# Strategic Analysis of RSS Content Ecosystems for Automated CEFR B1/B2 English Learning Pipelines

## 1. Introduction and Architectural Framework

The development of "Ovi English School" represents a convergence of three distinct technological and pedagogical domains: automated content syndication via RSS, Large Language Model (LLM) text adaptation, and automated audio synthesis (TTS). This report provides an exhaustive analysis of the content sourcing layer of this pipeline, specifically targeting the acquisition of high-quality, full-text English material suitable for adaptation to the CEFR B1/B2 proficiency levels for Japanese adult learners. The analysis is predicated on the understanding that the quality of the input data—the raw RSS feed—fundamentally dictates the efficacy of the downstream LLM adaptation. Without semantically rich, contextually complete input, the LLM cannot effectively simplify syntax or generate vocabulary explanations without risking "hallucination" or losing the narrative thread.

### 1.1 The Technical Imperative of Full-Text Syndication

The primary technical bottleneck identified in modern content aggregation is the distinction between "summary" and "full-text" RSS feeds. Historically, RSS (Really Simple Syndication) was designed to syndicate full content.1 However, as digital publishing models shifted towards ad-revenue and subscription maximization, major publishers (e.g., The New York Times, BBC, Wall Street Journal) transitioned to truncated feeds that provide only a headline and a teaser, forcing the user to click through to the website.3

For a human user, this click-through is a minor inconvenience. For an automated pipeline, it introduces significant latency and complexity. Relying on truncated feeds necessitates the implementation of a secondary scraping layer—often requiring headless browsers or DOM parsers to retrieve the full text.4 This not only increases computational overhead but also raises legal exposure regarding terms of service violations. Therefore, this research prioritizes sources that provide "machine-readable full text" natively within the XML feed (<content:encoded> or equivalent tags), or those hosted on platforms with permissive access policies like static engineering blogs or government repositories.6

### 1.2 Pedagogical Alignment: The B1/B2 Constraint

The target demographic—Japanese adult learners at the B1 (Intermediate) and B2 (Upper Intermediate) levels—requires content that balances authenticity with accessibility. B1 learners can understand the main points of clear standard input on familiar matters, while B2 learners can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in their field of specialization.

The source material must, therefore, possess specific linguistic characteristics even *before* LLM adaptation:

1. **Logical Structure:** The text must have a clear expository arc. Disjointed "breaking news" updates often lack the context a learner needs.
2. **Professional Vocabulary:** For the Tech, Medical, and Business verticals, the source must naturally contain the high-frequency domain-specific lexis (e.g., "scalability," "comorbidity," "market capitalization") that the learner intends to acquire.
3. **Cultural Neutrality:** While cultural nuance is important, hyper-local references (common in local news feeds) create cognitive overload for learners. Sources like *Voice of America (VOA)* and *The Conversation* are structurally superior here as they are written for a global or educated lay audience.8

### 1.3 Legal and Ethical Syndication Strategy

The automated nature of the Ovi pipeline necessitates a robust copyright strategy. While "Fair Use" provisions in US copyright law may protect the *transformative* use of copyrighted news for educational adaptation (creating a new work), the direct ingestion and storage of full-text articles from restrictive sources (like Reuters or AP) carries liability.10 Consequently, this report segments sources based on licensing, placing a premium on:

* **Public Domain:** Works of the US Federal Government (VOA, CDC, NIH, NASA).
* **Creative Commons:** Sources explicitly allowing republication (Global Voices, The Conversation, ProPublica).
* **Permissive Corporate:** Engineering blogs and consulting insights (Netflix TechBlog, McKinsey) which encourage dissemination for brand building.

## 2. Vertical 1: General News

The General News vertical serves as the foundational content stream for Ovi English School. The challenge lies in securing reliable, daily updates of significant global events without relying on the restrictive feeds of commercial wire services. The analysis reveals a stark dichotomy between commercial broadcasters (restrictive) and public/educational broadcasters (permissive).

### 2.1 Public Broadcasting and Educational News Sources

#### 2.1.1 Voice of America (VOA) Learning English

**Rating: ★★★★★ (Essential)**

**Status:** Full Text Available (often via specific scrapers or optimized feeds), Public Domain.

Voice of America Learning English stands as the single most critical resource for this pipeline. Unlike standard news outlets, VOA Learning English is explicitly designed for the target demographic (English learners), utilizing a limited vocabulary (approx. 1,500 words) and simplified syntax.

* **Feed Characteristics:** The primary feed URL https://learningenglish.voanews.com/rssfeeds leads to a directory of specialized programs.12 The content is text-heavy and accompanied by audio, providing a ground-truth reference for the TTS pipeline.
* **Pedagogical Fit:** The content is already adapted to a pre-intermediate/intermediate level. This reduces the "hallucination risk" of the LLM, as the model does not need to aggressively simplify complex grammar but can focus on extracting vocabulary lists and generating comprehension questions.
* **Legal Status:** As a funded entity of the US Agency for Global Media, VOA texts are generally in the public domain, allowing for unrestricted commercial use and adaptation.8

**Implementation Strategy:**

Ingest the specific "As It Is" (Daily News) and "Science & Technology" feeds. These provide the highest volume of daily content. The structure of VOA pages is highly consistent, allowing for 100% reliable text extraction if the RSS feed abstract is truncated.

#### 2.1.2 BBC Learning English and World Service

**Rating: ★★☆☆☆ (General News) / ★★★★☆ (Learning English)**

**Status:** Summary Only (News) / Audio-Centric (Learning).

The BBC represents a high-quality but technically challenging source. The standard news feeds (http://feeds.bbci.co.uk/news/rss.xml) are strictly summary-only and subject to strict copyright terms that prohibit commercial use.1

However, the **BBC Learning English** vertical is highly relevant. While they lack a single comprehensive RSS feed for *text*, they offer podcast RSS feeds (e.g., *6 Minute English* at https://podcasts.files.bbci.co.uk/p02pc9zn.rss).14

* **Analysis:** The value here is in the transcripts. The podcast RSS feeds are enclosures for audio. The pipeline would need to fetch the URL from the feed, parse the HTML page for the transcript, and process that.
* **Verdict:** Use BBC News only as a "signal" to identify trending topics, but source the actual text from VOA or Global Voices to avoid legal friction. Use BBC Learning English for specific "Listening" modules, but note that the text availability in the feed is low.

#### 2.1.3 NPR (National Public Radio)

**Rating: ★★★☆☆**

**Status:** Truncated / Topic Specific.

NPR utilizes a granular, ID-based RSS generation system. This allows for precise filtering (e.g., Topic ID 1013 for Education, ID 1019 for Technology).15

* **Feed Quality:** The feeds are truncated summaries. However, NPR articles are written in a conversational, articulate style that is ideal for B2 listening practice.
* **Technical Access:** NPR previously offered a robust API with full-text access for registered developers, though access policies have tightened.16 Without API access, scraping is required.
* **Recommendation:** Use the specific topic feeds to fill gaps in the "General News" vertical, specifically for US-centric cultural stories that VOA might miss.

### 2.2 Independent and Non-Profit Journalism (The "Safe Harbor" Sources)

#### 2.2.1 Global Voices

**Rating: ★★★★★**

**Status:** Full Text (often), Creative Commons (CC BY 3.0).

Global Voices is a community translation project that aggregates citizen media. It is distinct because it operates under a **Creative Commons Attribution 3.0** license, explicitly permitting commercial reuse and adaptation.17

* **Feed URL:** https://globalvoices.org/feed/.18
* **Content Depth:** The articles are deeply contextual, often explaining local political or cultural nuances to a global audience. This expository style is perfect for B2 learners who need to practice reading for context.
* **Implementation:** The feed is high-volume and categorized by region. The "English" feed provides a diverse range of accents and perspectives if used for TTS reference.

#### 2.2.2 The Conversation (US & UK)

**Rating: ★★★★★**

**Status:** Full Text (often), Creative Commons (CC BY-ND).

*The Conversation* publishes articles written by academics and edited by journalists. This results in content that is rigorously researched but accessible—ideal for "Academic English" preparation.

* **Licensing:** They publish under **CC BY-ND** (Attribution-NoDerivatives).9
  + *Constraint Check:* The "NoDerivatives" clause implies you cannot change the text. However, for a learning app, if you present the original text (for reading) and then provide *supplementary* LLM-generated vocabulary lists and quizzes, this likely complies. If the LLM *rewrites* the article to a lower level, this technically violates ND.
  + *Strategic Pivot:* Use *The Conversation* content for "Advanced Challenge" (B2+) segments where the text is presented verbatim, rather than simplified.
* **Feed Quality:** High integrity, structured metadata.

### 2.3 Wire Services (Reuters / AP)

**Rating: ★☆☆☆☆**

**Status:** Strict Copyright, Summary Only.

Reuters and AP form the backbone of global news, but their RSS feeds are designed for syndication clients who pay licensing fees.

* **Reuters:** Feeds (e.g., http://feeds.reuters.com/Reuters/worldNews) are strictly truncated and terms of use explicitly forbid commercial aggregation without a contract.10
* **AP:** Similar restrictions apply. The AP feed format is often custom and not standard RSS 2.0, complicating ingestion.21
* **Verdict:** Avoid direct ingestion. Use them only to validate the factual accuracy of other sources.

### 2.4 General News Source List & Priority

| **Priority** | **Source Name** | **Feed URL** | **Quality Rating** | **License** | **Full Text?** |
| --- | --- | --- | --- | --- | --- |
| **1** | **VOA Learning English** | https://learningenglish.voanews.com/rssfeeds | ★★★★★ | Public Domain | Yes (Scrape) |
| **2** | **Global Voices** | https://globalvoices.org/feed/ | ★★★★★ | CC BY 3.0 | Yes |
| **3** | **The Conversation** | https://theconversation.com/us/articles.atom | ★★★★☆ | CC BY-ND | Yes |
| **4** | **Breaking News English** | https://breakingnewsenglish.com/bne.xml | ★★★★☆ | Educational | Yes |
| **5** | **NPR Education** | https://feeds.npr.org/1013/rss.xml | ★★★☆☆ | Copyright | No |

## 3. Vertical 2: Tech English (Ovi Tech English)

This vertical targets IT professionals. The critical insight here is that **Engineering Blogs** from major tech companies are vastly superior to general tech journalism (like *The Verge* or *TechCrunch*) for educational purposes. Engineering blogs describe *how* systems are built, using the precise technical vocabulary (e.g., "latency," "throughput," "deployment pipeline," "microservices") that Japanese IT professionals need to master for their careers. Furthermore, these blogs often provide full-text feeds to attract engineering talent.

### 3.1 Corporate Engineering Blogs (The "Hidden Goldmine")

#### 3.1.1 Netflix TechBlog

**Rating: ★★★★★**

**Status:** Full Text Available.

The Netflix TechBlog is an industry standard for cloud infrastructure, chaos engineering, and data science.

* **Feed URL:** https://netflixtechblog.com/feed.22
* **Content Analysis:** High complexity (C1/C2), requiring LLM simplification for B2 learners. However, the structure is impeccable—usually following a "Problem -> Solution -> Implementation -> Result" narrative arc, which is easy for an LLM to parse and summarize.
* **Host:** Hosted on Medium. Medium feeds can be truncated, but the Netflix feed often provides substantial content. The URL structure medium.com/feed/@NetflixTechBlog is the canonical alternative.

#### 3.1.2 Uber Engineering

**Rating: ★★★★☆**

**Status:** Full Text Available.

Covers logistics, high-scale data processing, and AI.

* **Feed URL:** https://eng.uber.com/feed/.22
* **Reliability:** High update frequency. The articles are detailed and often include diagrams (which the LLM can describe in text if multimodal capabilities are used).

#### 3.1.3 Stripe Engineering

**Rating: ★★★★★**

**Status:** Full Text.

Stripe is known for "literary" engineering writing. Their posts are exceptionally clear and well-edited, reducing the burden on the adaptation pipeline.

* **Feed URL:** https://stripe.com/blog/feed.rss.25 Note that the main blog feed mixes product news with engineering; an LLM-based filter is needed to select "Engineering" tagged posts.

#### 3.1.4 GitHub Blog (Engineering Category)

**Rating: ★★★★★**

**Status:** Full Text.

Essential for software developers. It covers the tools developers use daily (Git, Actions, Copilot).

* **Feed URL:** https://github.blog/category/engineering/feed/.26
* **Relevance:** High. Content regarding "AI coding assistants" and "CI/CD" is trending and highly relevant for the target audience.

### 3.2 Developer Communities and Aggregators

#### 3.2.1 Hacker News (via hnrss.org)

**Rating: ★★★☆☆**

**Status:** Metadata / Links Only.

Hacker News (HN) itself does not host content; it links to it. The official RSS is sparse.

* **The Workaround:** Use **hnrss.org**. This third-party service allows for granular filtering (e.g., "Only posts with 100+ points").
* **Feed URL:** https://hnrss.org/newest?points=100.28
* **Pipeline Role:** Use this feed as a "discovery engine." When a post hits the threshold, the pipeline captures the *external link*, scrapes that page, and processes it. This ensures the content is currently relevant to the global tech community.

#### 3.2.2 Dev.to

**Rating: ★★★★☆**

**Status:** Full Text.

Dev.to is a community platform where developers write tutorials and opinion pieces.

* **Feed URL:** https://dev.to/feed.29
* **Pedagogical Fit:** Excellent for B1 learners. Many posts are tutorials ("How to build X with React") which use imperative verbs and simple sentence structures, making them easier to digest than abstract architectural discussions.
* **Licensing:** Content is often user-generated; checking individual post licenses is prudent, but generally, the platform encourages sharing.

### 3.3 Mainstream Tech Journalism (Navigating Restrictions)

#### 3.3.1 TechCrunch & The Verge

**Rating: ★★☆☆☆**

**Status:** Truncated / Summary Only.

* **TechCrunch:** https://techcrunch.com/feed/.30 Heavily truncated.
* **The Verge:** https://www.theverge.com/rss/index.xml.31
* **Issues:** These sites rely on ad revenue and are aggressive against scraping. Their content is also often colloquial or gadget-focused ("New iPhone Review"), which has lower long-term educational value than engineering concepts.
* **Recommendation:** Deprioritize these in favor of Engineering Blogs. Use only if "Consumer Tech" is a requested sub-vertical.

### 3.4 Tech News Source List & Priority

| **Priority** | **Source Name** | **Feed URL** | **Quality Rating** | **Note** |
| --- | --- | --- | --- | --- |
| **1** | **Netflix TechBlog** | https://netflixtechblog.com/feed | ★★★★★ | High-level Engineering |
| **2** | **GitHub Engineering** | https://github.blog/category/engineering/feed/ | ★★★★★ | Dev Tools Focus |
| **3** | **Dev.to** | https://dev.to/feed | ★★★★☆ | B1 Friendly / Tutorials |
| **4** | **Uber Engineering** | https://eng.uber.com/feed/ | ★★★★☆ | Backend/Data Science |
| **5** | **Stack Overflow Blog** | https://stackoverflow.blog/feed | ★★★★☆ | Career/Coding Culture |
| **6** | **Hacker News (Filtered)** | https://hnrss.org/newest?points=200 | ★★★☆☆ | Discovery Engine |

## 4. Vertical 3: Medical English (Ovi Medical English)

The Medical English vertical requires a careful bifurcation into two sub-tracks: **Patient-Centric** (General Medical English) and **Clinician-Centric** (Professional Medical English). For B1/B2 learners, Patient-Centric sources are often safer and more linguistically accessible, whereas Clinician-Centric sources (journals) are C2 level and risk inaccuracies during simplification.

### 4.1 Government and Public Health Agencies (The Gold Standard)

#### 4.1.1 Centers for Disease Control and Prevention (CDC)

**Rating: ★★★★★**

**Status:** Full Text / Public Domain.

The CDC offers the most comprehensive library of topic-specific RSS feeds.

* **Feed URL:** https://tools.cdc.gov/api/v2/resources/media/132608.rss (Press Release Feed).32
* **Full Text:** Press releases are often contained fully within the feed or are easily scrapable from the .gov domain.
* **Pedagogy:** The language is authoritative, clear, and defines medical terms explicitly. It is the perfect source for teaching public health vocabulary ("outbreak," "transmission," "vaccination," "prevention").
* **Licensing:** Public Domain (US Government work).

#### 4.1.2 National Institutes of Health (NIH)

**Rating: ★★★★★**

**Status:** Full Text / Public Domain.

NIH feeds cover research breakthroughs and grant news.

* **Feed URL:** https://www.nih.gov/news-events/news-releases/feed.34
* **Content:** Slightly more academic than CDC, focusing on "studies," "trials," and "discoveries." Excellent for distinguishing between "treatment" (clinical) and "research" (academic) vocabulary.

#### 4.1.3 World Health Organization (WHO)

**Rating: ★★★★☆**

**Status:** Full Text (often) / Open Access.

* **Feed URL:** https://www.who.int/feeds/entity/news/en/rss.xml.35
* **Context:** Provides a global perspective, useful for Japanese learners interested in international medical aid or global health policy.

### 4.2 Medical News Portals

#### 4.2.1 ScienceDaily (Health & Medicine)

**Rating: ★★★☆☆**

**Status:** Summary Only (Consistent Structure).

* **Feed URL:** https://www.sciencedaily.com/rss/top/health.xml.36
* **Value:** ScienceDaily publishes structured summaries of research papers. While the feed is not full text, the *format* of their articles (Summary -> Introduction -> Methodology) is so consistent that it is trivial to parse via a scraper.
* **Licensing:** Restrictive. Use for "fair use" adaptation of the *facts* rather than direct republication.37

#### 4.2.2 MedPage Today & Medscape

**Rating: ★★☆☆☆**

**Status:** Truncated / High Complexity.

* **MedPage Today:** https://www.medpagetoday.com/rss/headlines.xml.38
* **Medscape:** https://www.medscape.com/ (Feed URLs vary by specialty).
* **Analysis:** These are targeted at practicing US physicians. The language is extremely dense with jargon (C1/C2). While useful for a "Advanced Medical English" course, they pose significant challenges for B1/B2 adaptation. The content is also often gated (registration wall).
* **Recommendation:** Deprioritize in favor of CDC/NIH for the MVP phase.

### 4.3 Evidence-Based Medicine

#### 4.3.1 The Cochrane Library

**Rating: ★★★★☆**

**Status:** Abstract / Plain Language Summary.

Cochrane Reviews are the gold standard for evidence-based medicine.

* **Feed URL:** https://www.cochranelibrary.com/cdsr/table-of-contents/rss.xml.39
* **The "Plain Language Summary":** Cochrane articles explicitly include a "Plain Language Summary" designed for non-experts. **This is a hidden gem.** This section is practically *already adapted* to B2 English. Extracting this specific section via the RSS link provides high-quality, pre-simplified medical texts.

### 4.4 Medical News Source List & Priority

| **Priority** | **Source Name** | **Feed URL** | **Quality Rating** | **License** | **Target Audience** |
| --- | --- | --- | --- | --- | --- |
| **1** | **CDC Newsroom** | https://tools.cdc.gov/api/v2/resources/media/132608.rss | ★★★★★ | Public Domain | Public / Pro |
| **2** | **NIH Research** | https://www.nih.gov/news-events/news-releases/feed | ★★★★★ | Public Domain | Researcher |
| **3** | **Cochrane Reviews** | https://www.cochranelibrary.com/cdsr/table-of-contents/rss.xml | ★★★★☆ | Restrictive | Evidence-Based Pro |
| **4** | **WHO News** | https://www.who.int/feeds/entity/news/en/rss.xml | ★★★★☆ | Open Access | Global Health |
| **5** | **ScienceDaily** | https://www.sciencedaily.com/rss/top/health.xml | ★★★☆☆ | Restrictive | Student / Layperson |

## 5. Vertical 4: Business English

The Business English vertical presents the highest barrier to entry due to the aggressive paywalls of the "Big Three" (Wall Street Journal, Financial Times, Bloomberg). Automated pipelines cannot easily bypass these without violating laws or terms of service. Therefore, the strategy must pivot from "Financial News" to "Corporate Thought Leadership" and "Startup Management," where content is published for marketing purposes and is thus free and accessible.

### 5.1 Management Consulting (The "HBR" Alternatives)

Harvard Business Review (HBR) feeds are truncated and heavily policed.40 However, top-tier consulting firms publish comparable quality insights for free.

#### 5.1.1 McKinsey & Company (Featured Insights)

**Rating: ★★★★★**

**Status:** Full Text (Reports/Summaries).

* **Feed URL:** https://www.mckinsey.com/insights/rss.42
* **Content:** High-level strategic English. Vocabulary focuses on "transformation," "synergy," "value chains," and "digitalization."
* **Adaptability:** The articles are structured with clear headings and executive summaries, making them ideal for identifying "Business Concepts" for learners.

#### 5.1.2 Bain & Company

**Rating: ★★★★☆**

**Status:** Full Text.

* **Feed URL:** https://www.bain.com/insights/rss.43
* **Focus:** Similar to McKinsey but often more practical/operational. Good for "Case Study" style lessons.

### 5.2 Entrepreneurship and Startups

#### 5.2.1 First Round Review

**Rating: ★★★★★**

**Status:** Full Text.

* **Feed URL:** http://firstround.com/review/feed.xml.44
* **Why it works:** First Round publishes definitive guides on startup management. The content is essentially a "textbook" on modern business management (hiring, product-market fit, culture). It is widely shared and explicitly designed to be educational for founders.

#### 5.2.2 Y Combinator Blog

**Rating: ★★★★☆**

**Status:** Full Text.

* **Feed URL:** https://blog.ycombinator.com/rss/.46
* **Content:** Funding news, startup advice, economic outlooks. The language is direct and concise (Silicon Valley style), which is easier for B1 learners than the flowery prose of the Financial Times.

### 5.3 Financial Markets (The "Paywall" Workaround)

#### 5.3.1 CNBC

**Rating: ★★★☆☆**

**Status:** Summary / Scrape-Friendly.

CNBC is one of the few major financial news outlets that maintains a relatively open RSS ecosystem compared to WSJ/Bloomberg.

* **Feed URL:** https://www.cnbc.com/id/10001147/device/rss/rss.html (Business News).47
* **Usage:** Use this for "Market Updates" (vocabulary: "bull/bear market," "inflation," "interest rates"). While the feed is a summary, the site structure is standard and scrapable.

#### 5.3.2 Project Syndicate (Economics)

**Rating: ★★★★☆**

**Status:** Licensing Required for Republication, but RSS is rich.

* **Feed URL:** https://www.project-syndicate.org/rss.48
* **Content:** Op-eds by Nobel laureates (Stiglitz, Sachs).
* **Level:** C2 (Academic).
* **Use Case:** Excellent for "Advanced Business English." The RSS often provides a substantial excerpt.

### 5.4 Business English Source List & Priority

| **Priority** | **Source Name** | **Feed URL** | **Quality Rating** | **Domain** | **Full Text?** |
| --- | --- | --- | --- | --- | --- |
| **1** | **McKinsey Insights** | https://www.mckinsey.com/insights/rss | ★★★★★ | Strategy | High (Reports) |
| **2** | **First Round Review** | http://firstround.com/review/feed.xml | ★★★★★ | Management | Yes |
| **3** | **The Conversation (Biz)** | https://theconversation.com/us/business/articles.atom | ★★★★★ | Economy | Yes (CC-ND) |
| **4** | **Y Combinator** | https://blog.ycombinator.com/rss/ | ★★★★☆ | Startups | Yes |
| **5** | **CNBC Business** | https://www.cnbc.com/id/10001147/device/rss/rss.html | ★★★☆☆ | Markets | No (Summary) |
| **6** | **Financial Times** | N/A | ★☆☆☆☆ | Finance | No (Paywall) |

## 6. Feed Quality & Technical Implementation

### 6.1 The "Scraping Gap" Analysis

This research confirms that a "pure RSS" pipeline (where XML contains all necessary text) is only viable for about 40% of the desired high-quality sources (primarily Government, Engineering Blogs, and Non-Profits). For the remaining 60% (Commercial News), the RSS feed serves only as a **notification trigger**.

**The Required Architecture:**

1. **Ingestion:** Poll RSS Feed for new <item>.
2. **Filtering:** Check constraints (Is it relevant? Is it too short?).
3. **Expansion:**
   * *If content:encoded exists:* Pass to LLM.
   * *If truncated:* Use the <link> tag. Initiate a headless browser session (e.g., Puppeteer) to fetch the URL. Use a readability parser (like Mozilla's Readability.js) to strip navigation/ads and extract the main article text.
4. **Adaptation:** LLM processes the raw text.

### 6.2 Licensing & Copyright Risk Matrix

The greatest threat to "Ovi English School" is not technical, but legal.

* **High Risk:** Scraping WSJ, NYT, FT, or BBC News. These organizations aggressively defend their IP. **Avoid.**
* **Medium Risk:** Scraping TechCrunch, The Verge, CNBC. Generally tolerated for personal use or snippets, but systematic commercial republication (even adapted) is risky.
* **Low Risk (Green Light):**
  + **VOA/CDC/NIH:** Public Domain.
  + **Global Voices/The Conversation:** Creative Commons.
  + **Engineering Blogs:** Implicit license to share (marketing), generally minimal enforcement.

### 6.3 Update Frequency Strategy

* **Hourly:** CNBC (Markets), Global Voices (News). Use for "Daily Briefings."
* **Daily:** VOA, CDC. Use for "Core Lessons."
* **Weekly:** McKinsey, Netflix TechBlog. Use for "Deep Dive" weekend episodes.

### 6.4 Conclusion & Priority Recommendation

To launch "Ovi English School" with the highest quality and lowest risk, prioritize the **Public Sector** and **Engineering Blog** sources. They provide the "Full Text" gold standard required for LLM adaptation without the legal and technical headaches of bypassing paywalls.

**Recommended Implementation Order:**

1. **Phase 1 (The Foundation):** Integrate **VOA Learning English**, **CDC**, and **Global Voices**. These provide safe, daily, full-text content across General and Medical verticals.
2. **Phase 2 (The Tech Specialist):** Integrate **Netflix**, **Uber**, and **GitHub** engineering blogs. This instantly creates a high-value "Tech English" product that competitors (using generic TechCrunch feeds) cannot match.
3. **Phase 3 (The Business Expansion):** Integrate **McKinsey** and **First Round Review** for high-level business concepts.
4. **Phase 4 (The Gap Filler):** Build the scraping infrastructure for **CNBC** and **NPR** to fill the "breaking news" gaps, ensuring robust error handling for paywalls/truncation.

By following this prioritized list, Ovi English School can establish a robust, automated content pipeline that delivers high-value, level-appropriate English learning material to the Japanese market.

#### Works cited

1. RSS Feeds explained - Fullmeasure, accessed February 1, 2026, <https://fullmeasure.co.uk/rssdoc>
2. A guide to using RSS feeds, the files that contain real-time updates from websites, accessed February 1, 2026, <https://africa.businessinsider.com/news/a-guide-to-using-rss-feeds-the-files-that-contain-real-time-updates-from-websites/1qn9k9z>
3. RSS feed - full text or summary? - Webweaving - 32-Bit Cafe, accessed February 1, 2026, <https://discourse.32bit.cafe/t/rss-feed-full-text-or-summary/436>
4. BBC News - Content only - Full-Text RSS - FiveFilters.org Community, accessed February 1, 2026, <https://forum.fivefilters.org/t/bbc-news-content-only/357>
5. Full-Text RSS - FiveFilters.org, accessed February 1, 2026, <https://www.fivefilters.org/full-text-rss/>
6. quartz/docs/features/RSS Feed.md · main · Luise Matthes / publishing - Git / GitLab, accessed February 1, 2026, <https://gitlab.uni-weimar.de/tego6278/publishing/-/blob/main/quartz/docs/features/RSS%20Feed.md?ref_type=heads>
7. Retrieve Full Content of Posts from RSS Feed - Stack Overflow, accessed February 1, 2026, <https://stackoverflow.com/questions/77672272/retrieve-full-content-of-posts-from-rss-feed>
8. Voice of America - Learn American English with VOA Learning English, accessed February 1, 2026, <https://learningenglish.voanews.com/>
9. Climate Quitting: the People Leaving their Fossil Fuel Jobs Because of Climate Change – Opening Conversations - Open Books, accessed February 1, 2026, <https://openbooks.library.umass.edu/englwrit112reader/chapter/climate-quitting-the-people-leaving-their-fossil-fuel-jobs-because-of-climate-change-ware-and-augustine/>
10. Terms of use | Thomson Reuters, accessed February 1, 2026, <https://www.thomsonreuters.com/en/terms-of-use>
11. Terms and conditions | The Associated Press, accessed February 1, 2026, <https://www.ap.org/terms-and-conditions/>
12. Rss - VOA - Voice of America English News, accessed February 1, 2026, <https://learningenglish.voanews.com/rssfeeds>
13. English Copyright and Intellectual Property Law: Using RSS news feed content in third party app : r/LegalAdviceUK - Reddit, accessed February 1, 2026, <https://www.reddit.com/r/LegalAdviceUK/comments/177ld0y/english_copyright_and_intellectual_property_law/>
14. 6 Minute English - Podnews, accessed February 1, 2026, <https://podnews.net/podcast/i6w2>
15. NPR RSS News Feeds - Reddit, accessed February 1, 2026, <https://www.reddit.com/r/NPR/comments/16t5102/npr_rss_news_feeds/>
16. List of news media APIs - Wikipedia, accessed February 1, 2026, <https://en.wikipedia.org/wiki/List_of_news_media_APIs>
17. Republishing Guidelines - Global Voices, accessed February 1, 2026, <https://globalvoices.org/about/global-voices-attribution-policy/>
18. RSS Feeds - Global Voices, accessed February 1, 2026, <https://globalvoices.org/feeds/>
19. Top 10 The Conversation RSS Feeds, accessed February 1, 2026, <https://rss.feedspot.com/the_conversation_rss_feeds/>
20. 1 of 4 MEDIA SERVICES TERMS AND CONDITIONS Please read these terms carefully before submitting a signed Order Form for a Service - Reuters, accessed February 1, 2026, <https://about.reuters.com/>
21. RSS feeds for Associated Press News - Web Applications Stack Exchange, accessed February 1, 2026, <https://webapps.stackexchange.com/questions/113549/rss-feeds-for-associated-press-news>
22. RSS - GitHub Gist, accessed February 1, 2026, <https://gist.github.com/PierreZ/7aebca8c13e7a9005356>
23. Netflix TechBlog, accessed February 1, 2026, <https://netflixtechblog.com/>
24. Follow Uber Engineering Blog | Feeder – RSS Feed Reader, accessed February 1, 2026, <https://feeder.co/discover/7c06255aed/eng-uber-com>
25. Stripe Blog: Online Payment Solutions Blog, accessed February 1, 2026, <https://stripe.com/blog>
26. Top 45 GitHub RSS Feeds - RSS Database, accessed February 1, 2026, <https://rss.feedspot.com/github_rss_feeds/>
27. New Blog RSS Feed Location - The GitHub Blog, accessed February 1, 2026, <https://github.blog/news-insights/new-blog-rss-feed-location/>
28. Hacker News RSS, accessed February 1, 2026, <https://hnrss.org/>
29. Publishing from RSS or Atom - Editor Guide - DEV Community, accessed February 1, 2026, <https://dev.to/p/publishing_from_rss_guide>
30. Top 45 TechCrunch RSS Feeds, accessed February 1, 2026, <https://rss.feedspot.com/techcrunch_rss_feeds/>
31. Daily RSS let-down: The Verge has an encoding issue. - Reddit, accessed February 1, 2026, <https://www.reddit.com/r/rss/comments/1jhzskg/daily_rss_letdown_the_verge_has_an_encoding_issue/>
32. CDC Online Newsroom, accessed February 1, 2026, <https://tools.cdc.gov/api/v2/resources/media/132608.rss>
33. Subscribe Link - CDC Media Library, accessed February 1, 2026, <https://tools.cdc.gov/medialibrary/app/landing/layouts/feeds.html>
34. News Releases | National Institutes of Health (NIH), accessed February 1, 2026, <https://www.nih.gov/news-events/news-releases>
35. RSS Feeds - WHO Regional Office for Africa, accessed February 1, 2026, <https://www.afro.who.int/rss-feeds>
36. Health & Medicine News - ScienceDaily, accessed February 1, 2026, <https://www.sciencedaily.com/news/health_medicine/>
37. RSS Newsfeeds - ScienceDaily, accessed February 1, 2026, <https://www.sciencedaily.com/newsfeeds.htm>
38. RSS Feeds - Information Center, accessed February 1, 2026, <https://support.medpagetoday.com/hc/en-us/articles/206269823-RSS-Feeds>
39. What's new | Cochrane Library, accessed February 1, 2026, <https://www.cochranelibrary.com/about/releases>
40. Top 25 HBR RSS Feeds (Harvard Business Review), accessed February 1, 2026, <https://rss.feedspot.com/hbr_rss_feeds/>
41. HBR RSS Feeds - hbr.org RSS feeds generator - New Sloth, accessed February 1, 2026, <https://newsloth.com/popular-rss-feeds/hbr-rss-feeds>
42. Explore our insights | McKinsey & Company, accessed February 1, 2026, <https://www.mckinsey.com/featured-insights>
43. Insights - Bain & Company, accessed February 1, 2026, <https://www.bain.com/bain-leadership/publications/>
44. First Round Review, accessed February 1, 2026, <https://review.firstround.com/>
45. Podcast - First Round Review, accessed February 1, 2026, <https://review.firstround.com/podcast/>
46. Also, don't forget to set up an RSS or Atom feed for your website. Contrary to t... | Hacker News, accessed February 1, 2026, <https://news.ycombinator.com/item?id=46469710>
47. Follow CNBC | Feeder – RSS Feed Reader, accessed February 1, 2026, <https://feeder.co/discover/fe348be4a4/cnbc-com>
48. A Practical Guide to RSS - Huey Lee, accessed February 1, 2026, <https://huey.xyz/posts/2022-01-06-practical-rss/>