**

Crucially, whilst there is better data enrichment than with the simplest solutions, natural language processing (NLP) to classify and categorize content is either limited or non-existent. This in turn means that the process of finding what matters still runs into familiar difficulties. As above, the central challenge: turning huge volumes of news into usable insight, remains unsolved.

Click here to read more about what distinguishes AYLIEN News API from an intermediate News API provider.

> Whilst there is better data enrichment than with the simplest solutions, natural language processing (NLP) to classify and categorize content is either limited or non-existent.

4. Legacy News APIs

Perhaps the most familiar type of News APIs are, understandably, the legacy media monitoring solutions. These are the 'big beasts' of the industry, some of which have been around for decades, slowly (very slowly) adding functionality on top of what could be considered a digital clippings service.

Firstly and most importantly, it is important to understand that even here NLP is limited, and thus what we believe to be the core value of AYLIEN is missing. This approach will provide a good number of sources, but falls down when it comes to making sense of them. Searches rely on complex and unwieldy Boolean queries that take a long time to craft and tune, and require regular updates.

Perhaps more importantly, many of the challenges that exist when working with a large legacy vendor apply. These include:

- High licensing costs, and arcane pricing structures that tend to move in one direction only (upwards)

- Slow development cycles, meaning features taken for granted in contemporary solutions (such as clustering to determine significance, or sentiment analysis) are either non-existent or extremely limited

- Poor customer support and general lack of responsiveness

- Difficulty in influencing roadmap, inflexible or slow to respond to market feedback around business requirements

To conclude, whilst these solutions may offer reasonably comprehensive news and content coverage (although still more limited than AYLIEN), any customer will find themselves dealing with 'big company problems' that will ultimately undermine any attempt to build a working solution with real support from the vendor.

Click here to find out more about what distinguishes AYLIEN News API from a legacy News API provider.

5. AYLIEN News API

At AYLIEN, we firmly believe that our News API is the best on the market. We also get this as feedback from our customers. That's because we provide the winning combination of easy access to the world's news and industry leading natural language processing (NLP). Or to put it another way, we give you all the news, and enable you to find exactly what you're looking for, using cutting-edge technolgy.

Easy access to the world's news coverage:
Having instant and easy access to 80,000+ global news publishers is an incredibly powerful position to be in. This equates to, on average, 1.2 million news articles every day, across 16 languages. AYLIEN gives you the option to access articles in their native language, and also auto-translates every article into English. All you have to do is find what you're looking for. Easy, right?

Industry beating NLP capabilities:
Every news article that AYLIEN ingests goes through our NLP engine, which adds 26 data points, including entity, category, and industry tags, sentiment analysis, and a rich amount of metadata. We recognise over 5.6 million entities, and have a proprietary taxonomy of 4,500+ category and industries. Simply put, AYLIEN creates the most detailed structured news data on the market, that makes creating powerful queries easy, so that you can find the news that matters, when it matters.

> "
>
> With AYLIEN we have a lot more coverage, which was one of our goals. But AYLIEN's categorization helps to reduce the noise. We feel like we can control it to catch more incidents, even those on the edge that we wouldn't catch otherwise.
>
> Thomas Ebenstein, CTO, ESG Plus
>
> "

Fast time to value

Time is money, and investing in a new technology can often take much longer than expected to start delivering value. AYLIEN is comitted to helping our customers get up and running as fast as possible, thanks to our support and customer success teams, and complemented by our meticulous technical documentation and getting started guides. And if you've got any questions, we're just a message away at all times.

The easiest way to build and test queries

Signing up for AYLIEN News API also gives access to Search UI, which enables developers to build, test, and iterate queries within a matter of seconds. All it takes is to input your parameters and click 'Search'. Results can be viewed instantly as a news feed, and you can investigate more thoroughly through our suite of auto-generated visualizations, which utilize AYLIEN's time series, trends, and clusters endpoints. Once you are happy with your query, you can copy it into your production environment, safe in the knowledge that your query is providing quality news results.
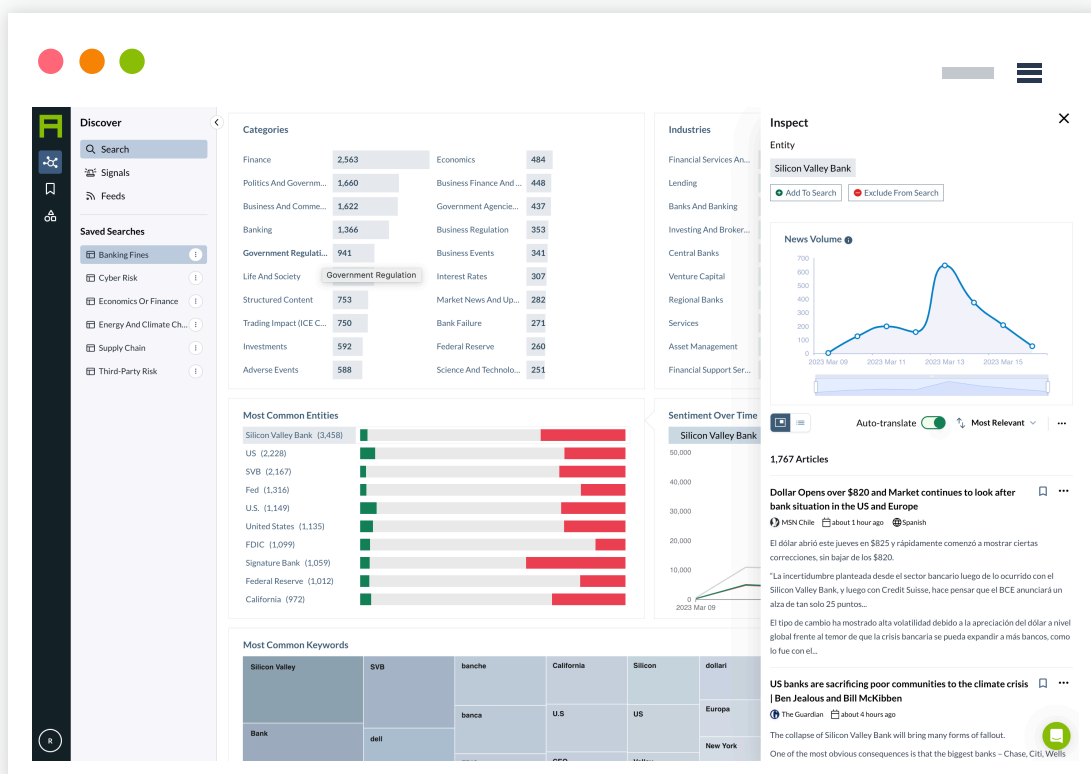
See Live Demo

Start Free Trial

Further reading

- 11 crucial things to look for in a News API

- AYLIEN vs LexisNexis

- AYLIEN vs Webhose/Webz.io

- AYLIEN technical documentation

- AYLIEN News API product page

About AYLIEN

AYLIEN, a Quantexa company, is the leader in applying natural language processing (NLP) to the world's news. We do this because there is an abundance of business signals and insights to be mined from news articles, but due to the overwhelming scale of daily news, these often critical pieces of information are missed. By transforming all the world's unstructured news articles into structured news data, AYLIEN enables organizations like Wells Fargo, IBM, Fidelity, and Deloitte to provide their intelligence teams with the quality data they require to make better decisions.

AYLIEN

A Quantexa Company

Sign up for a 14 day free trial today

Start Free Trial