The CEE Journalism Lecturer's Advanced AI Toolkit (2026 Edition)

Table of Contents

Title	Description
Introduction	An overview of why AI matters for journalism, and why governance and judgment matter more than tools.
How to Use This Toolkit	Guidance for lecturers on how to teach with this toolkit in a modular, problem-based way.
Three Core Pillars of the Curriculum	The foundational principles that guide all AI use in this toolkit: sovereignty, reinvestment of time, and balance.
The AI Audit Rubric	A framework for evaluating AI tools before using them in journalistic work or teaching.
Tool Cluster 1: Transcription & Translation	Using AI to process interviews, recordings, and multilingual material accurately and safely.
Tool Cluster 2: Verification & Investigations	Tools and techniques for verifying content, tracking changes, and uncovering networks.
Tool Cluster 3: Writing & Analysis	Using AI to support drafting, summarising, and working with documents and data.
Tool Cluster 4: Audio, Video & Social	Applying AI to multimedia storytelling and social distribution while maintaining editorial control.
Tool Cluster 5: Security Challenges	Understanding the security risks of AI tools and how to protect data, sources, and journalists.

Tool Cluster 6: Building Your Own Tools ("Vibe Coding")	Using AI-assisted coding to build simple newsroom tools without relying on vendors.
Tool Cluster 7: AI Agents & Automated Workflows	Using AI agents to coordinate tasks, monitor information, and assist editorial workflows.
Addendum A: Caution List: Tools with significant risks	A cautionary list of commonly used tools that pose significant ethical, legal or security risks.
Addendum B: Strategic Reinvestment of Time	How time saved through AI should be reinvested into better journalism rather than more output.
Addendum C: Resources & Research	Key readings, policies, and research for further teaching and study.
Addendum D: Resources & Research Database	
Addendum E: Who to follow online	

Disclaimer
The Toolkit contains links and references to resources provided by third parties. These links are provided for informational purposes only and do not constitute professional advice. While the Toolkit may be used for teaching or training, it should not be relied upon as a sole source of guidance. Although we make reasonable efforts to source reliable content from third parties, we do not guarantee the accuracy of, nor do we endorse, the views, recommendations, or opinions expressed by any third-party content provider linked in the Toolkit. We also have no control over the content of those links or resources.
We accept no liability for any decisions made or actions taken by readers based on this Toolkit. You should not act or refrain from acting based on the information in the Toolkit without first seeking appropriate professional advice regarding your specific facts and circumstances.

Objectives & learning outcomes

Objectives

1: Equip lecturers to teach AI as a set of editorial, ethical and security decisions that shape journalistic practice.

2: Build critical AI literacy grounded in journalistic judgment by enabling students to evaluate, audit, and challenge AI systems, with particular attention to verification, bias and hallucination.

3: Strengthen data independence and source protection in AI workflows by introducing student journalists to open-source and local alternatives that reduce dependency on global platforms and protect sensitive material in high-risk environments.

4: Reframe efficiency gains as a reinvestment opportunity into investigation, verification, audience engagement and human reporting.

5: Develop the capacity for student journalists to build, adapt and critique simple newsroom tools using AI-assisted coding, reducing reliance on vendors, and increasing editorial independence.

Learning outcomes

1: Apply structured frameworks, like the Cost, Difficulty, Invasiveness (CDI) scores[1], to evaluate AI tools for cost, difficulty, invasiveness, security and contextual relevance before use.

2: Select appropriate AI tools and workflows based on the sensitivity of the material, the vulnerability of sources and the potential harm of misuse.

3: Identify hallucinations, translation distortions, analytical weaknesses and framing bias in AI-generated text, audio, visuals and code.

4: Demonstrate editorial accountability when using AI by explaining how AI was used in reporting or production, including verification steps, human oversight, and disclosure decisions.

5: Prototype simple, purpose-built newsroom tools with clear editorial intent using AI-assisted coding to build or adapt small tools while articulating their journalistic value, limitations, and ethical risks.

Introduction

We are at an inflection point in journalism education. By 2025, the challenge is no longer *access* to Artificial Intelligence, but the *governance* of it. Research indicates that over 42% of journalists are already using "Shadow AI" (unapproved tools used without editorial oversight) to write

[1] Created for this toolkit

headlines, transcribe interviews, and generate code.[2]

Your students are likely already using AI tools. The danger is that they are using them blindly. They could be feeding sensitive data into opaque models, risking source confidentiality and relying on "black box" algorithms. This toolkit is designed to help you elevate your teaching from simple "tool training" to critical AI literacy. It does not merely list software. Here our goal is to train journalists who are not just *users* of technology, but *auditors* of it.

Artificial Intelligence is already embedded in journalism, whether it is formally acknowledged or not. Students and early-career journalists are using AI systems to transcribe interviews, summarise documents, translate material, write code, generate visuals, and search for information. The question facing journalism education is therefore not whether AI should be introduced, but how it should be understood, governed and used responsibly.

In the newsroom, AI is currently used in six broad ways.

1) It is used to support reporting and research by transcribing interviews, translating material, and searching large document collections.

2) It is used in verification and investigations to analyze images and videos, track changes to official sources and identify networks of actors.

3) It is used in writing and summarising long texts, cleaning data and assisting with drafting.

4) It is used in audio, video and visual storytelling to improve accessibility and reach audiences in new formats.

5) It is used in security and source protection through encryption, document sanitization and local processing.

6) It is increasingly used to build small, custom tools and this allows journalists to create their own scrapers, calculators, or explainers without relying on external vendors.

How to Use This Toolkit

This toolkit is designed as a teaching companion. It assumes that lecturers are working with students who may be curious, sceptical or already experimenting independently with AI tools.

Rather than organising teaching around tools themselves, the toolkit is structured around journalistic problems. Each tool cluster begins with a real newsroom challenge, for example, handling multilingual interviews, verifying social media content, or analysing messy government data. Tools are introduced only as responses to these problems, not as ends in themselves.

[2] https://digiday.com/media/journalists-are-using-generative-ai-tools-without-company-oversight-study-finds/

For teaching purposes, it is important to demystify AI early. Artificial Intelligence is not a single system and it does not "think [3]." It is a collection of techniques that recognise patterns in data and generate outputs based on probability. In journalism, this means AI is very good at handling scale (large volumes of text, audio, images or data) and very poor at understanding meaning, intent, harm or truth. This imbalance is central to how it should be taught.

Throughout the toolkit, lecturers are encouraged to slow students down rather than speed them up. Every tool is accompanied by a Cost Difficulty Invasiveness (CDI) score [4], which helps students understand trade-offs. A tool that is easy to use may expose sensitive data. A free tool may monetize user content. A powerful tool may quietly train on journalistic work. These considerations are treated as editorial decisions, not technical details.

A central concept used throughout the toolkit is the Time Dividend [5]. AI saves time, but time saved is not neutral. If reclaimed time could be used to simply produce more content faster and then journalistic quality erodes. The toolkit therefore pairs each tool with how time will be saved and then suggestions for how saved time should be reinvested into deeper reporting, verification, source development or audience engagement. This reinforces the idea that AI's value lies in strengthening journalism, not hollowing it out.

Using the term AI can be fraught at the best of times. It is unavoidable to use it throughout this toolkit, but using terms like "an AI" often makes people think of AI as a sentient agent or a conscious being [6]. It also implies there's one discrete object being called AI, which isn't accurate. When talking to your students, try to make sure that all references position AI as a field of practice or as a specific tool, not as a grander intelligence. AI works with patterns, not meaning[7], and it is important to stress that.

Finally, the toolkit is intentionally comparative. Wherever possible, commercial tools are shown alongside open-source or locally run alternatives. This allows students to understand what convenience buys them (and what it costs them) and reduces dependency on a small number of global platforms. In regions where resources are limited or security risks are high, this comparative approach is essential.

The core pillars of this toolkit:

---

[3] AI Is Not Intelligent (research article). *Statistical models excel at pattern recognition but lack essential human intelligence traits*, 27 Jan 2025.
https://www.researchgate.net/publication/388410056_AI_Is_Not_Intelligent

[4] Created for this toolkit
[5] Created for this toolkit
[6] https://www.techpolicy.press/we-need-to-talk-about-how-we-talk-about-ai/

[7] *Artificial Intelligence Works With Patterns, Not Meaning*. Medium (2025). https://medium.com/core-ai/artificial-intelligence-works-with-patterns-not-meaning-and-that-changes-everything-e736da3250c9

- **1) Sovereignty [8].** Whenever possible, use tools that keep data on the journalist's own device. The important note to keep in the back of your head at all times is that using AI can always results in a security risk and that is why when deciding to use AI it isn't just a decision to improve efficiency, but also a decision that takes security into account [9]. AI sovereignty is about allowing the media to control its own data pipelines, train and operate its own AI models that reflect local context, build their own tools and run them on their own infrastructure. They need this to save money, to make money and also to not get swallowed by companies that don't have their best interests at heart. It is a way for countries or communities to operate without global reliance on big tech. For our purposes it is always better to teach the commercial tool *alongside* the technical alternative. This demystifies the "magic" of AI and provides a free, secure backup for students with limited resources or high security needs. Commercial tools change pricing, alter privacy policies, or get banned. By teaching students the "harder," technical route, you give them a skill that is free, secure, and permanent. This is vital for investigative journalists handling sensitive leaks in hostile information environments.

- **2) Time reinvestment.** AI saves time. But in journalism, saving time is dangerous if that time is treated as leisure. This toolkit introduces the concept of the "Time Dividend." *The Rule:* If an AI tool saves you 4 hours of transcribing, you owe the story 4 hours of deeper human work. The key to saving time with AI is also figuring out how to use that time to make the piece of work better. What you should do with this extra time isn't always obvious. And arguably it needs to be on tasks that can in no way be automated, like interviewing one more source. The goal of saving time with AI is to spend more time on *human* connection and investigation. .

- **3) Balance.** To make it easier to compare AI tools in a newsroom or teaching context, this toolkit introduces a simple, original rating system called the CDI score. The CDI score was created specifically for this guide to help lecturers and students quickly assess whether a tool is practical, accessible, and appropriate for their environment.

  Rather than ranking tools as "good" or "bad," the CDI score highlights barriers to entry. It asks three questions every newsroom should consider before adopting an AI tool: Can we afford it? Can we use it? And what does it cost us in terms of privacy and control?

---

[8] *The State of Sovereign AI* August 2025
https://www.linuxfoundation.org/hubfs/Research%20Reports/lfr_sovereign_ai25_082525a.pdf

[9] *AI, Data Governance and Privacy: Synergies and Areas of International Co-operation.* June 20, 2024
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

Each tool in this toolkit is rated on the following three dimensions:

Cost (0–10). 0 means free. 10 means expensive. This measures affordability. Is this realistic for a freelance journalist or small newsroom or is it priced for a well-funded organisation?

Difficulty (0–10)
0 means "click and go." 10 means advanced technical or coding skills are required.
This indicates how much technical knowledge is needed to use the tool effectively.

Invasiveness (0–10)
0 means the tool runs locally or offline and keeps data on your own device.
10 means your data is uploaded, stored, or used to train external models.
This reflects the privacy and security trade-offs involved in using the tool.

An example of how to read the CDI score:
A cloud-based transcription service might have a CDI score of 3–1–6:
Cost: 3 — affordable for most journalists.
Difficulty: 1 — simple, drag-and-drop interface.
Invasiveness: 6 — audio files are uploaded to external servers.
This tells students and lecturers that the tool is easy to use and affordable, but may be unsuitable for sensitive interviews.

By contrast, an open-source transcription tool that runs locally might score 0–8–0:
Cost: 0 — free.
Difficulty: 8 — requires considerable technical setup.
Invasiveness: 0 — data never leaves the device.
This makes it a better option for high-risk reporting, even though it is harder to use.

The CDI score is not a verdict. It is a conversation starter. When teaching classes, it helps students understand trade-offs, compare tools responsibly and make informed decisions based on their resources, skills and security needs.

Ethical Guidelines for the Sustainable Newsroom

As AI systems become embedded in reporting, editing, distribution and audience interaction, the central ethical challenge for journalism is no longer whether to use AI, but how to use it without surrendering responsibility, credibility or source protection.

There is another issue which is linked to ethical AI and that is striving for independence from large tech companies. As you can imagine, this is easier said than done. Where does financial dependence on a certain platform start to infringe on your ethical standing.

These guidelines are designed for a *sovereign newsroom.* In short, this means that one treats AI as infrastructure, not authority. You need to think of automation as assistance, not authorship and efficiency as a means to strengthen accountability, not weaken it.

The principles below establish non-negotiable red lines for editorial control, integrity and data security. In this context, editorial control means that humans and not tools, platforms, or algorithms retain final authority over journalistic decisions. Having a bedrock in the ethical principles is good preparation for when selecting and deciding on AI tools.

1. The "Human Command" Principle [10]

The Rule: *AI is a tool for suggestion, never for decision. No AI output is published without human verification.*

- **The "Human-in-the-Loop" workflow [11]:**
  - **Drafting vs. Publishing:** AI may generate 100 headline ideas (Drafting), but a human editor must select the one that is accurate and fair (Publishing).
  - **The "10-Second" Audit:** For every AI-generated summary or translation, the journalist must trace at least three claims back to the original source text. If the AI cannot provide a citation, the claim is treated as false until proven otherwise.
  - **The Kill Switch:** Human editors must retain the technical ability to immediately retract or modify AI-driven content (automated chatbots) if they begin hallucinating or are hijacked by bad actors.

Tip: Ask students to treat AI like a "Junior Intern" who is incredibly fast but prone to pathological lying. You would never publish the intern's work unread so then do not publish the AI's.

2. No "Synthetic Reality" [12]

The Rule: *Never use AI to simulate the physical reality of news events.*

In conflict zones a photograph is evidence of a war crime. Using AI to generate "realistic" images of protests, soldiers, or destruction—even for "illustration"—corrodes the public's

---

[10] https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

[11] https://ai-act-service-desk.ec.europa.eu/en/ai-act/article-14

[12] https://www.ap.org/about/newsroom/artificial-intelligence/

ability to believe *real* photos.

- **Prohibited Uses (Red Lines):**
  - Generating images of specific people (politicians, activists) doing things they never did.
  - Generating "photos" of events (protests, bombings, meetings).
  - Altering the content of real photos (for example, remove a person from a meeting).

- **Permitted Uses (Green Lines):**
  - **Abstract Concepts:** Visualizing non-physical stories like "cyberwarfare," "inflation," or "mental health" where no specific reality is being claimed.
  - **Clearly Stylized Art:** Using AI to create cartoons or sketches that obviously look like art, not photography.

By flooding the zone with fake realistic images, we give dictators an excuse to dismiss *real* evidence as "just AI." Journalists must not contribute to this.[3]

3. The "Zero Trust" Data Rule [13]

The Rule: *Assume every cloud tool is compromised. Protect your sources from your software.*

Newsrooms can face high-level state surveillance. Their Terms of Service often allow them to review user data for "safety" or training purposes.

- **Green Data (Public Info):** Press releases, published articles, public speeches.
- **Yellow Data (Internal/Low Risk):** Drafts of opinion pieces, planning documents, non-sensitive interviews.
- **Red Data (Sensitive/Source Material):** Whistleblower documents, interviews with soldiers/victims and unreleased corruption evidence.

The AI Audit Rubric

Instructors are advised to encourage students to run every tool through this grid before use.

Criterion	Questions you need to ask	Score 1 – 5
1. Data Sovereignty. This is	1) Where does my	0 - No sovereignty (worst

[13] https://cpj.org/resources/digital-safety/

the idea that determines who controls data, where it is processed, which laws apply to it and who can access it at every stage of its use.	data go when I use this tool? 2) Is my data stored, logged, or reused after the task is completed?	case). Data is uploaded to the cloud, stored, logged, and reused for model training. 5 - Full sovereignty (best case). Local processing only. User controls storage, deletion and access.
2. Exportability. This is the ability to retrieve, move and reuse your data without lock-in.	Can I export my data in open, machine-readable formats (such as a CSV file for spreadsheets, a JSON file for structured data, or just plain text for subtitles)?	0 – No exportability (worst case). Data cannot be exported and outputs are locked inside a proprietary interface. 5 – Full exportability (best case). All data can be exported instantly, no restrictions and the data does not degrade when you export it.
3. Business Model. Specifically, how the company makes money and whether your data is part of the product.	Is the pricing clear and predictable? Does the company sell, share, or monetize user data or metadata?	0 - extractive model (worst case). This can include vague or shifting terms of service and user data used for training, analytics or resale 5 – Aligned and transparent model (best case). Revenue comes from clear, upfront payment (subscription, licence, or one-time fee). User data is not sold, shared, or reused beyond delivering the service.

		Terms of service are stable, readable, and explicit about data use.
4. Security. The tool should protect your accounts, content and access from others.	Does the tool support strong authentication, such as two-factor or multi-factor authentication (2FA/MFA)? This means access requires more than just a password. Perhaps also a code from a phone, an app or a physical security key.	0 – Insecure (worst case). Single-password login only, no clear encryption practices and very weak access controls. 5 – Strong security (best case). End-to-end encryption or strong at-rest encryption and this coupled with clear access control, logging and recovery mechanisms.
Score out of 20		

Glossary of AI in journalism

Artificial Intelligence (AI)

Computer systems are designed to perform tasks that normally require human judgment, such as recognising speech, identifying patterns or generating text. AI does not understand meaning, but instead it predicts likely outputs based on data.

Algorithm

A set of rules or instructions a computer follows to perform a task. In journalism, algorithms often determine how content is processed, ranked or generated.

Large Language Model (LLM)

A type of AI trained on vast amounts of text to generate language. LLMs predict what word is likely to come next, based on patterns, not truth.

Hallucination

When an AI system produces information that sounds confident but is false or invented.

Hallucinations are a known limitation, not a rare bug.

Training Data

The data used to teach an AI system how to behave. Biases and gaps in training data are reflected in outputs.

Open Source

Software whose code is publicly available and can be inspected, modified and run locally. Often safer and more transparent for journalism.

Cloud-Based Tool

A tool that processes data on external servers rather than on the user's device. This can create security and privacy risks.

Local Processing

Running software entirely on a user's own computer, without sending data to the internet. Often preferred for sensitive material.

Metadata

Information about data, such as time, location, device type or author. Metadata can expose sources even when content seems anonymous.

Retrieval Augmented Generation (RAG)

A technique that allows AI to answer questions using a specific set of documents, rather than general internet data. Useful for private archives.

Automatic Speech Recognition (ASR)

Technology that converts spoken audio into written text. Used by journalists to transcribe interviews, press briefings, podcasts and field recordings at speed and scale.

Whisper

An open-source speech-to-text model developed by OpenAI, known for high accuracy across accents, noisy audio and multiple languages. Widely used as the underlying engine in transcription tools for journalism, podcasts and interviews.

Speech-to-Text

A category of AI systems that convert spoken language into written text. Used by journalists to rapidly turn interviews, meetings, and audio recordings into editable transcripts.

Graphics Processing Unit (GPU)

A specialized processor designed to handle large numbers of calculations in parallel. AI tools use GPUs to transcribe audio, generate text, or process images much faster than standard

CPUs.

Command Line Interface (CLI)

A text-based way of interacting with software by typing commands instead of clicking buttons. Common in advanced AI workflows, but usually unnecessary for everyday newsroom use.

End-to-End Encryption (E2EE)

A security method where only the sender and recipient can read the content of messages or files. Critical for protecting sources, sensitive communications and investigative material.

Integrated Development Environment (IDE)

A software workspace that combines code editing, testing, and debugging in one place. Used mainly by developers building or customizing AI tools, not by most journalists directly.

Large Language Model (LLM)

An AI system trained on vast amounts of text to understand, generate and analyze language. Powers tools for writing, summarization, translation, research assistance, and chat interfaces.

Optical Character Recognition (OCR)

Technology that converts scanned documents or images of text into machine-readable text. Essential for working with PDFs, court records, archival material, and photographed documents.

Open-Source Intelligence (OSINT)

Information gathered from publicly available sources such as social media, websites, and databases. Used in investigations, verification, and background research without relying on confidential sources.

Retrieval-Augmented Generation (RAG)

A technique where an AI model retrieves information from specific documents before generating an answer. Improves accuracy and reduces hallucinations when working with trusted newsroom data.

Text-to-Speech (TTS)

Technology that converts written text into spoken audio. Used for accessibility, audio articles, voice previews and rapid multilingual publishing.

Virtual Private Network (VPN)

A tool that encrypts internet traffic and masks a user’s location. Important for journalist

safety, privacy, and working in high-risk or censored environments.

Stable Diffusion

An AI system that generates images from written descriptions, such as illustrations, photos, or visual concepts. It can be used on personal computers or through online services, giving users more control over how images are created and where the data is processed.

Open-Weight Models

AI models that anyone can download and run themselves, rather than only accessing through a company's website or app. This means organizations can use the model on their own computers or servers, with more control over privacy, cost and how the tool behaves.

RSS Feeds

A simple way for websites to automatically send updates when new content is published. Used by journalists to monitor multiple news sites or government pages without visiting each one manually.

APIs (Application Programming Interfaces)

A structured way for one system to share data directly with another system. Used in newsrooms to automatically pull information such as election results, weather data or public records into tools and workflows.

Top 10 AI Use Cases of AI in Journalism

Here are the primary ways the AI is transforming journalism and real-world examples to illustrate.

1. Audience Engagement via Chatbots and Interactive Tools

News organizations are using AI-driven chatbots and interactive assistants to engage readers in conversation, answer questions and provide personalized information. These tools can make the news more accessible and interactive for audiences:

Example: Four small newsrooms launched custom AI chatbots [14] in about a month. These are local outlets using lightweight chatbots specifically to keep readers on-site and answer practical questions instead of bouncing to search and social media.

2. Automated News Writing for Routine Coverage

[14] https://www.niemanlab.org/2025/08/local-newsrooms-are-building-ai-chatbots-fast-and-cheap/

Many newsrooms are leveraging AI to automate the writing of routine news items – things like weather updates, sports recaps, stock reports, or local crime blotters.[15]

Example: In Slovakia, some media outlets used AI to generate content based on real-time results during recent elections, with *AI helping to produce hundreds, sometimes thousands, of articles covering the results in each municipality, freeing editorial staff from manually writing repetitive, data-driven stories.

3. AI-Assisted Writing and Editing

Rather than writing entire articles autonomously, AI is also used as a smart assistant for journalists and editors. These tools can proofread text, suggest better headlines, improve clarity or even propose interview questions and sources.

Example: Gubbi Labs (India) This is a small science news organization in India, built an AI-assisted workflow tool ("Babbler") to speed up writing and editing of science stories.[16] The system can take dense academic research articles and automatically transform them into draft news stories in a more engaging, layperson-friendly style. Journalists then polish these drafts, which has significantly increased the team's content output while maintaining quality.

4. Fact-Checking and Misinformation Detection

Verifying information and debunking false claims is a critical part of journalism. AI systems can assist fact-checkers by scanning content for claims, comparing them against databases, and flagging potential falsehoods or inconsistencies.

Example: Nest Center (Mongolia). Mongolia's *Nest Center for Journalism Innovation and Development* built the country's first AI-driven fact-checking and sentiment analysis system.[17] Integrated into their Pluma news platform, this tool automatically flags suspicious claims and hostile narratives in Mongolian social media content. It scans large volumes of posts for disinformation patterns and even gauges sentiment, alerting the small fact-checking team to potential fake news so they can investigate and debunk them promptly.

5. Investigative Research and Data Analysis

AI is helping journalists find the needle in the haystack when digging through large datasets, public records, or documents. Small investigative teams can use machine learning and NLP (natural language processing) to uncover patterns or leads that would be hard to spot manually.

Example: Dataphyte (Nigeria). Investigative outlet *Dataphyte* created Nubia AI, which combines data analytics with story writing to assist deep dives.[18] Reporters can upload large datasets (e.g. government budgets or health records) and ask Nubia questions; the AI then analyzes every row for trends or anomalies and generates draft narratives. For example, in 2025 a Dataphyte journalist used Nubia to crunch Nigeria's electricity distribution data – the AI

[15] https://www.thomsonfoundation.org/media/269005/tf_ai_in_v4_newsrooms.pdf
[16] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/gubbi-labs
[17] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/nest-center
[18] https://ijnet.org/en/story/debunking-disinformation-turning-datasets-stories-ai-changing-newsrooms-nigeria

spotted patterns and produced analytical text, which the reporter refined in his own style. This saved hours of manual number-crunching and helped uncover hidden insights for investigative stories.

6. Personalized Content Recommendations

To keep readers engaged in the era of information overload, news companies are turning to AI recommendation engines. These systems learn from user behavior and preferences to suggest news stories or deliver tailored newsletters that match individual interests. Personalization can increase reader loyalty and content consumption.

Example: Makedonia (Greece) [19]. Historic regional newspaper *Makedonia* reinvented itself with an AI-powered personalized newsletter system. Their tool analyzes each reader's behavior and preferences on the site and then automatically curates a customized email newsletter and mobile push alerts for that individual. For example, a local reader interested in sports and city politics will get a different mix of stories than someone who clicks mainly on culture pieces. This personalisation initiative has re-engaged lapsed readers and attracted younger audiences, creating a more loyal community by delivering news "they want to know" while still including editor-picked must-reads to avoid filter bubbles.

7. Transcription, Translation and Content Summarization

AI excels at converting content from one form to another, which can greatly speed up newsroom workflows and expand audience reach. This category includes transcribing audio/video to text, translating between languages, and summarizing long texts into briefs.

Example 1: ICIR (Nigeria). The *International Centre for Investigative Reporting* in Nigeria built NativeAI for Newsrooms [20], a speech transcription and translation assistant tailored to local needs. This AI tool automatically transcribes audio or video interviews into text and then translates the text into Nigeria's three major languages (Hausa, Igbo, Yoruba). Unlike generic tools, it was trained to handle Nigerian accents and dialects, solving a huge pain point – reporters no longer spend hours manually transcribing recordings, and stories can be quickly produced in multiple languages. This has broadened the reach of ICIR's reporting to non-English speakers and made their journalism more inclusive.

8. AI-Generated Visuals and Media

Visual journalism is also benefiting from AI. Newsrooms can use generative AI tools to create images, graphics, or even synthetic voice and video, especially when resources are limited or a quick illustration is needed.

Example: Scroll.in (India). Digital news site *Scroll.in* introduced an AI system called Factivo to automatically repurpose text articles into short videos. [21] During India's 2024 general election, they fed election news stories into Factivo, which generated concise video summaries complete

[19] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/makedonia#:~:text=
[20] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/icir-nigeria#:~:text=The%20problem%20they%20wanted%20to,understand%20and%20serve%20local%20languages
[21] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/scroll

with graphics and subtitles. These AI-made explainer videos (clearly labeled as AI content) were pushed to Instagram, resulting in a sharp spike in follower engagement and sharing. The small Scroll team saw this as a way to meet audiences on social media with visual content without hiring a full video crew. Encouraged by the results, they are expanding Factivo to turn articles into other formats (audio streams, interactive charts) to enrich storytelling.

9. Comment Moderation and Community Interaction

Managing the community conversation around news is a challenge that AI can help solve. Small newsrooms often struggle to moderate comments or engage readers in civil discourse due to limited staff. AI tools can detect toxic or spam comments, assist moderators in reviewing posts faster, and even prompt productive discussions by guiding commenters.

Example: Rappler (Philippines). Investigative outlet *Rappler* built an AI chatbot named "Rai" to interact with readers and leverage its archives.[22] Initially trialled as a moderator in private chat forums, Rai graduated in late 2024 into a public-facing Q&A assistant on Rappler's app. Readers can ask Rai questions about the news ("What's the latest on the election fraud case?") and the bot responds with concise answers drawn exclusively from Rappler's 400,000+ published stories. The AI cites its sources and sticks to verified information, building trust while giving the audience a new way to explore reporting. This kind of AI-driven community interaction, essentially a virtual journalist answering queries 24/7, has increased user engagement and time spent on Rappler's platform.

10. Business and Revenue Optimization (Subscriptions, Paywalls & Analytics)

Beyond editorial tasks, AI is also being applied to the business side of news media – helping outlets increase reader revenue and financial sustainability. Machine learning can analyze user behavior and conversion data to optimize paywalls, predict subscriber churn, personalize marketing offers, and even identify new product opportunities.

Example: Sowt (Jordan). Amman-based podcast company *Sowt Media* built an AI-powered analytics platform to drive subscriptions and ads revenue.[23] The system aggregates user data from across their products (podcast listens, website visits, newsletter clicks) into a single dashboard and then uses machine learning to predict user behavior. For example, within a day or two of a new marketing campaign launch, the AI can forecast that campaign's conversion performance – allowing Sowt to adjust messaging or budgeting in near-real-time. It also segments audiences (by interests or geography) and suggests optimal pricing and content bundles for each group. This data-driven approach has made Sowt's marketing far more efficient: early results in 2025 showed higher subscriber sign-ups and reduced customer churn, as the company can proactively target those at risk of cancelling. These AI insights into audience habits and pricing are helping a mid-sized Middle Eastern media outlet compete with larger players on the business front.

[22] https://gijn.org/stories/newsrooms-using-ai-chatbots-leverage-reporting/
[23] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/sowt

Tool cluster 1: Transcription & Translation

Using AI to process interviews, recordings, and multilingual material accurately and safely.

Newsrooms now routinely record far more material than they can manually process. A reporter covering a corruption trial may record several days of court proceedings; a local newsroom may record dozens of community interviews; an investigative team may collect interviews across multiple countries and languages. Transcription tools are used to turn hours of audio into searchable text, so journalists can find quotes, cross-check testimony, and share material with editors. Translation tools allow reporting done in one language to be used by colleagues and audiences in another, as seen in cross-border investigations where interviews conducted in Ukrainian, Romanian, or Arabic are translated into English for collaborative reporting. These tools matter because they determine which reporting is usable, which voices make it into stories, and how accurately spoken evidence is represented.

1. TurboScribe [24]

Cloud-based transcription tool built on Whisper that allows bulk uploads and fast, multilingual transcription.

The core engine of TurboScribe is built upon OpenAI's Whisper technology, widely regarded as the industry standard for open-source speech-to-text accuracy. However, TurboScribe wraps this raw model in a user-friendly, GPU-accelerated infrastructure. This allows for features such as automatic speaker recognition, built-in translation to over 134 languages [25], and the ability to upload 50 files simultaneously for bulk processing.

TurboScribe was created by Leif, an independent developer who focused on making the power of the Whisper model accessible and less technical. The tool is not interested in harvesting your data, however, it does run in the cloud (rather than securely on your machine) so you do need to have trust in their encryption standards and retention policies.

- **The Tool's Purpose:** Handling massive archives for low-budget newsrooms.
- **Relevance to journalism:** . This tool is vital for investigating corruption where journalists are dumped with hundreds of hours of municipal hearing recordings. It supports 98 languages including Romanian and Georgian.
- **CDI Score:** Cost: 3/10 | Difficulty: 1/10 | Invasiveness: 5/10
    - **Comments:** Low flat fee, vital for freelance students. You can simply drag and drop files into the tool for them to be processed. There is cloud processing, so that means

[24] *https://turboscribe.ai*
[25] https://turboscribe.ai/

it is acceptable for public data, but not ideal for confidential sources.

- **Time Dividend:**
  - *How is time saved:* Weeks of listening time.
  - *Ways to reinvest the time:* Use the saved time to cross-reference names across multiple transcripts to find recurring actors in local government corruption.

2. HappyScribe [26]

Transcription and subtitle platform combining AI output with human editing workflows for publication-ready text.

HappyScribe operates as a hybrid transcription and subtitling platform [27] that bridges the gap between AI automation and human precision. Its defining feature is the integrated "Human Review" service and a highly sophisticated interactive editor designed specifically for subtitling workflows. It supports over 120 languages and dialects [28], catering to a global user base that requires nuance often missed by purely algorithmic approaches.

The platform's interactive editor is its standout technical achievement. It allows users to visually sync text with audio waveforms, adjust timing for subtitles, and correct speaker labels dynamically. This makes it superior for post-production video work where timing precision is paramount.

HappyScribe is a European company that has carved a niche by focusing on the workflow of transcription rather than just the raw AI output.

- **The Tool's Purpose:** Video subtitles and translation for social distribution.
- **Relevance to journalism:** Critical for newsrooms producing content for EU audiences. It provides a visual "timeline" for subtitles, which is essential for correcting timing errors in translation.
- **CDI Score:** Cost: 6/10 | Difficulty: 2/10 | Invasiveness: 6/10
- **Time Dividend:**
  - *How is time saved:* 2 hours per minute of video (synchronization is tedious).
  - *Ways to reinvest the time:* Cultural Localization. Don't just translate the content, adapt it. Use the time to ensure the English subtitles convey the *irony* or *sarcasm* common in political discourse, for example.

[26] https://www.happyscribe.com
[27] https://www.happyscribe.com/
[28] https://fritz.ai/happyscribe-ai-review/#:~:text=HappyScribe%20AI%20Review%3A%20Is%20It,But

3. DeepL Pro [29]

High-accuracy machine translation and writing refinement tool designed for professional and editorial use.

DeepL Pro is widely considered the gold standard for AI-driven machine translation [30], consistently outperforming competitors like Google Translate in linguistic nuance and grammatical accuracy [31]. It offers a suite of tools including a web translator, document translation (preserving formatting for.docx,.pptx,.pdf), and an API for integration into other software.

The "DeepL Write" feature acts as an AI copyeditor [32], refining style and tone in addition to translating, which aids in producing professional-grade communications in non-native languages.

Based in Cologne, Germany, DeepL adheres to strict European Union data protection (GDPR) standards.[33] This provenance is significant for users concerned with data sovereignty and US-based surveillance or data mining.

- **The Tool's Purpose:** High-nuance translation for wire reports.
- **Relevance to journalism:** Significantly outperforms Google Translate on Slavic and Romance languages (Romanian, Polish, Russian). Preserves the "formal" vs. "informal" distinction crucial in CEE languages.
- **CDI Score:** Cost: 5/10 | Difficulty: 1/10 | Invasiveness: 4/10
  - **Comments:** Pro version deletes text after processing; Free version uses it for training
- **Time Dividend:**
  - *How time is saved:* Speeds up reading and working with foreign-language material by producing fast, high-quality translations of articles and documents, including preserved formatting in files.
  - *Ways to reinvest the time:* Use the saved time to verify key terms, names, and quotes against the original language (or with a fluent speaker) so tone, legal wording, and diplomatic phrasing are not unintentionally altered.

---

[29] https://www.deepl.com/pro
[30] https://www.linguise.com/blog/guide/deepl-vs-google-translate/#:~:text=DeepL%20vs%20Google%20Translate%20comparison%3A,results%2C%20especially%20for%20European
[31] https://www.smartling.com/blog/google-translate-vs-deepl#:~:text=Although%20Google%20Translate%20is%20the,more%20accurate%20in%20specific%20situations
[32] https://www.deepl.com/en/features/write
[33] https://www.noota.io/en/best-european-ai-guide#:~:text=The%209%20Best%20European%20AI,data%20privacy%2C%20residency%2C%20and

4. Narakeet [34]

Text-to-speech and slide-to-video tool for creating narrated audio and video content from documents.

Narakeet is a specialized tool designed to convert text into spoken audio and to transform PowerPoint slides into narrated videos automatically. It streamlines the creation of voiceovers for educational content, YouTube videos, and internal training materials without the need for recording equipment. Narakeet supports 100 languages and offers 800 voices.

- **The Tool's Purpose:** Specialized TTS for Georgian and Armenian.
- **Relevance to journalism:** One of the few tools with decent support for Georgian and Armenian accents/scripts, which major platforms often neglect.
- **CDI Score:** Cost: 4/10 | Difficulty: 1/10 | Invasiveness: 3/10
- **Time Dividend:**
  - *How time is saved:* Eliminates the need to record, re-record, and edit voiceovers by automatically generating narration directly from text or slides.
  - *Ways to reinvest the time:* Script quality and accuracy—use the saved time to tighten scripts, check pronunciations and names, and ensure the narration matches editorial tone and factual detail before publishing.

5. THE SOVEREIGN ALTERNATIVE: OpenAI Whisper [35]

Open-source automatic speech transcription model that can be run locally for security.

Developed by OpenAI and released as open-source [36] Whisper represents a rare moment where a frontier AI lab released a state-of-the-art model weights-and-all. While OpenAI is a corporate entity, the open sourcing of Whisper empowers the sovereign individual.

.

Trained on 680,000 hours of multilingual, multitask supervised data, Whisper excels at robustness against accents, background noise, and technical jargon. Because it runs via Python or command-line interface (CLI) locally,[37] there are no file size limits or upload caps—only the limitations of the user's hardware.

- **The Tool's Purpose:** The gold standard for secure, high-accuracy transcription.
- **Relevance to journalism:** Whisper is currently the only model that handles code-switching [38] (mixing different languages in one sentence) effectively. It supports

---

[34] https://www.narakeet.com
[35] https://github.com/openai/whisper
[36] https://openai.com/index/whisper/
[37] https://blog.spheron.network/local-installation-guide-for-openai-whisper-step-by-step-instructions
[38] https://huggingface.co/openai/whisper-large-v3/discussions/207#:~:text=Code%20Switching%20using%20Whisper,tuning

Georgian and Armenian with reasonable accuracy, though dialects remain challenging.

- **CDI Score:** Cost: 0/10 | Difficulty: 8/10 | Invasiveness: 0/10
  - **Comments:** The app is free. It requires command Line and terminal knowledge. And there is a zero chance of data leakage, it runs entirely offline.
- Time Dividend:
  - How time is saved: Eliminates the need to record, re-record, and edit voiceovers by automatically generating narration directly from text or slides.
  - *Ways to reinvest the time:* Use the saved time to review the original audio more carefully, note hesitations or emphasis, verify quotes, and focus on editorial judgment rather than mechanical transcription.

Teaching guidance

Emphasise that a transcript is not evidence until it has been checked against the original audio. Students often assume text is authoritative. This cluster is an opportunity to reinforce listening, accuracy, and responsibility.

Draw attention to which voices are transcribed poorly and why. Accents, background noise, and speech patterns are not technical edge cases but core issues of representation and inclusion.

Be explicit when transcription and translation tools are unsafe to use, particularly for sensitive interviews. This cluster is an early place to introduce source protection without fear-based teaching.

Exercise 1: Meaning loss in translation

Students translate a short interview excerpt into another language using AI, then back into the original language.[39] They must identify shifts in tone, meaning, or emphasis and explain how those shifts could affect a published story.

Exercise 2: Accent and exclusion test

Students are given short interview recordings featuring different accents, speech patterns, or levels of audio quality. They run all recordings through the same transcription tool and compare accuracy rates. Students then write a short reflection on which voices are transcribed least accurately and why this matters for fairness, representation, and whose voices are likely to be excluded from published journalism.

[39] https://www.youtube.com/watch?v=S1FhGaxXt0w

Where should you start? (cost, difficulty and what it teaches)

Start with: TurboScribe + DeepL Pro

Why: Extremely low learning curve (drag-and-drop, copy-paste). Immediate, visible payoff for students. Affordable or free tiers suitable for universities and students

What it teaches well: The value of transcription as a *working document*, not a finished artefact. How translation choices affect meaning, tone, and power. The limits of "AI accuracy" and the need for human verification

Best teaching use: Introductory AI literacy, interview processing and cross-border collaboration exercises.

Introduce next: HappyScribe (selectively)

Why: Slightly higher cost, but unmatched for subtitle timing and video workflows. Visual interface makes it easy to teach to non-technical students.

What it teaches well: The difference between transcription *for reporting* and transcription *for publication.* How translation and timing choices shape audience perception.

Tool cluster 2: Verification & Investigations

Tools and techniques for verifying content, tracking changes, and uncovering networks.

Verification tools are now essential in everyday reporting, not just investigations. When a video circulates on social media claiming to show a war crime, protest, or police abuse, newsrooms use video verification tools to check where and when it was filmed and whether it has appeared before. When governments quietly update official websites after public backlash, journalists use change-tracking tools to document what was altered and when. In longer investigations, network-analysis tools are used to map company ownership, political connections, or financial flows, such as tracing how a local company links back to offshore entities. These tools matter because they allow journalists to prove claims, document accountability, and resist denial in environments where misinformation is deliberate and systematic.

6. NotebookLM [40]

[40] https://notebooklm.google

AI-powered research and synthesis tool grounded in user-provided source documents.

NotebookLM is a research assistant designed to work *only* from materials you upload, such as PDFs, transcripts, reports, notes, or web articles. Unlike general chatbots, it does not answer from the open internet by default; it reasons strictly within the provided sources and cites them directly. This makes it particularly useful for sense-making, summarisation, and comparison across large document collections.

NotebookLM excels at helping users interrogate complex material: asking questions across documents, generating summaries, extracting themes, and explaining contradictions while pointing back to the original text. It is not a writing or publishing tool; it is a thinking and reading companion. As part of Google's ecosystem, it prioritises ease of use and rapid onboarding, but uploaded materials are processed on Google servers.

The Tool's Purpose: Source-grounded research, synthesis, and sense-making.

Relevance to journalism: Ideal for journalists working through long reports, court judgments, policy documents, or interview transcripts who need to understand *what the material says* before deciding what to report.

- **CDI Score:** Cost: 0/10 | Difficulty: 1/10 | Invasiveness: 6/10

Time Dividend:

How time is saved: Reduces hours of skimming and re-reading by quickly surfacing summaries, key passages, and relevant sections across multiple documents.

Ways to reinvest the time: Interpretation and framing—use the saved time to read critical passages closely, compare competing claims, and decide what is genuinely newsworthy rather than relying on surface-level summaries.

7. Google Pinpoint [41]

Document analysis tool for journalists that makes large collections of PDFs, images, and audio searchable.

Google Pinpoint is a research tool specifically engineered for investigative journalists and academics dealing with massive document dumps. It uses Google's OCR (Optical Character Recognition) [42] and speech-to-text technologies to make thousands of PDFs, images, handwritten notes, and audio files searchable.

Pinpoint's strength lies in its ability to handle "unstructured" data. It can transcribe audio files

---

[41] https://journaliststudio.google.com/pinpoint

[42] https://journaliststudio.google.com/pinpoint/about/#:~:text=Search%20through%20thousands%20of%20documents

and automatically index them alongside PDF text. A key feature is entity extraction [43]. It automatically identifies and highlights people, organizations, and locations across the entire dataset, revealing connections that a human reader might miss.

As part of the "Google News Initiative [44]" Pinpoint is free for journalists, reflecting Google's strategy to integrate itself into the news gathering ecosystem [45]. While powerful, this provenance is a double-edged sword. Using Pinpoint means uploading potentially sensitive investigative material to Google's servers. While Google promises privacy "by default," the threat model of an investigation targeting Google partners or government entities might preclude its use.

- **The Tool's Purpose:** Investigative document analysis (OCR & Entity Extraction).
- **Relevance to journalism:** Can process thousands of PDFs (handwritten archives, leaked contracts) in mixed languages.
- **CDI Score:** Cost: 0/10 | Difficulty: 3/10 | Invasiveness: 6/10
  - It is free for journalists, but your metadata is still visible to Google.
- **Time Dividend:**
  - *How time is saved:* Instantly makes large volumes of documents, images, and audio searchable, eliminating the need for manual reading just to find relevant names, dates, or terms.
  - *Ways to reinvest the time:* Use the saved time to select the right documents to be processed.

8. InVID / WeVerify [46]

Browser-based video and image verification toolkit for social media and conflict reporting.

The InVID/WeVerify plugin is a browser extension designed to help journalists [47] and fact-checkers verify videos and images from social media platforms like Facebook, YouTube and X. The tool offers a suite of forensic modules. It can fragment a video into keyframes, allowing users to perform reverse image searches on specific moments to find the original source. It includes metadata readers, optical character recognition (OCR) for text within images, and

---

[43] https://gijn.org/stories/making-video-audio-files-searchable/#:~:text=Drive%29
[44] https://newsinitiative.withgoogle.com/en-gb/resources/trainings/video-getting-started-with-pinpoint/#:~:text=Initiative%20newsinitiative,its%20new%20functionality%20%27Structured
[45] https://knightcenter.utexas.edu/google-artificial-intelligence-tools-notebooklm-pinpoint-offer-creative-techniques-to-increase-efficiency-in-the-newsroom/#:~:text=,complex%20ideas%20and%20generate%20connections
[46] https://www.invid-project.eu
[47] https://www.invid-project.eu/tools-and-services/invid-verification-plugin/#:~:text=This%20toolkit%20is%20provided%20by,when%20verifying%20videos%20and%20images

forensic filters (like magnifying lens and error level analysis) to detect image manipulation. Recent updates added [48] a "CheckGIF" tool to compare manipulated images against originals and a "Database of Known Fakes".

It is comprehensive, free, and trusted by major newsrooms, consolidating multiple forensic tools into one interface. However, it relies on third-party APIs [49] (like Twitter/YouTube) which can break or change access rules, disrupting workflows.

- **The tool's purpose:** Video forensics and keyframe extraction.[50]
- **Relevance to journalism:** This is the standard EU-funded tool for verifying footage from conflict zones. It is essential for checking "reverse image search" across Yandex, Baidu, and Google simultaneously.
- **CDI Score:** Cost: 0/10 | Difficulty: 5/10 | Invasiveness: 1/10
- **Time Dividend:**
  - *How time is saved:* Automates multiple technical verification steps—keyframe extraction, reverse image searches, metadata checks, and basic forensic analysis—within a single interface.
  - *Ways to reinvest the time:* Use the saved time to compare results across sources, assess consistency or contradictions, and document clearly why a piece of visual evidence is credible or not.

9. Visualping [51]

Website monitoring service that alerts users when visual or textual changes occur on web pages.

Visualping is a simple yet potent website change detection tool [52]. It monitors web pages for visual or text changes and sends alerts via email, Slack, or SMS. It is widely used for competitive intelligence, regulatory monitoring, and tracking disparate digital footprints [53] that do not offer RSS feeds or APIs.

- **The Tool's Purpose:** Monitoring changes on government websites.
- **Relevance to journalism:** Governments in the region often stealthily delete reports or

[48] https://weverify.eu/verification-plugin/#:~:text=Assistant%20guides%20users%20between%20the,Checking%20Innovative%20Initiative
[49] https://liveproxies.io/blog/x-twitter-scraping
[50] https://www.invid-project.eu/tools-and-services/invid-verification-plugin/

[51] https://visualping.io

[52] https://www.techradar.com/pro/visualping-web-content-monitoring-review
[53] https://visualping.io/blog/regulatory-compliance-monitoring

change politician biographies. Visualping alerts you to pixel-level changes.

- **CDI Score:** Cost: 4/10 | Difficulty: 2/10 | Invasiveness: 2/10
- **Time Dividend:**
  - *How time is saved:* Automates continuous monitoring of multiple web pages, removing the need for repeated manual checks to see whether content has changed.
  - *Ways to reinvest the time:* Use the saved time to confirm what changed, capture evidence (screenshots, timestamps), and maintain a clear change log before deciding whether further reporting or outreach is warranted.

10. THE SOVEREIGN ALTERNATIVE:
Bellingcat Online Investigation Toolkit [54]

Curated collection of open-source tools and methods for digital investigations.

The Bellingcat Toolkit is not a single piece of software but a curated, dynamic repository of open-source tools maintained by the world's premier citizen investigation collective. [55] It serves as the definitive map for the OSINT landscape, organizing tools by function: geolocation, data verification, archiving, and transport analysis.

The toolkit includes guides and links [56] to satellite imagery services, shadow analysis tools and social media scrapers. Bellingcat has revolutionized journalism by proving that open-source data can solve war crimes and expose state secrets. Their toolkit is the crystallization of this methodology. Using their toolkit aligns the researcher with a methodology of transparency and verifiability. It is a community-driven resource, democratizing access to intelligence capabilities previously reserved for state agencies.

- **The Tool's Purpose:** Methodology and geolocation tools.
- **Relevance to journalism:** Not a single "AI," but a suite of AI-assisted tools (like SunCalc and satellite imagery analysis) vital for holding power to account in the region.
- **CDI Score:** Cost: 0/10 | Difficulty: 9/10 | Invasiveness: 0/10
- **Time Dividend:**
  - *How is time saved:* Prevents "reinventing the wheel" for investigations.
  - *Ways to reinvest the time:* Determin the *exact time* an event happened to disprove

---

[54] *https://bellingcat.gitbook.io/toolkit*

[55] https://medium.com/@VerifyHQ/top-10-osint-tools-for-image-and-video-verification-in-2025-139f06458172

[56] https://www.bellingcat.com/resources/2024/09/24/bellingcat-online-investigations-toolkit/#:~:text=Toolkit%20www,and%20maps%2C%20social%20media%2C

official narratives.

11. THE SOVEREIGN ALTERNATIVE: Aleph (OCCRP) [57]

Investigative database for searching public records, leaks, and corporate ownership data worldwide.

Aleph is a massive, open-access [58] data platform created by the Organized Crime and Corruption Reporting Project (OCCRP). It acts as a "global archive of research material for investigative reporting," indexing public records, leaks, and databases to help trace assets and corporate ownership worldwide. The OCCRP is a non-profit media organization dedicated to exposing crime. Unlike commercial compliance databases, Aleph is free for journalists and activists [59], democratizing access to financial intelligence.

- **The Tool's Purpose:** Follow-the-money database.
- **Relevance to journalism:** The "Google" for corruption. Connects entities across borders.
- **CDI Score:** Cost: 0/10 | Difficulty: 4/10 | Invasiveness: 1/10
- **Time Dividend:**
  - *How is time saved:* Aleph answers basic investigative questions, "Has this person or company appeared in other leaks or records?" in seconds, replacing days of manual cross-checking across registries and prior investigations.
  - *Ways to reinvest the time:* Use the time saved for primary verification: retrieving certified documents, conducting interviews, and validating findings for legal or public scrutiny. Aleph surfaces leads, but credibility still comes from traditional reporting.

12. Neo4j [60]

Graph database used to model and analyse complex networks such as financial flows or social relationships.

---

[57] https://aleph.occrp.org

[58] https://www.occrp.org/en/announcement/occrp-announces-a-new-chapter-for-its-investigative-data-platform-aleph-pro#:~:text=Created%20by%20OCCRP%20nearly%2010,collaborate%20on%20vast%20data%20investigations

[59] https://www.occrp.org/en/announcement/aleph-pro-frequently-asked-questions-on-the-future-of-occrps-investigative-data-platform#:~:text=OCCRP%20www,journalists%2C%20researchers%2C%20and%20the%20public

[60] https://neo4j.com

Neo4j is the world's leading graph database management system [61]. Unlike traditional relational databases (SQL) that store data in tables, Neo4j stores data as "nodes" and "relationships." This makes it the ultimate tool for mapping complex networks, such as money laundering schemes, social media interactions, or cyber infrastructure.

The Tool's Purpose: Network mapping and relationship analysis.

Relevance to journalism: Essential for investigations where connections matter more than individual facts (company ownership, political patronage, shell structures).

- **CDI Score:** Cost: 3/10 | Difficulty: 8/10 | Invasiveness: 0/10
  - **Comments:** The community edition free, but hosting costs apply. And it can run fully locally.

Time Dividend:

- *How is time saved:* Replaces weeks of manual relationship mapping.
- *Ways to reinvest the time:* Hypothesis testing. Use the graph to ask "who connects these two entities?" rather than reading documents sequentially.

13. Changedetection.io [62]

Self-hosted, open-source website change monitoring tool offering privacy and unlimited checks.

Changedetection.io lets you keep an eye on websites and get alerts [63] when something changes. For example, when a government page is quietly updated, a report is removed, or wording is altered after public criticism. Unlike commercial services, this tool runs under your own control instead of on someone else's platform. That means no company can see which websites you are monitoring, and there are no limits on how many pages you track or how often you check them. Because it does not rely on a third-party service, there are no monthly fees and no risk that your monitoring activity is logged or shared. It is especially useful for journalists who need to watch sensitive or politically important websites without drawing attention. It gives you a way to monitor website changes quietly, privately, and for free, without handing that information to another company.

The Tool's Purpose: Private, independent monitoring of website changes.

Relevance to journalism: Enables journalists to monitor sensitive government or corporate

[61] https://github.com/neo4j/neo4j
[62] https://changedetection.io

[63] https://changedetection.io/as-featured-in#:~:text=use

pages without revealing interest to third-party services.

- **CDI Score:** Cost: 0/10 | Difficulty: 4/10 | Invasiveness: 0/10

Time Dividend:

- *How is time saved:* Eliminates repetitive manual checking.
- *Ways to reinvest the time:* Verification writing. Document what changed, when, and why it matters.

14. Dangerzone [64]

Local document sanitisation tool that converts untrusted files into safe, non-executable PDFs.

Dangerzone is a specialized security tool that sanitizes potentially malicious documents. It takes untrusted PDFs, office documents, or images and converts them into safe, inert PDFs. Dangerzone works by creating a sandbox (a secure, isolated environment) where you can open the suspicious file and convert it [65] into raw pixel data (images), and then reconstruct a new PDF from those images. Developed by Micah Lee at the Freedom of the Press Foundation, Dangerzone is explicitly designed for journalists handling leaks from anonymous sources.

The Tool's Purpose: Safe handling of leaks and untrusted documents.

Relevance to journalism: Prevents malware hidden in PDFs, Word files, or images from compromising a journalist's device—one of the most common and least visible attack vectors against newsrooms.

- **CDI Score:** Cost: 0/10 | Difficulty: 3/10 | Invasiveness: 0/10

Time Dividend:

- *How is time saved:* Avoids catastrophic compromise rather than speeding up workflow.
- *Ways to reinvest the time:* Secure verification. Always sanitise documents before reading, sharing, or analysing their contents.

15. OpenStreetMap [66]

[64] https://dangerzone.rocks
[65] https://gijn.org/resource/introduction-investigative-journalism-digital-security/
[66] https://www.openstreetmap.org

Open geographic database and query interface for extracting detailed mapping data.

OpenStreetMap is a website where volunteers around the world add and update information about roads, buildings, paths, villages, and landmarks. Think of it as a shared public map that anyone can improve and anyone can use.

While Google Maps is mainly about directions, traffic, and businesses. OpenStreetMap is about recording what actually exists in a place. Because it is community-driven, it often shows things Google Maps misses, especially in rural areas, poorer regions, or conflict zones. OpenStreetMap often records things like footpaths, dirt roads, and informal routes that people actually use, but that are not "official". Small or non-commercial buildings such as clinics, water points, refugee camps, schools, mosques, or community halls that have no business listings.

For journalists, OpenStreetMap is useful when reporting on places that are under-mapped or ignored, because it can show whether basic infrastructure exists even when there are no reviews, photos, or commercial listings.

The Tool's Purpose: Open geospatial investigation and mapping.

Relevance to journalism: Often more accurate than Google Maps in conflict zones and rural areas due to humanitarian mapping contributions.

- **CDI Score:** Cost: 0/10 | Difficulty: 6/10 | Invasiveness: 0/10

Time Dividend:

- *How is time saved:* Replaces field mapping and incomplete commercial datasets.
- *Ways to reinvest the time:* Geolocation verification. Cross-check satellite imagery, eyewitness claims, and timelines.

Teaching guidance

Reinforce that verification tools surface leads, not truth. Students should always be required to explain how a tool's output supports or weakens a claim.

Encourage methodological transparency. Ask students to document each verification step so the process can be assessed, not just the conclusion.

Use this cluster to teach patience. Verification takes time, and resisting speed is part of professional discipline in investigative journalism.

Exercise 1: Prove or disprove

Students are given a viral image or video with a strong claim attached. Their task is to verify it

using appropriate tools and produce a short verification memo stating what can be confirmed, what cannot, and what evidence supports each conclusion. Marks are awarded for method, not outcome.

Exercise 2: Mapping power

Students are given a list of companies or individuals involved in a public issue (e.g. housing, mining, elections). Using public data and network-mapping tools, they create a basic relationship map and submit a short explanation of what the map reveals that would not be obvious from reading documents alone.

Where should you start? (cost, difficulty and what it teaches)

Start with: InVID / WeVerify + Visualping

Why these first

- These tools allow lecturers to focus on **journalistic reasoning** (what to check, why it matters) rather than technical setup.
- Browser-based or very simple interfaces
- Immediately applicable to everyday reporting
- Teach verification *as a habit*, not a specialist skill

What they teach well

- How to verify social media content under time pressure
- The idea that claims must be proven, not repeated
- How small digital traces (a frame, a timestamp, a page change) become evidence

Introduce next: Google Pinpoint + Aleph (OCCRP)

Why these next

- Both are free for journalists
- Designed explicitly for investigative work
- Handle large, messy datasets that overwhelm manual methods

What they teach well

- Once students understand *why* verification matters, these tools teach scale and structure.
- How investigations move from *documents* to *patterns*
- Entity recognition, cross-referencing, and follow-the-money thinking
- The difference between a document archive and an investigation

Tool cluster 3: Writing & Analysis

Using AI to support drafting, summarising, and working with documents and data.

Journalists increasingly work with material that is too large for any one person to read closely. When a government releases a 400-page policy document late at night, reporters use AI to generate an initial summary so they can identify key sections to investigate further. In court reporting, journalists use AI to scan judgments for precedents or contradictions. In data journalism, AI-assisted tools help clean spreadsheets full of errors or inconsistencies before analysis begins. These tools matter because they help journalists cope with scale without abandoning scrutiny, allowing them to focus their limited time on verification, interviews, and editorial judgment rather than mechanical processing.

16. Claude [67]

Large language models optimized for long documents, reasoning and analytical writing tasks.

Claude Opus 4.5 (released November 2025 [68]) is positioned as the world's premier model for complex reasoning, coding, and agentic tasks. It is the flagship model from Anthropic, designed to outperform competitors in deep research and extended cognitive tasks.

Anthropic was founded by former OpenAI safety researchers. Their "Constitutional AI" approach prioritizes safety and steerability [69]. For investigators, this means Claude is often less prone to sycophancy [70] (agreeing with the user falsely) and better at nuanced, neutral analysis than ChatGPT.

- **The Tool's Purpose:** The "Editor" and "Coder."
- **Relevance to journalism:** Superior nuance and less "robotic" tone than ChatGPT. The 200k context window allows pasting entire EU regulation PDFs for summary.
- **CDI Score:** Cost: 5/10 | Difficulty: 4/10 | Invasiveness: 3/10
- **Time Dividend:**
  - *How is time saved:* Drafting headlines, summarizing reports, writing Python code.
  - *Ways to reinvest the time:* Fact-Checking. Always assume the AI is hallucinating.

---

[67] https://www.anthropic.com/claude
[68] https://www.anthropic.com/news/claude-opus-4-5
[69] https://www.datastudios.org/post/claude-4-5-vs-deepseek-s-in-november-2025-full-report-and-comparison-on-features-performance-pri
[70] https://www.anthropic.com/research/towards-understanding-sycophancy-in-language-models

Check every name and date.

17. ChatGPT Team Plan [71]

Collaborative version of ChatGPT offering shared workspaces and enhanced data controls.

The ChatGPT Team plan is the professional tier of the world's most famous chatbot. It bridges the gap between the individual "Plus" user and the "Enterprise" tier, offering a collaborative workspace with data privacy guarantees. The Team plan is their answer to corporate security concerns. There have been useful studies on benchmarking the Performance of LLMs in processing of the Ukrainian language [72].

- **The Tool's Purpose:** Data Analysis and Excel formulas.
- **Relevance to journalism:** Enables teams to collaboratively analyse datasets and documents, quickly cleaning spreadsheets and generating charts during reporting. Useful for internal analysis and planning, but not suitable as a primary source or for highly sensitive material.
- **CDI Score:** Cost: 5/10 | Difficulty: 2/10 | Invasiveness: 8/10
  - **Comments:** The team plan is safer, but it is still invasive. It is one of the safest cloud AI tools available, but in this toolkit, invasiveness is about exposure depth, not vendor trust.
- **Time Dividend:**
  - *How time is saved:* Automates routine data tasks such as cleaning messy spreadsheets, generating basic charts, and writing formulas, reducing time spent wrestling with Excel mechanics.
  - *Ways to reinvest the time:* Use the saved time to interrogate what the data actually shows, identify gaps or anomalies, and check assumptions with the data's creators or subject-matter experts.

18. Perplexity [73]

AI-powered search engine that synthesises answers with citations from live web sources.

Perplexity is an engine that challenges the traditional Google Search paradigm [74]. Instead of a

---

[71] https://chat.openai.com

[72] https://apps.ucu.edu.ua/wp-content/uploads/2025/06/Eduard_Pekach_contribution.pdf

[73] https://www.perplexity.ai

[74] https://scet.berkeley.edu/aravind-srinivas-lessons-from-building-perplexity-ai/#:~:text=Perplexity%20Decoded%3A%20Lessons%20from%20the,and%20synthesize%20information%20through%20AI

list of links, it uses LLMs (ChatGPT, Claude) to read search results and synthesize a direct answer with citations.

Perplexity was created by people who previously worked at major tech companies like OpenAI and Google, with the goal of making online search less cluttered and more useful. However, it has also been criticised for collecting [75] information from websites in ways that some publishers do not agree with, which has caused tension with parts of the media industry. But they are not the only LLM doing this [76].

- **The tool's purpose:** The "Journalistic" Search Engine.
- **Relevance to journalism:** Provides *citations* for every claim. Excellent for background research on complex geopolitical topics.
- **CDI Score:** Cost: 3/10 | Difficulty: 1/10 | Invasiveness: 4/10
- **Time Dividend:**
  - *How time is saved:* Quickly summarizes and compares information from multiple web sources in one view, reducing the time spent opening, skimming, and synthesizing many separate pages.
  - *Ways to reinvest the time:* Verification and triangulation—use the saved time to open the cited sources, cross-check key claims against primary documents or trusted outlets, and build a short source list you would be comfortable publishing from.

19. Wordtune [77]

Writing assistant that rewrites sentences to adjust tone, clarity, and emphasis.

Wordtune is an AI-powered writing companion focused on phrasing and tone. Unlike a grammar checker, it rewrites sentences to be more casual, formal, concise, or expanded, acting as a "co-writer" rather than a proofreader. Excellent for overcoming writer's block or refining non-native English.

The tool's purpose: To help journalists quickly improve sentence clarity, tone, and flow without changing the underlying meaning of their reporting.

---

[75] https://techcrunch.com/2025/08/04/perplexity-accused-of-scraping-websites-that-explicitly-blocked-ai-scraping/#:~:text=AI%20startup%20Perplexity%20is%20crawling,to%20internet%20infrastructure%20provider%20Cloudflare

[76] https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-scrape-publisher-sites-licensing-2024-06-21/#:~:text=Exclusive%3A%20Multiple%20AI%20companies%20bypassing,scraping%20of%20their%20content

[77] https://www.wordtune.com

Relevance to journalism: Particularly useful for journalists writing in a second language or under deadline pressure, Wordtune helps refine copy while keeping editorial control with the reporter, not the tool.

- **CDI Score:** Cost: 4/10 | Difficulty: 1/10 | Invasiveness: 3/10
- **Time Dividend:**
  - *How is time saved:* Reduces editing rounds for English copy.
  - *Ways to reinvest the time:* Reporting. More time investigating, less time wrestling with syntax.

20. Gemini [78]

Google's large language model integrated into Docs, Drive, and multimedia analysis tools.

Gemini is Google's multimodal large language model, deeply integrated into Google Docs, Drive, Gmail, and YouTube. Its primary strength lies in working directly with content already stored in the Google ecosystem, allowing journalists to summarise long documents [79], analyse emails, or extract timelines without exporting files. However, this tight integration also creates dependency on Google's infrastructure, making it unsuitable for sensitive investigations or source material.

The Tool's Purpose: To help users quickly analyse, summarise, and work with documents, emails, and media already stored in Google's products.

Relevance to journalism: Useful for background research, desk reporting, and making sense of large document collections, but should be avoided for confidential reporting due to data control and platform dependency.

- **CDI Score:** Cost: 5/10 | Difficulty: 2/10 | Invasiveness: 7/10
- **Time Dividend:**
  - *How is time saved:* Analyzing personal archives stored in Google Drive.
  - *Ways to reinvest the time:* Chronology. Building timelines of events based on the emails/docs found.

21. OpenRefine [80]

Open-source tool for cleaning, clustering, and transforming messy datasets.

---

[78] https://gemini.google.com

[79] https://workspaceupdates.googleblog.com/2025/09/gemini-summaries-on-drive-home-page.html#:~:text=You%20can%20now%20see%20Gemini,in%20the%20Gemini%20side%20panel

[80] https://openrefine.org

OpenRefine is a free tool that helps journalists work with messy or poorly formatted data. It is especially useful for cleaning spreadsheets, fixing inconsistencies, and preparing data so it can be properly analysed. It is widely used in data journalism because many official datasets are incomplete, inconsistent, or badly structured.

Essential for cleaning government datasets where "Chisinau", "Chișinău", and "Kishinev" need to be merged into one entry.

The tool looks similar to a spreadsheet, but it is much more powerful. Originally developed with Google's support and later released to the public, OpenRefine is now maintained by a global community and is considered a standard tool for cleaning and preparing data before reporting.

- **The Tool's Purpose:** Cleaning messy data.
- **Relevance to journalism:** It allows you to spot patterns, merge different versions of the same name or place, and make large-scale changes safely.
- **CDI Score:** Cost: 0/10 | Difficulty: 6/10 | Invasiveness: 0/10
- **Time Dividend:**
  - *How is time saved:* Days of manual "find and replace."
  - *Ways to reinvest the time:* Once the data is clean, you can actually *interrogate* it to find the story.

22. Tabula [81]

Tool for extracting tables from PDF documents into CSV or Excel formats.

Tabula is a specialized tool designed to solve one specific, painful problem: extracting data tables from PDF files into CSV or Excel formats. It is widely used by investigative journalists to liberate data locked in government reports.

Tabula runs locally in a web browser. Users upload a PDF, draw a box around the table, and Tabula attempts to parse the layout and extract the text into cells.

- **The Tool's Purpose:** Liberating data from PDFs.
- **Relevance to journalism:** Governments love to publish budgets as scanned PDFs to hide data. These tools turn them back into Excel.
- **CDI Score:** Cost: 0/10 | Difficulty: 4/10 | Invasiveness: 0/10
- **Time Dividend:**
  - *How is time saved:* No manual re-typing of budget tables.
  - *Ways to reinvest the time:* Verify the government's totals. They often make

[81] https://tabula.technology

"mistakes" in the PDF summaries.

23. Pandas [82]

Python library for data analysis, manipulation, and statistical computation.

Pandas is the top alternative to Excel and OpenRefine for serious data analysis. It is an open-source Python library that provides high-performance, easy-to-use data structures and analysis tools. Pandas is better than Excel when data becomes large, complex, or needs to be handled responsibly. Excel is a visual spreadsheet designed for manual work, while Pandas is a data analysis engine built for scale and repeatability. With Pandas, every step of cleaning, filtering, or analysing data is written as code, which means the process is transparent, reusable, and auditable, a critical requirement for serious journalism. Pandas can handle millions of rows where Excel and OpenRefine often fail, and because it runs locally, journalists retain full control over their data.

The Tool's Purpose: Sovereign large-scale data analysis.

Relevance to journalism: Required when datasets exceed spreadsheet limits or require reproducible analysis.

- **CDI Score:** Cost: 0/10 | Difficulty: 7/10 | Invasiveness: 0/10

Time Dividend:

- *How is time saved:* Automates repetitive calculations and cleaning.
- *Ways to reinvest the time:* Story discovery. Use free time to explore anomalies instead of fixing formatting.

24. LanguageTool [83]

Grammar and style checker that can be run locally or self-hosted for privacy.

LanguageTool is the sovereign alternative to Grammarly. It is an open-source proofreading software that checks for grammar, style, and spelling errors in over 20 languages. Crucially, it can be hosted on a local server, ensuring no text leaves the organization.

The Tool's Purpose: Privacy-safe language correction.

Relevance to journalism: Allows proofreading without exposing drafts to third-party cloud

[82] https://pandas.pydata.org
[83] https://languagetool.org

services.

- **CDI Score:** Cost: 0/10 | Difficulty: 2/10 | Invasiveness: 0/10

Time Dividend:

- *How is time saved:* Automatically catches grammar, spelling, and consistency errors at draft stage, reducing back-and-forth between reporters and editors and shortening final edit cycles.
- *Ways to reinvest the time:* Editors can focus on story logic, sourcing, and narrative strength rather than mechanical corrections, improving overall journalistic quality.

25. Ollama [84]

Local LLM runtime for running open-weight language models on personal hardware.

Ollama is a lightweight runtime that makes it easy to download, manage, and run AI open-weight large language models locally on a personal computer. With a single terminal command, users can run models such as Llama 3, Mistral, Gemma, or DeepSeek entirely offline, without sending prompts or data to external servers. Ollama abstracts away much of the complexity of model hosting (weights, inference servers, model switching), making local AI experimentation accessible to non-specialists.

The Tool's Purpose: To run open-weight large language models directly on a personal computer, enabling local inference without reliance on cloud-based AI services.

Relevance to journalism: Allows journalists, researchers, and technologists to use full language models on their own hardware, avoiding data transmission to external servers and enabling experimentation, testing, and learning under complete local control.

- **CDI Score:** Cost: 0/10 | Difficulty: 6/10 | Invasiveness: 0/10
  - **Comments:** Models are open-weight and downloadable at no charge. It requires comfort with terminal commands and an understanding of local system resources, like RAM and storage. The bonus is that models execute entirely on the user's machine. So, prompts and data do not leave the device unless explicitly configured otherwise.

Time Dividend:

How time is saved: Removes dependence on cloud access by allowing an LLM to run directly on a personal computer, enabling instant local prompts, testing, and iteration without uploads or network latency.

*Ways to reinvest the time: U*se the saved time to learn how models behave, compare outputs

[84] https://ollama.com

across models, and prototype workflows before deciding what (if anything) belongs in the cloud.

Teaching guidance

Make authorship and accountability explicit. Students should never submit AI-assisted writing without being able to explain how the output was shaped, checked, and edited.

Watch for overconfidence. Fluent AI output often masks shallow or incorrect analysis, so require students to justify key claims with source references.

Use this cluster to distinguish between understanding and producing text. The learning goal is comprehension, not speed.

Exercise 1: Question the document

Students upload a complex document (policy, budget, judgment) and ask an AI system ten structured questions about it. They submit the answers alongside a critique of which questions produced useful insight and which produced misleading or shallow responses.

Exercise 2: Test the AI Detectors

This exercise is designed to test the AI detection tools and realize their shortcomings.

1) Generate a piece of AI text in CHATGPT (or another LLM)
2) Get a piece of text written definitely by a human.
3) Run both through the AI detection tool, GPTZERO and see the results.
4) Try and humanize the AI text either with a tool called The Humaniser or by changing it manually (by adding errors and words to make the AI text seem more "human").

Relevant links:

https://gptzero.me/
https://chatgpt.com/g/g-2azCVmXdy-ai-humanizer

Where should you start? (cost, difficulty and what it teaches)

Start with: Perplexity + ChatGPT Team (or equivalent controlled LLM access)

Why these first

- Extremely low technical barrier
- Immediate usefulness for overwhelmed students

- Familiar interfaces reduce fear and friction
- Suitable for short classroom exercises

What they teach well

- How to quickly orient oneself in a complex topic. Always require students to click citations (Perplexity) or trace claims back to the original document. This is where editorial discipline is taught.
- The difference between *background understanding* and *original reporting*
- How AI-generated summaries must be interrogated, not trusted

Introduce next: Claude (Opus or Sonnet) + Wordtune

Why these next

- Claude excels at long documents and nuanced tone
- Wordtune is intuitive and especially valuable for non-native English speakers
- Both tools reinforce the idea of the journalist as **editor**, not generator

What they teach well

- How framing, tone, and emphasis change meaning
- How to work with AI as a drafting assistant, not an author
- How "good writing" can still contain factual errors

Tool cluster 4: Audio, Video & Social

Applying AI to multimedia storytelling and social distribution while maintaining editorial control.

Many newsrooms now produce one story across multiple formats: an article, a podcast segment, a short video, and social clips. AI tools are used to generate subtitles for videos, create audio versions of articles for visually impaired audiences, or cut long interviews into short, shareable clips. A regional newsroom might turn a written investigation into a narrated video explainer to reach younger audiences on social platforms. These tools matter because they shape who can access journalism and how stories travel, while also raising important questions about transparency when synthetic visuals or voices are used.

26. ElevenLabs [85]

[85] https://elevenlabs.io

AI voice synthesis and dubbing platform for realistic text-to-speech and multilingual audio.

ElevenLabs is the market leader in AI voice synthesis. It creates hyper-realistic text-to-speech (TTS) and offers voice cloning capabilities that are virtually indistinguishable from human speech.

The platform offers "Voice Design" to create custom voices and "Voice Cloning" to replicate a specific person's voice from a short sample. Its "Dubbing Studio" automatically translates video audio into other languages while preserving the original speaker's voice characteristics and syncing the lips (in some modes).

ElevenLabs has set the benchmark for quality but also sparked ethical debates regarding deepfakes. Their "Voice Isolator" and safety measures attempt to mitigate misuse, but the tool remains the primary engine for both high-end content creation and unauthorized voice cloning.

The languages offered depends on the model of ElevenLabs that you are using [86]. For Multilingual v2 you have 29 languages. These include Arabic, Bulgarian, Chinese (Mandarin), Croatian, Czech, Danish, Dutch, English, Filipino, Finnish, French, German, Greek, Hindi, Indonesian, Italian, Japanese, Korean, Malay, Polish, Portuguese, Romanian, Russian, Slovak, Spanish, Swedish, Tamil, Turkish and Ukrainian. For Flash v2.5 you get 32 languages (with Hungarian, Norwegian and Vietnamese added). For Eleven v3 you get a full 74 languages. Adding Afrikaans, Armenian, Assamese, Azerbaijani, Belarusian, Bengali, Bosnian, Catalan, Cebuano, Chichewa, Estonian, Galician, Georgian, Gujarati, Hausa, Hebrew, Icelandic, Irish, Javanese, Kannada, Kazakh, Kirghiz (Kyrgyz), Latvian, Lingala, Lithuanian, Luxembourgish, Malayalam, Marathi, Nepali, Pashto, Persian, Punjabi, Sindhi, Somali, Swahili, Telugu, Thai, Urdu and Welsh.

- **The Tool's Purpose:** Text-to-Speech (TTS) for accessibility and audio reporting.
- **Relevance to journalism:** Excellent for creating audio versions of articles for the visually impaired or for "radio" broadcasts in regions with low literacy.
- **CDI Score:** Cost: 6/10 | Difficulty: 3/10 | Invasiveness: 5/10
- **Time Dividend:**
    - *How time is saved:* Eliminates the need for a recording studio and voice actor.
    - *Ways to reinvest the time:* Audio requires a different writing style than print. Use the time to rewrite the article for the *ear* (shorter sentences, active verbs). And also your specific voice and way of speaking.

---

[86] https://help.elevenlabs.io/hc/en-us/articles/13313366263441-What-languages-do-you-support

27. Descript [87]

Audio and video editor that allows users to edit media by editing text transcripts.

Descript reimagines video and audio editing by treating it like a word processor. Instead of cutting waveforms on a timeline, users edit the transcript text, and the AV is cut automatically. Descript pioneered the text-based editing paradigm. It lowers the barrier to entry for podcast and video production significantly, democratizing high-quality editing.

Incredible speed for dialogue editing. Text-based workflow is intuitive. However, it can be buggy/resource-heavy on large projects, and the "text-first" approach is less precise for frame-perfect visual cuts than Premiere Pro.

- **The tool's purpose:** To make audio and video editing faster and more accessible by allowing journalists to edit recordings through text rather than complex timelines.
- **Relevance to journalism:** Ideal for podcasts, interviews, and explainer videos, Descript enables reporters and editors to produce polished audio and video without specialist editing skills, while retaining editorial control over what is cut and kept.
- **CDI Score:** Cost: 5/10 | Difficulty: 3/10 | Invasiveness: 4/10
- **Time Dividend:**
  - *How time is saved:* Rough cuts take minutes instead of hours.
  - *Ways to reinvest the time:* Improving the sound design and add ambient noise and music to make the story immersive.

28. OpusClip [88]

AI tool that extracts short, shareable video clips from long-form content.

OpusClip is a specialized generative AI tool that repurposes long-form video (podcasts, webinars) into viral-style short clips (Shorts, Reels, TikToks).

It uses NLP to analyze the video transcript and identify "hooks" or viral moments. It automatically reframes horizontal video to vertical (9:16), keeping the speaker in focus (active speaker detection).
OpusClip targets the "creator economy" need for volume. It automates the tedious task of finding highlights, allowing creators to scale content output without a dedicated editing team.

---

[87] https://www.descript.com

[88] https://www.opus.pro

- **The Tool's Purpose:** Viral short creation.
- **Relevance to journalism:** Automatically cuts long Zoom interviews into vertical TikToks with captions. Critical for reaching youth globally.
- **CDI Score:** Cost: 4/10 | Difficulty: 1/10 | Invasiveness: 5/10
- **Time Dividend:**
- *How time is saved:* 90% of editing time for socials.
- *Ways to reinvest the time:* More community management. Take the time to reply to the comments on TikTok.

29. Midjourney [89]

AI image generation tool accessed via Discord, focused on high-quality visual illustration.

Midjourney is widely regarded as the most artistically capable AI image generator. Accessible primarily via Discord, it generates high-fidelity, creative, and photorealistic images from text prompts.

- **The tool's purpose:** Conceptual illustration.
- **Relevance to journalism:** Creating visuals for abstract topics like "inflation," "cyberwar," or "corruption" when all the stock photos you find are boring.
- **CDI Score:** Cost: 5/10 | Difficulty: 6/10 | Invasiveness: 5/10
- **Time Dividend:**
  - *How time is saved:* No need to buy stock photos.
  - *Ways to reinvest the time:* Workflow reliability—use the saved time to run a proper visual sign-off process (editor approval, keep originals, document any edits) and to publish faster across multiple formats without losing accuracy or trust.

30. Adobe Firefly [90]

Commercially safe generative AI integrated into Adobe's creative tools.

Adobe Firefly is integrated into Photoshop and Illustrator and offers image generation that is safe for commercial use, trained on Adobe Stock images rather than scraped internet data [91].

---

[89] https://www.midjourney.com

[90] https://www.adobe.com/products/firefly.html

[91] https://www.adobe.com/products/firefly/discover/firefly-vs-stable-diffusion.html

Firefly powers the "Generative Fill" tool in Photoshop, allowing users to extend images or add/remove objects with perfect lighting matching. In this context, “Generative Fill” means using generative AI to create new pixels that complete or modify an image based on a user’s intent or description. You are not just cloning or patching an image like you would with traditional Photoshop tools. It analyses the existing image and produce new content that matches the surrounding lighting, perspective and style wherever you make a selection. Its "Structure Reference" allows users to upload a sketch and generate an image that matches the layout. It now includes a video model.

- **The Tool's Purpose:** "Safe" generative fill.
- **Relevance to journalism:** Extending the background of a photo to fit a website banner. "Copyright safe" for commercial newsrooms.
- **CDI Score:** Cost: 6/10 | Difficulty: 2/10 | Invasiveness: 4/10
- **Time Dividend:**
  - *How time is saved:* Automates complex visual edits—background extension, object removal, and layout adjustments—that would normally require advanced Photoshop skills and significant manual time.
  - *Ways to reinvest the time:* Stronger visual storytelling—use the saved time to select better images, improve captions and context, and collaborate more closely between editors and designers on how visuals support the story.

31. Buffer AI Assistant [92]

Social media scheduling and caption-generation tool integrated into Buffer.

Buffer is a social media management platform, and its AI Assistant is integrated to help draft, repurpose, and schedule content across platforms.

The AI helps generate post ideas, rewrite content for different networks (e.g., shorten for X, expand for LinkedIn), and summarize long articles into threads. It uses OpenAI's API but is tuned for social media best practices.

The Tool’s Purpose: To help newsrooms plan, adapt, and publish content across multiple social platforms efficiently from a single interface.

Relevance to journalism: Supports consistent, responsible distribution of reporting by adapting headlines and summaries for different platforms, while allowing journalists to retain

[92] https://buffer.com

editorial judgment over tone, timing, and context rather than chasing engagement manually.

- **CDI Score:** Cost: 3/10 | Difficulty: 2/10 | Invasiveness: 3/10
- **Time Dividend:**
  - *How time is saved:* Writing 10 variations of a tweet for different platforms.
  - *Ways to reinvest the time:* Analyzing *when* your audience is online and *what* they engage with.

32. Runway [93]

AI-powered video generation and editing platform for text-to-video and visual effects.

Runway is a research company pioneering AI video generation. Their Gen-2 and Gen-3 Alpha models allow for text-to-video, video-to-video, and image-to-video creation, effectively acting as an AI film studio.

Runway created the original Stable Diffusion model. They are a core innovator in the space, often releasing features months before Adobe or OpenAI. They focus on professional filmmakers, not just casual users.

- **The Tool's Purpose:** B-roll generation.
- **Relevance to journalism:** Generating "generic" video (e.g., "a gloomy street in a city") when filming is impossible.
- **CDI Score:** Cost: 7/10 | Difficulty: 5/10 | Invasiveness: 5/10
- **Time Dividend:**
  - *How time is saved:* Rapid generation of illustrative video and visual sequences without filming, location shoots, or sourcing stock footage.
  - *Ways to reinvest the time:* Use the saved production time to verify facts, add context, and apply clear labeling so synthetic visuals are not mistaken for real footage.

Teaching guidance

Stress that format choices are editorial decisions. Changing a story's medium changes how it is understood, and students must be able to explain those choices.

Be clear about transparency requirements when synthetic media is involved. Students should

[93] https://runwayml.com

practice writing disclosure notes alongside content, not as an afterthought.

Encourage reflection on audience impact. This cluster is an opportunity to discuss reach, accessibility, and responsibility together. [94]

Exercise 1: Format reshaping

Students take one reported story and use AI tools to produce an audio version and a short social video. They submit all versions and a short reflection on how tone, emphasis, and meaning changed across formats.

Exercise 2: The misleading clip test

Students are given a long video interview or press conference and asked to use AI tools to generate three short clips suitable for social media. They must then analyse each clip and explain how editing choices (length, captions, framing) could change audience interpretation or introduce misrepresentation. The submission includes the clips and a written justification of which version, if any, is publishable.

Where should you start? (cost, difficulty and what it teaches)

Start with: Descript + Buffer AI Assistant

Why these first

- Moderate cost, low technical barrier
- Text-based workflows feel familiar to journalists
- Immediate relevance to everyday newsroom output
- Allow experimentation without synthetic media risks

What they teach well

- How one piece of reporting travels across formats
- The relationship between transcript, audio, and video
- How tone and framing shift between article, clip, and post
- That distribution is an editorial decision, not a marketing afterthought

Introduce next: OpusClip (with guardrails)

Why this next

[94] https://www.presshop.com/post/top-ai-tools-every-journalist-should-master-for-2025-and-beyond

- Extremely easy to use, OpusClip is ideal for teaching **scale and reach**, but must be framed carefully.
- Demonstrates how AI identifies “hooks” and attention patterns
- Highly relevant for reaching younger audiences

What it teaches well

- How algorithms prioritise moments over context
- The trade-off between reach and nuance
- Why viral formats can distort meaning

Tool cluster 5: Security Challenges

Understanding the security risks of AI tools and how to protect data, sources, and journalists.

Journalists often upload sensitive material into AI systems without fully understanding where that data goes. For example, a reporter might upload an interview transcript containing a whistleblower’s name into a cloud-based writing assistant, or a student newsroom might process sensitive documents through an online summarisation tool. Security-focused tools and practices help journalists encrypt files, remove metadata, and process information locally rather than on external servers. This cluster matters because even a strong investigation can be undermined if sources are exposed, communications are intercepted, or data is unintentionally shared with third parties.

33. VeraCrypt [95]

Open-source disk encryption software for protecting files and entire drives.

VeraCrypt is the gold standard for open-source disk encryption. It allows users to create encrypted file containers or encrypt entire hard drives/partitions, rendering data inaccessible without the key.

Maintained by IDRIX (France), VeraCrypt is audited and trusted by the security community. In an era of border checks and device seizures, it is the primary defense for data at rest.

The Tool’s Purpose: To protect files and entire devices by encrypting them so that data cannot be accessed without the correct password or key.

Relevance to journalism: Essential for safeguarding sensitive reporting materials, source

[95] https://www.veracrypt.fr

identities, and investigations if a laptop, hard drive, or USB stick is lost, stolen, or seized.

- **CDI Score:** Cost: **0/10** | Difficulty: **6/10** | Invasiveness: **0/10**
- **Time Dividend:**
  - *How is time saved:* Prevents time-consuming crisis response after device loss or seizure, such as damage control, source notification, or legal risk mitigation.
  - *Ways to reinvest the time:* Journalists can work on sensitive stories without constantly worrying about physical device security or worst-case scenarios.

34. Dangerzone [96]

Tool for safely opening and sanitising documents from untrusted sources.

Dangerzone is a security tool designed specifically for journalists handling untrusted documents from anonymous sources. It converts potentially malicious PDFs, Word files, or images into safe, inert PDFs by opening them in a sandboxed environment and reconstructing them as images. Developed by Micah Lee at the Freedom of the Press Foundation, Dangerzone assumes all leaked documents are hostile until proven otherwise.

The Tool's Purpose: Safe handling of leaks and anonymous submissions.

Relevance to journalism: Prevents malware attacks embedded in documents—one of the most common attack vectors against journalists.

- **CDI Score:** Cost: 0/10 | Difficulty: 3/10 | Invasiveness: 0/10

Time Dividend:
How is time saved: Avoids catastrophic compromise rather than saving time.
Ways to reinvest the time: Secure verification. Clean documents first, investigate second.

35. ProtonMail [97]

Encrypted email service designed for secure communications.

ProtonMail is an encrypted email service designed with privacy in mind. Based in Switzerland, it provides end-to-end encryption (E2EE) for emails between ProtonMail users, meaning the messages are encrypted on the sender's device and can only be decrypted by the recipient. Even when emailing non-ProtonMail contacts, ProtonMail allows you to send a password-protected encrypted message that the recipient can decrypt with a shared password. The service is open source and has a user-friendly interface similar to mainstream email, making secure email accessible without special technical knowledge.

The Tool's Purpose: Secure email for sensitive journalistic communications.

---

[96] https://dangerzone.rocks
[97] https://proton.me/mail

Relevance to journalism: Prevents third parties—including email providers—from accessing message content, reducing exposure when handling confidential information or whistleblower contact.

- **CDI Score:** Cost: 2/10 | Difficulty: 3/10 | Invasiveness: 2/10

Time Dividend:

- *How is time saved:* Reduces time spent mitigating email breaches or compromised inboxes.
- *Ways to reinvest the time:* Editorial focus. Journalists can concentrate on reporting instead of digital damage control.

36. Tor Browser [98]

Privacy-focused browser that anonymizes internet traffic via the Tor network.

Tor Browser is a modified Firefox browser that routes your internet traffic through the Tor network, which is a decentralized system of volunteer-run servers. It encrypts and bounces your communications through multiple nodes, making it very difficult for anyone to trace your browsing or messaging back to you. This allows journalists to anonymize their web activity and access sites that may be censored or blocked on local networks. For journalists, Tor is invaluable for protecting the confidentiality of online research and communications. It enables reaching information on the “open” or dark web anonymously – for example, accessing whistleblower platforms or publishing outlets via their .onion (Tor-only) sites. It helps keep browsing habits and source communications hidden from government surveillance or ISP tracking, which is crucial in investigative reporting and in authoritarian regimes.

The Tool's Purpose: Anonymous research and censorship-resistant web access.

Relevance to journalism: Protects journalists from network surveillance when researching sensitive topics, accessing blocked websites, or visiting whistleblower platforms.

- **CDI Score:** Cost: 0/10 | Difficulty: 4/10 | Invasiveness: 2/10

Time Dividend:

- *How is time saved:* Avoids repeated access blocks, monitoring alerts, or compromised research sessions.
- *Ways to reinvest the time:* Investigative depth. Journalists can research freely without self-censorship.

37. Bitwarden [99]

[98] https://www.torproject.org
[99] https://bitwarden.com

Open-source password manager with end-to-end encryption.

Bitwarden is a popular open-source password manager that securely stores all your passwords in an encrypted vault. It helps generate strong, unique passwords for each account and auto-fills them when needed, so you don't have to remember or reuse passwords. All data stored in Bitwarden is end-to-end encrypted (zero-knowledge), meaning even Bitwarden's servers can't read your passwords. The service has undergone independent security audits in 2023 and 2024 to ensure its integrity. Journalists often manage numerous online accounts (email, social media, cloud services, databases). A password manager is vital for protecting these accounts from hacking. Using strong, unique passwords for every account is one of the simplest yet most effective security steps

The Tool's Purpose: Secure password storage and account protection.

Relevance to journalism: Prevents account compromise caused by weak or reused passwords—one of the most common ways journalists lose access to email, cloud storage, and social accounts.

- **CDI Score:** Cost: 0/10 | Difficulty: 2/10 | Invasiveness: 1/10

Time Dividend:

- *How is time saved:* Eliminates password resets, account lockouts, and recovery processes.
- *Ways to reinvest the time:* Operational stability. Journalists spend time reporting, not regaining access to tools.

Teaching guidance

Avoid teaching security as a technical specialty. Frame it instead as a situational judgment that every journalist must exercise.

Encourage students to think in terms of worst-case consequences rather than average use. Security decisions are shaped by who might be harmed if something goes wrong.

Reinforce that convenience is not neutral. Choosing an easier tool can be an editorial risk, even when the reporting is strong.

Exercise 1: Compare the security of two tools

Each student selects a function of AI tools and then chooses two examples. They then investigate the data security of these tools. Ask the questions: what data is entered, where it goes, who could access it, and what the worst-case consequences would be in a sensitive reporting scenario. Give a conclusion of which of your two tools is safer and why?

Exercise 2: Safe vs unsafe workflow

Students are given the same reporting task twice — once involving a public figure and once involving a vulnerable source. They must design two different AI workflows and explain which tools are safe in each case and why.

Where should you start? (cost, difficulty, and what it teaches)

Start with: Bitwarden + ProtonMail

This pairing establishes a foundation: journalists cannot protect sources if they cannot protect their own accounts.

Why these first:

- Free or low-cost tools with a very low technical barrier
- Immediately relevant to how journalists already work (email and accounts)
- Easy to adopt without changing hardware or operating systems
- Address the most common real-world security failures in newsrooms

What they teach well:

- That weak or reused passwords are a major source of compromise
- How encrypted communication differs from ordinary email
- The limits of encryption (what is protected vs what metadata remains exposed)
- That basic digital hygiene is a professional responsibility, not an optional extra

Introduce next: Tor Browser.

Tor helps students understand that research itself can be risky, even before anything is published.

Why this next:

- Moderate increase in conceptual complexity, but still accessible
- Introduces anonymity and surveillance awareness without technical setup
- Directly relevant to investigative research and censored environments

What it teaches well:

- How browsing habits can expose investigations
- The difference between secure content and observable behaviour
- When anonymity is necessary and when it is excessive
- How state or network-level monitoring affects journalism

Tool cluster 6: Building your own tools with "vibe-coding"

Using AI-assisted coding to build simple newsroom tools without relying on vendors.

Some newsrooms now build small, purpose-specific tools rather than waiting for commercial products. Journalists have used AI-assisted coding to build scrapers that monitor government tenders, calculators that show how policy changes affect household budgets, or internal databases that track ongoing investigations. A student newsroom might build a simple tool to collect and visualise housing complaints from public records. These tools matter because they allow journalism to respond directly to editorial needs, reduce dependence on external platforms, and help journalists understand — and control — the infrastructure behind their reporting.

The most profound shift in the AI era is the democratization of software development. Journalists and editors, previously dependent on scarce engineering resources or third-party vendors, can now build their own tools using low-code platforms and AI coding assistants. This shift from "buying" to "building" allows newsrooms to create bespoke micro-tools that solve specific, local problems.

The huge drawback of vibe-coding is that the cost seems manageable, but it can run away from you as you can easily max out your monthly quota (but sending more prompts) and so run up a high bill. The paradox is that even though vibe-coding is cheaper than employing a developer, plenty of the costs you incur can be from trying to fix mistakes that the model you are prompting did by mistake. Vibe-coding is a process of proposing an idea and maybe a solution and then hurrying after the model as it does it amazing work and then dozens of mistakes that you need to identify and fix. It can feel like you are running eternally in circles.

Before we dive into the tools, let's outline examples of successful tools that were coded and created in-house to give you a sense of the potential.

1. La Silla Vacía (Colombia)

La Silla Vacía developed [100] a centralised hub of AI-powered assistants tailored for newsroom workflows [101]. Modules include a style editor aligned with their editorial guidelines, a generator for daily news round-ups, a draft writer, and tools for building X (formerly Twitter) threads — all integrated into one custom system. This is a *bespoke newsroom platform*, not a generic tool.

This helps a mid-size newsroom automate repetitive editorial tasks and maintain consistent

[100] https://www.youtube.com/watch?v=K4gcU4rb52w
[101] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024/la-silla-vacia

quality while freeing reporters to focus on reporting and engagement.

2. Nawaat (Tunisia)

Nawaat, a long-standing digital outlet in Tunisia, created Nawaat AI [102], a content platform that enhances audience engagement with Tunisian history and current affairs, including archive exploration features and new interactive formats. This custom system helps readers *interact with context and background*, moving beyond static articles to deeper engagement — something no off-the-shelf CMS did for them.

3) The Quint (India)

The Quint has publicly discussed [103] AI-driven features they built in-house to enhance reader experience, including AI-generated article summaries and interactive modules that are *created internally* for their editorial workflow (not simply adopting a consumer tool).

4) Briefly News's "Editorial Eye" (South Africa)

This South African newsroom [104] supported by an AI adoption mentorship program developed tools like "Editorial Eye", an internal AI assistant for grammar, style, and proofreading, created in collaboration with newsroom teams rather than simply licensing a product.

5. Agencia Mural (Brazil)

Agencia Mural developed Local Climate Alert [105], an AI-based system that delivers concise, geo-targeted climate risk updates to local communities. It monitors climate data and creates actionable alerts relevant to specific neighbourhoods. It's a newsroom-created, real-world alert tool that translates complex data into direct, time-sensitive information for residents — an example of *AI for civic engagement and safety*, not just content production.

38. Custom GPTs [106]

Configurable AI assistants with fixed instructions, optional tools, and persistent context for repeated newsroom tasks.

First, this is not strictly "coding", but Custom GPTs are an entry point to progressing from standard LLMs. Custom GPTs are AI assistants created and used inside ChatGPT. They allow journalists to define a stable role, upload reference materials, and reuse the same instructions across multiple sessions. Unlike agents, Custom GPTs do not run on schedules,

[102] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024
[103] https://wan-ifra.org/2025/06/ai-tools-help-the-quint-drive-engagement-subscriptions/
[104] https://iafrica.com/south-african-newsrooms-adopt-ai-to-boost-productivity-and-strengthen-editorial-standards/
[105] https://www.journalismai.info/programmes/innovation/innovation-challenge-2024
[106] https://openai.com/index/introducing-gpts/

monitor information, or act independently; they respond only when prompted by a human.

Their strength lies in consistency rather than autonomy: a Custom GPT behaves the same way every time it is used, making it suitable for repeatable analytical or editorial tasks.

The tool's purpose: Reusable editorial assistants with clearly bounded scope.

Relevance to journalism: Useful for editors and investigative teams who need consistent analytical support over time, such as archive interrogation, policy comparison, style consistency checks, or structured document analysis.

CDI Score: Cost: 4/10 | **Difficulty:** 2/10 | **Invasiveness:** 6/10

Comments: Runs entirely within the OpenAI platform. Must be paired with strict internal rules governing what documents, notes, or unpublished material may be uploaded.

Time Dividend:
How time is saved: Removes the need to repeatedly re-prompt, re-explain context, or restate editorial constraints.

Ways to reinvest the time: Focused verification, interpretation, and editorial decision-making rather than repetitive setup work.

39. Cursor [107]

AI-first code editor that allows users to build software using natural language prompts.

Cursor is a desktop application that lets you build and modify digital tools by talking to the computer in natural language. Instead of writing complex code line by line, you describe what you want the tool to do, and Cursor generates or edits the technical parts for you.

Cursor "reads" all the files in your project, so when you ask it to make a change, it understands the full context — not just one screen or document. This makes it possible to build or adjust tools gradually, even if you don't fully understand how they work under the hood.

Cursor can work with different AI systems (such as ChatGPT or Claude) and can also be set up to run entirely on your own computer, which is important for sensitive journalism.

- **The Tool's Purpose:** A professional-grade tool for building custom newsroom software using plain language.
- **Relevance to journalism:** Cursor makes it possible for journalists to build or adapt their own investigative tools instead of waiting for a developer.
- **CDI Score:** Cost: 5/10 | Difficulty: 4/10 | Invasiveness: 2/10
    - Comments: Free tier available and the Pro plan is a monthly subscription. It

[107] https://cursor.sh

requires installing a code editor. And there is a "Privacy Mode" where no code is stored on their servers, crucial for sensitive investigations.

- **Time Dividend**:
    - *How is time saved:* Weeks of waiting for a developer to build a simple tool.
    - *Ways to reinvest the time:* Understanding and maintaining your tools. Instead of treating software as a black box, you can ask Cursor to explain what the tool does and update it when a website, dataset, or source changes, keeping your reporting infrastructure alive and under your control.

40. Replit [108]

Browser-based coding platform for rapid prototyping and deployment of small applications.

Replit is a collaborative, browser-based integrated development environment (IDE). It allows users to write, run, and deploy code instantly from any device without setup.

Founded by Amjad Masad, Replit aims to democratize coding. It is particularly popular for rapid prototyping and education.

- **The Tool's Purpose:** Zero-setup web app builder.
- **Relevance to journalism:** Perfect for quick, public-facing projects. Example: A Moldovan newsroom wants to build a simple "Inflation Calculator" for their readers. Replit's "AI Agent" can build and host this in 10 minutes.
- **CDI Score:** Cost: 5/10 | Difficulty: 2/10 | Invasiveness: 6/10
    - **Comments:** Core plan needed for the best AI Agent features. And it is cloud-hosted, so code and data live on the Replit servers.
- **Time Dividend:**
    - **How time is saved:** You don't have to worry about publishing the tool online. Replit handles all of that automatically. No servers, no technical setup, no launch delays.
    - *Ways to reinvest the time:* Use the time you saved to test the tool with real users, improve the questions it answers, and make sure it genuinely helps people understand the story, instead of wrestling with technical problems.

41. Google Colab [109]

Cloud-hosted Jupyter Notebook environment for running Python and data analysis code.

Google Colab is a hosted Jupyter Notebook service that provides free access to computing resources, including GPUs (Graphics Processing Units) and TPUs (Tensor Processing Units). Colab allows researchers to write and execute Python code in the browser. Colab has lowered the barrier to entry for AI research globally. It is the "playground" where much of the

[108] https://replit.com
[109] https://colab.research.google.com

open-source AI revolution is tested and shared.

No setup. Standard for sharing reproducible research. However, runtime disconnects when you close the browser (so you can't use this for long production jobs) and it has privacy issues (code runs on Google servers).

- **The Tool's Purpose:** Data Science & Visualization "Vibes."
- **Relevance to journalism:** Journalists can upload a messy government dataset and use the built-in "Gemini Code Assist" to write the Python code to visualize it (e.g., "Make a heatmap of this corruption data").
- **CDI Score:** Cost: 0/10 | Difficulty: 3/10 | Invasiveness: 5/10
- **Time Dividend:**
    - *How is time saved:* "I don't know how to make a map in Python" is solved in seconds.
    - *Ways to reinvest the time:* Data Verification. The AI might hallucinate the map coordinates. Check the outliers manually.

Teaching guidance

Make it clear that coding skill is not the learning outcome. Problem definition, editorial purpose, and ethical reasoning matter more than working code.

Resist tool worship. Students should be encouraged to critique their own tools as rigorously as they would critique a source.

Use this cluster to teach independence. The goal is to help students see that journalism infrastructure can be shaped, not just consumed.

This cluster should not be taught as "learning to code." That framing scares students. Instead, teach this as editorial problem-solving with software, where AI acts as a translator between journalistic intent and code. The learning objective is *agency*, not technical mastery.

Students should leave understanding:

- When it makes sense to build a small tool instead of using a platform
- How AI-assisted coding lowers the barrier to experimentation
- Why owning your tooling matters for independence, safety, and sustainability
- How to critically inspect AI-generated code, not blindly trust it

Exercise 1: Build a minimum prototype

Using AI-assisted coding, students build a very basic prototype (e.g. form, scraper, calculator). Assessment focuses on whether the tool does what it claims, not on technical

elegance.

Exercise 2: Ethics and failure analysis

Students analyse their own prototype (or a classmate's) and write a short report on how it could fail journalistically — including data errors, bias, misuse, or harm — and how those risks could be mitigated.

Where should you start? (cost, difficulty and what it teaches)

Start with: Replit (AI Agent)

Why these first

- Very low technical barrier; no installation required
- Browser-based environment feels safe and reversible for beginners
- Fast feedback loop — students see a working tool within minutes
- Encourages experimentation without needing to understand infrastructure
- Ideal for short workshops and mixed-ability classrooms

What it teaches well

- How to translate an editorial problem into a software request
- That tools can be small, specific, and purpose-built
- The difference between *describing* a problem and *solving* it
- How user inputs, logic, and outputs connect
- That journalists can prototype public-interest tools without vendors

Introduce next: Google Colab

Why this next

- Builds on familiarity with data and spreadsheets
- Low cost and widely used in journalism education
- Excellent for showing how AI assists with analysis, not just writing
- Encourages reproducibility and transparency when used correctly

What it teaches well

- How AI can help structure messy data into stories
- The relationship between datasets, code, and visual output
- Why analysis choices shape editorial conclusions
- The importance of checking AI-generated logic and visuals
- That "seeing the data" is an editorial act

Tool cluster 7: AI Agents & Automated Workflows

You can use AI agents to coordinate tasks, monitor information and assist editorial workflows. The key is to do this without surrendering judgment or too much control.

AI agents are a step beyond single-task AI tools. Instead of responding to one prompt and stopping, an agent can be instructed to pursue an objective over time: checking sources regularly, combining multiple tools, maintaining memory and triggering actions when conditions are met.

If a large language model is best thought of as a very capable "brain" that answers questions when you speak to it, an AI agent is that same brain given a role, a goal, and permission to act repeatedly in the background. This is the moment when you may think this sounds like prompting. It is important to distinguish an AI agent from simply giving a large language model a role in a prompt. A prompted LLM is reactive: it responds once, with no persistence, no awareness of time, and no ability to act unless a human asks again. An agent, by contrast, is a system built around the model that runs continuously or on defined triggers. This feature of understanding time is important. It has external memory, can check the same sources repeatedly, decide when something has changed. The language model provides reasoning, but the agency comes from the surrounding infrastructure (schedules, state, tools and thresholds) not from the prompt itself.

Instead of waiting for a prompt, an agent runs in a loop: it checks what it is supposed to watch, decides whether anything has changed, uses tools if needed (search, databases, scrapers, internal archives), remembers what it has already seen, and only surfaces something to a human when it meets a defined threshold. In practice, this makes an agent feel less like a chatbot and more like a junior researcher or assistant who has been given standing instructions and reports back when something genuinely matters.

Here are some examples of what agents are capable of:

- An agent can monitor court registries or procurement portals and alert reporters when new documents appear.
- An agent can be used to cross-reference names across previously published stories, leaked datasets and company registries.
- An agent can track whether published articles are being republished, plagiarized or misquoted elsewhere online.
- Maintain a "beat assistant" trained on a specific topic (e.g. energy policy, housing law).
- Run structured checks on draft articles (e.g. "flag unclear sourcing", "identify unverifiable claims").
- Interrogate large internal document collections repeatedly without re-prompting from scratch.

Agents can influence agenda-setting, tip discovery, and the allocation of human attention.

Used well, they create time for reporters to do good work. Used poorly, they create opaque systems that newsroom staff no longer fully understand.

42. ChatGPT Atlas [110]

AI-enhanced web browser with built-in agents for goal-driven, multi-step web tasks.

ChatGPT Atlas is a browser developed by OpenAI [111] that integrates ChatGPT directly into the browsing experience. In addition to page-level assistance (summarisation, explanation, comparison), Atlas includes an Agent Mode that allows the AI to carry out multi-step tasks across the web, such as navigating sites, opening links, collecting information, and reporting results back to the user.

Unlike traditional AI assistants, Atlas can act within the browser itself, executing a sequence of actions toward a defined goal. These actions remain bounded to the current session and require human instruction and oversight; Atlas does not operate persistently or in the background.

The tool's purpose: Goal-driven browser automation under direct human supervision.

Relevance to journalism: Useful for research and verification tasks that involve navigating complex or fragmented web interfaces, such as tracking regulatory updates, locating public records across multiple sites, comparing disclosures, or gathering background information that would otherwise require repetitive manual browsing.

CDI Score: Cost: 4/10 | Difficulty: 2/10 | Invasiveness: 4/10

Comments: Runs entirely within the OpenAI ecosystem and performs actions on behalf of the user inside the browser. Should only be used with publicly accessible information and non-sensitive accounts. Journalists remain responsible for verifying outputs and ensuring compliance with terms of service and ethical standards.

Time Dividend:
How time is saved: Automates multi-step web navigation and information gathering that would otherwise require manual searching and copying.
Ways to reinvest the time: Validating findings, cross-checking sources, and deciding which surfaced information warrants deeper reporting.

43. Apify [112]

[110] https://chatgpt.com/atlas/
[111] https://www.theguardian.com/technology/2025/oct/21/openai-chatgpt-web-browser-atlas
[112] https://apify.com/

Prebuilt agents for monitoring and scraping websites, collecting data, and detecting changes over time.

Apify provides ready-made and customisable agents ("actors") that run on schedules, monitor websites for changes, extract structured data, and trigger downstream actions. Unlike conversational assistants, Apify agents operate continuously in the background and surface results without repeated human prompting.

When paired with simple rules or an LLM for interpretation, Apify enables true agentic behaviour.

The tool's purpose: Background agents for continuous web monitoring and data collection.

Relevance to journalism: Highly suitable for monitoring court rolls, procurement portals, company disclosures, government notices, platform policy pages, and other sources that change unpredictably. Particularly valuable for investigations that depend on catching updates early.

CDI Score: Cost: 4/10 | **Difficulty:** 3/10 | **Invasiveness:** 3/10

Comments: Easier to deploy than agent frameworks and does not require deep engineering skills. Ethical and legal boundaries around scraping must be clearly understood. Best used with human review queues rather than automatic publication.

Time Dividend:
How time is saved: Automates continuous checking of high-value websites and surfaces changes immediately.
Ways to reinvest the time: Responding faster to emerging documents and changes, rather than discovering them late.

44. AutoGen [113]

Open-source framework for Agentic AI

AutoGen is a system for setting up multiple AI assistants that talk to each other, divide up work and report back to a human when a task is complete. This system enables multiple agents to collaborate toward a goal, each with defined roles (in a newsroom this could be a researcher, verifier or summariser).

The key part here is that the agents can negotiate with each other, critique each other's outputs, call tools and escalate results to a human editor when confidence thresholds are

[113] https://github.com/microsoft/autogen

met.

- **The tool's purpose:** Coordinated agent teams with internal checks.
- **Relevance to journalism:** Useful for complex investigations where agents perform different bounded tasks (document scanning, cross-referencing, contradiction detection) before presenting findings for human review.
- **CDI Score:** Cost: 1/10 | Difficulty: 7/10 | Invasiveness: 2/10
- **Comments:** AutoGen is open-source with an MIT license. It runs on your own infrastructure with the only hard cost being the usage of your LLM API. It's powerful but dangerous if misused. It must be deployed with strict limits on scope, recursion and output authority.
- **Time Dividend:**
  How time is saved: Runs multiple AI roles in parallel to surface contradictions, missing evidence, and follow-up questions in the first pass.
  Ways to reinvest the time: Early human review of conflicts, gaps, and follow-up leads flagged by collaborating agents.

45. n8n [114]

Open-source workflow automation platform that can manage AI tools, scrapers and databases.

n8n is not an AI agent by itself, but it is one of the most important agent foundations available to newsrooms. It allows teams to build workflows where AI tools are triggered automatically: scraping a website, sending text to an LLM, storing results, and notifying a journalist.

It is used for activities like monitoring government websites for updates, pulling new documents into a review queue and automatically tagging and categorizing incoming material.

- **The tool's purpose:** A workflow control layer that enables automation without surrendering oversight.
- **Relevance to journalism:** Essential for investigative and data teams that want agent-like behaviour without surrendering control.
- **CDI Score:** Cost: 1/10 | Difficulty: 7/10 | Invasiveness: 1/10
- **Comments:** Because n8n can be self-hosted, it is one of the few ways to build sovereign agents that never expose data externally. Requires technical support. Not suitable for casual users.
- **Time Dividend:**
  How time is saved: Eliminates manual monitoring and repetitive checks.
  Ways to reinvest the time: Pattern recognition and editorial decision-making.

---

[114] https://n8n.io

46. Zapier AI [115]

Low-code automation platform with AI-powered actions and conditional workflows.

Zapier AI allows users to build simple, rule-based workflows that combine AI actions with everyday newsroom tools such as email, Google Docs, Slack, Airtable, and CMS platforms. While not autonomous in the research sense, Zapier workflows function as **bounded agents**: they watch for triggers, apply logic, and take predefined actions repeatedly and reliably.

Unlike experimental agent frameworks, Zapier workflows are *explicit*, *auditable*, and *non-recursive*. The system never "decides" to do something new, but rather executes what editors have approved in advance.

- **The tool's purpose:** Safe automation of repetitive newsroom processes.
- **Relevance to journalism:** Common uses include monitoring RSS feeds, tagging incoming material, routing tips to the right desk, and generating first-pass summaries for editors.
- **CDI Score:** Cost: 5/10 | Difficulty: 2/10 | Invasiveness: 2/10
- **Comments:** Zapier AI can become costly at scale, especially with high-volume workflows and AI actions tied to external APIs. The 5/10 cost rating is adequate for small or occasional use, but it doesn't reflect pricing when you start automating frequently or adding large numbers of AI steps. Zapier is ideal for teaching automation without losing editorial control.

Time Dividend:

- **How time is saved:** Removes manual copying, pasting, forwarding, and status checking.
- **Ways to reinvest the time:** Editorial judgment, follow-up reporting, and source development.

Teaching guidance

When chatting to your students, emphasise that agents do not replace editorial judgment, they reallocate attention. Students should understand that deciding what an agent watches is itself an editorial act.

And be sure, when talking to your students, to stress the difference between automation which is safe and predictable and autonomy which is unpredictable and risky.

---

[115] https://zapier.com

Also, when talking to students, stress that they need to map workflows before introducing agents. This can be even with a pen and paper. And this is good advice for vibe-coding as well. Even if AI can handle the details and code, it still struggles with pulling out to the macro. So, tell students if they cannot explain a process clearly, they are not ready to automate it.

Students should leave this cluster understanding:

1. **Agents can appear consistent while being wrong.** When multiple agents agree or reinforce each other's outputs, errors can look more credible rather than less. Journalists must recognise false consensus and step in before agreement is mistaken for verification.
2. **Agents follow instructions, not intent.** An agent will execute rules exactly as defined, even when circumstances change or ethical context shifts. Journalists must recognise when rigid automation is producing technically correct but editorially inappropriate results and intervene.

Exercise 1: Define the handover point

Students are given a short description of an agentic workflow (e.g. an agent that monitors council agendas and flags agenda items that match predefined keywords).

Students must identify:

- What the agent is allowed to do on its own
- The exact moment a human must take over and this speaks to the point above around needing to design workflows (even if just on pen and paper) before you can start automating and part of this is knowing where these handover points are.
- Where is the point that should force human review.

Learning outcome: Teaches boundary-setting and intervention points without system design.

Exercise 2: Map a failure to a decision

Students are given a brief failure scenario. For example, an agent misses a late-posted document or flags a routine item as urgent.

Students need to determine:

- *What decision the agent made correctly*
- *What decision it made incorrectly*
- *Which human decision could have prevented harm*

Learning outcome: Shifts focus from "what went wrong" to "where responsibility sits".

Where should you start?

Start with: Claude Projects or Custom GPTs. Why: Low technical barrier, immediate usefulness. What it teaches well: How persistent context changes analysis. Why instructions matter more than prompts. Best teaching use: Investigations, policy reporting, archive work.

Introduce next: n8n (selectively). Why: Shows how agents are actually built in practice. What it teaches well: Systems thinking, accountability, and auditability. Avoid initially: Fully autonomous agents. Why: They teach the wrong lessons first, spectacle over responsibility.

Addendum A: Tools with Risks

1) The Transcription Trap

In the high-pressure environment of a newsroom, the transcription of interviews is a bottleneck. For decades, this was a manual, labour-intensive process. The arrival of AI has promised to reclaim hours of productivity. However, this convenience (in certain cases and with certain tools) comes with a severe hidden cost: the loss of custody over source data.

Otter.ai

Tool Overview: Otter.ai is the market leader in consumer AI transcription, known for its "OtterPilot" feature that automatically joins meetings to record and transcribe conversations.

What is the risk? The primary risk with Otter.ai is its architecture of aggressive data ingestion. The "OtterPilot" bot can join meetings automatically if calendar permissions are granted, potentially recording off-the-record pre-ambles or conversations where not all parties have consented. [116] In 2024 and 2025, the company faced class-action litigation[117] in the United States alleging violations of wiretapping laws, specifically regarding the non-consensual recording of voice biometrics. The data resides on US-based servers, subject to foreign subpoenas. And more critically, the privacy policy allows for the use of audio data to train the company's AI models. This means a whistleblower's voice is processed, tokenized, and potentially retained within the model's training corpus.

2) The Writing Process

Grammarly

Tool Overview: A cloud-based writing assistant that checks spelling, grammar, and tone. Grammarly functions by transmitting every keystroke in an active text field to its cloud servers for analysis. For a journalist investigating state corruption or organized crime, this is an operational security failure. The text of a draft investigation, containing names, dates, and allegations, is exposed to Grammarly's infrastructure before it is even published.

---

[116] https://www.workplaceprivacyreport.com/2025/08/articles/artificial-intelligence/ai-notetaking-tools-under-fire-lessons-from-the-otter-ai-class-action-complaint/

[117] https://roninlegalconsulting.com/hey-who-let-the-ai-in-a-closer-look-at-the-otter-ai-lawsuit/

What is the risk? The 2024–2025 integration of generative AI features increases the likelihood that user data feeds into broader language models. [118] Universities and corporations with high security requirements often ban Grammarly for this exact reason.

3) Visual Journalism

Visual storytelling is dominated by short-form video (TikTok, Reels). The tools required to compete in this algorithmic arena—offering auto-captions, filters, and snappy transitions are predominantly mobile-first and data-hungry. The risk here is biometric harvesting.

CapCut

Tool Overview: The most popular mobile video editor, owned by ByteDance (parent of TikTok).

What is the risk? CapCut is a consumer video-editing application whose data collection and licensing practices have raised significant privacy and surveillance concerns among regulators and legal commentators. In 2024-2025, it faced class-action lawsuits in the US (Illinois) for violating the Biometric Information Privacy Act (BIPA). [119] The lawsuits allege that CapCut collects biometric identifiers, including facial and voice data, as well as device and location information, without obtaining consent required under Illinois law. [120]. CapCut updated its Terms of Service on June 12, 2025, which have drawn legal and public criticism[121] because they grant broad rights to user-generated content, including rights to use, edit, reproduce, and exploit content worldwide, with implications for biometric features (these include facial recognition and voice patterns). This update expanded the scope of what CapCut can legally do with uploaded content without separate user consent. For a journalist in the CEE region, the geopolitical risk is acute. ByteDance is subject to Chinese security laws.[122]

4) The Communication Infrastructure

For certain parts of the world, Telegram is not just a messenger app, it is the internet. It is the primary source of news, alerts, and social organization. However, for a journalist, relying on it for private communication is a dangerous fallacy.

---

[118] https://www.ndsmcobserver.com/article/2024/11/university-classifies-grammarly-as-generative-ai
[119] https://www.mediapost.com/publications/article/403907/bytedances-capcut-must-face-privacy-suit.html
[120] https://news.bloomberglaw.com/privacy-and-data-security/capcut-users-advance-lawsuit-over-bytedance-data-collection
[121] https://2b-advice.com/en/2025/07/04/capcut-trouble-over-new-terms-of-service-legal-risks-lurk-here/
[122] https://jsis.washington.edu/news/u-s-tiktok-ban-national-security-and-civil-liberties-concerns/

Telegram

Tool Overview: The fatal flaw of Telegram is that Cloud Chats are not End-to-End Encrypted by default. Telegram holds the decryption keys. [123] This means the company, and by extension, any entity that can coerce the company, can read the messages. Only "Secret Chats" are true end-to-end encrypted (E2EE) and must be manually enabled[124].

The Better Alternative: Signal

Tool Overview: An encrypted messenger owned by a non-profit foundation. Signal minimizes metadata retention [125] . Even if subpoenaed, Signal can typically only produce the date of account creation and the date of last login. It is the gold standard for contact.

Addendum B: AI & The Strategic Reinvestment of Time

The most dangerous outcome of AI adoption is the Jevons Paradox: where increased efficiency simply leads to increased consumption (or in this case, production of low-value content), leaving staff just as burned out as before. If AI saves a reporter 5 hours a week, and the newsroom simply demands 5 more articles of the same quality, the opportunity for transformation is lost.

The strategic goal of AI in the newsroom is not to reduce headcount (fire staff) but to *upgrade the value of human labour*. [126] This section outlines a strategy for the "AI Dividend". This is the deliberate reinvestment of saved time into high-value activities that bolster the newsroom's offering.

Quantifying the Dividend

[123] https://onerep.com/blog/is-telegram-safe
[124] https://blog.invitemember.com/how-do-telegram-secret-chats-work/
[125] https://freedom.press/digisec/blog/metadata-102/
[126] https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2025

The Thomson Reuters 2024 Future of Professionals report indicates that AI tools can save professionals approximately 12 hours per week [127] by 2029. The report likened this to adding a colleague for every 10 team members, highlighting how productivity gains can aggregate across an organisation.

The "Stop Doing" List: A Prerequisite for Reinvestment

Reinvestment cannot happen without a deliberate cessation of low-value tasks [128]. Leadership must sanction a "Stop Doing" list to create the capacity for new work. This requires a rigorous audit of current newsroom activities. The critical management challenge is tracking this time and ensuring it is not absorbed by "shallow work" (emails, Slack, browsing).

Strategic Reinvestment Areas

Strategy 1: Investigative & Enterprise

Time saved from churning out "commodity news" (press release rewrites, basic event coverage) should be directed toward *original* reporting.

- If a reporter saves 5 hours a week on transcription, mandate that those 5 hours are blocked off for time dedicated solely to long-term investigation, source building, or public records requests.

Strategy 2: Community Listening and Engagement

AI cannot empathize. It cannot have coffee with a source, attend a town hall just to "read the room," or comfort a grieving family member.

- Reinvest time into physical presence. Send reporters out of the newsroom not just to cover stories, but to *find* them. Organize town halls, listener panels, and meet-and-greets.
- Trust is rebuilt through human connection. If AI handles the digital noise, humans must handle the analog signal. This creates a "moat" around the publication that AI aggregators cannot cross.

Strategy 3: New Product Development

Use the technical capacity built (via Section 3) to launch new editorial products that were

---

[127] https://www.thomsonreuters.com/en/press-releases/2024/july/ai-set-to-save-professionals-12-hours-per-week-by-2029

[128] https://www.forbes.com/sites/cherylrobinson/2025/04/26/high-impact-leadership-starts-with-a-stop-doing-list-not-a-to-do-list/

previously too resource-intensive.

- Launch a hyper-local newsletter, a niche podcast, or a data dashboard tracking local economic indicators.
- These products drive subscriptions and loyalty, moving the business model away from reliance on high-volume, low-value programmatic ad traffic. With AI assisting in the production (e.g., summarizing articles for a newsletter), a small team can support a product that previously required a large staff.

Addendum C: Futureproofing & Updating the Curriculum

The AI Toolkit is intentionally not a fixed list of tools. AI systems change frequently. Features are added or removed, pricing models shift, companies are acquired and data-use policies evolve. This addendum provides a practical, low-burden approach to curriculum maintenance that does not require technical expertise.

1. A Layered Approach to Curriculum Updates

Rather than constant monitoring, educators are encouraged to use a layered update cycle that balances awareness with sustainability.

Ongoing informal awareness

Educators should maintain general awareness of changes in the AI and journalism landscape through everyday signals:

- Colleagues mentioning new tools or features
- Students referencing emerging platforms
- Headlines, social media posts, or conference discussions

These informal cues are often the earliest indicators that something has changed. At this stage, no action is required beyond noticing.

Monthly

Once per month, educators should conduct a short review (15–30 minutes) focused on the tools actively referenced in their teaching. Key questions include:

- Have any tools changed pricing, access levels, or ownership?
- Have any tools been discontinued, merged, or significantly altered?
- Have there been notable changes to data-use or privacy policies?

Some institutions now formalize this as a brief “tool sanity check” to catch disruptions before they affect students.

Annual deep review

Once per year, a deeper review should be undertaken to reassess the toolkit as a whole. This is the appropriate moment to:

- Retire tools that are obsolete, unstable, or no longer trustworthy
- Introduce tools that have matured and proven their value
- Revisit ethical, legal, and data-protection considerations
- Update warnings, caveats, or teaching notes where needed

2. Tracking Tools and Trends Responsibly

Educators do not need to track the entire AI ecosystem. The focus should remain on tools that serve journalistic work and align with editorial values

Example of a tool that compares LLMs

LLM Arena (also known as LMArena, formerly *Chatbot Arena*) is a public, crowd-sourced platform for evaluating and comparing large language models (LLMs) [129]. On the platform, users can submit the same prompt to *two anonymous AI models* and vote on which response they think is better. These pairwise comparisons are aggregated into a leaderboard using an Elo-style rating system, allowing models to be ranked by performance based on *real human preference data* rather than just static metrics.

Monitoring tool stability

Simple habits are sufficient:

- Subscribe to one or two general AI or journalism newsletters
- Set basic alerts for the names of tools used in class
- Bookmark official blogs or status pages for key tools and review them periodically

If a tool loses functionality, changes its business model, or disappears, it should be reassessed or removed from teaching materials.

Evaluating new tools

When a new AI tool comes to attention, educators should assess it using three criteria:

1. **Usability.** Can students and educators learn it quickly?
2. **Relevance.** Does it support real journalistic tasks such as research, writing, editing, or verification?
3. **Responsibility.** Does it respect privacy, consent, accuracy, and transparency?

Educators are encouraged to test tools briefly using classroom-relevant examples rather than relying on marketing claims. Popularity alone is not a reason to include a tool.

---

[129] https://lmarena.ai/

Maintaining trust and compliance

Educators should remain alert to:

- Changes in how tools handle user data
- New requirements for uploads, permissions, or third-party sharing
- Shifts from free access to paid or restricted models

If a tool's data practices change in ways that conflict with classroom expectations, its use should be paused and reconsidered. Maintaining trust with students is essential.

Addendum D: Resources & Research Database

A curated collection of the most recent and impactful research papers, policy briefs, and training courses regarding AI in the media, updated for the 2025 academic year.

1. Selected Research on the Impact of AI on Media

- The State of The Bots (Q1 2025). "RAG bot scrapes now exceed Training bot scrapes." [130] A critical report on how AI bots are aggressively scraping news publisher data, often ignoring robots.txt protocols.
- AI Across America: Attitudes On AI Usage (August 2025) [131]. A massive survey (Report #116) revealing that 50% of US adults now use AI tools, with significant data on public trust and fear of job displacement.
- Generative AI Outlook Report (2025)[132] . European Commission JRC. Maps the path for AI policy and innovation in Europe, warning that without ethical oversight, the EU risks falling behind.
- Artificial Intelligence in Eastern Africa Newsrooms (March 2025) – CIPESA.[133] An analysis of perceived prevalence, ethics, and impact of AI on media freedom in Ethiopia, Kenya, Tanzania, and Uganda.
- Generative AI & Journalism: Content, Perceptions, and Audiences (Feb 2025) – RMIT University. [134] A study across seven countries (including Germany and UK) examining

[130] https://www.scribd.com/document/944556714/TollBit-State-of-the-Bots-Q1-2025
[131] https://www.chip50.org/reports/ai-across-america-attitudes-on-ai-usage-job-impact-and-federal-regulation
[132] https://digital-strategy.ec.europa.eu/en/news/generative-ai-set-transform-eu-economy-requires-further-policy-action
[133] https://www.eaeditors.org/publications
[134] https://figshare.com/articles/online_resource/Generative_AI_and_Journalism_Content_Journalistic_Per

how audiences perceive AI-generated news content and the transparency paradox.

- Generative AI in Journalism: The Evolution of Newswork (April 2024) – Associated Press. [135] The definitive report on how AP and other major newsrooms are adapting ethical guidelines for the generative age.
- Hallucination Mitigation Using Agentic AI (Jan 2025). [136] Technical research on using multiple AI agents to check each other's work to reduce "hallucinations".

2. Policy, Ethics & Legal Frameworks

- AI in the Work of an Attorney-at-Law (2025).[137] Recommendations on how legal professionals (and by extension, investigative journalists dealing with legal data) should use AI tools securely.
- M20 Policy Brief 4: AI, Africa and the G20 (July 2025) A critique of the "power concentration" of AI in the Global North and recommendations for protecting African data sovereignty. [138]
- The Paris Charter on AI and Journalism (Nov 2023). [139] A foundational ethical document signed by 16 partner organizations (RSF) defining principles for AI in news.
- Exposing Technology-Facilitated Gender-Based Violence (2023) – UNESCO. [140] How generative AI is being weaponized to harass women journalists and public figures.

3. Online Courses & Educational Resources

- Google News Initiative: Pinpoint & Data Tools. Using AI to search through thousands of PDFs and handwritten documents.[141]
- MIT: Introduction to Deep Learning (6.S191). A technical deep dive for those wanting to understand the architecture behind the models. [142]

4. Tools & Open Source Repositories

- Hugging Face: Journalists' Toolkit[143] . Finding open-source models to run locally for

---

ceptions_and_Audience_Experiences/28068008
[135] https://www.trust.org/wp-content/uploads/2025/01/TRF-Insights-Journalism-in-the-AI-Era.pdf
[136] https://arxiv.org/abs/2501.13946
[137] https://arxiv.org/abs/2501.13946
[138] https://media20.org/2025/07/10/m20-policy-brief-4-power-politics-and-economics-ai-africa-and-the-g20/
[139] https://rsf.org/en/paris-charter-ai-and-journalism
[140] https://www.unesco.org/en/articles/your-opinion-doesnt-matter-anyway
[141] https://newsinitiative.withgoogle.com/resources/trainings/google-ai-tools/
[142] https://www.youtube.com/watch?v=alfdI7S6wCY
[143] https://huggingface.co/spaces/JournalistsonHF/ai-toolkit

privacy.

- Journalism Cloud Alliance (OCCRP/GFMD)[144] . Shared infrastructure for investigative newsrooms to lower cloud costs and increase security.

5. Organisations Leading AI Innovation in Global South

- Centre for Information Integrity in Africa (CINIA).[145] Combatting disinformation and safeguarding democratic values in the African digital space.
- Vambo AI.[146] Bridging the language gap with AI solutions for indigenous African languages.
- SADiLaR (South African Centre for Digital Language Resources). [147] Creating digital resources for under-resourced languages.
- Lelapa AI. "AI for Africans, by Africans" building resource-efficient language models. [148]

6. Podcasts & Multimedia

- Reuters Institute: Digital News Report 2025 Launch (Africa). Discussion on trust, AI usage, and news consumption trends [149] .
- Inside the FT's AI-Powered Story Finding. [150] How the Financial Times uses AI not just to write, but to find stories in data.

7. Legacy Research (Foundational Reading)

- New Powers, New Responsibilities (2019) – LSE Polis. The seminal global survey that launched the JournalismAI initiative.[151]
- News Automation: The Rewards, Risks and Realities (2019) – WAN-IFRA. An early but relevant guide to the practicalities of "robot journalism." [152]
- The Next Wave of Disruption (2021) – IMS. Emerging market media use of AI

---

[144] https://gfmd.info/cloud-alliance/
[145] https://cinia.africa/
[146] https://www.vambo.ai/
[147] https://sadilar.org/en/
[148] https://lelapa.ai/
[149] https://www.youtube.com/watch?v=LM9JwQ1jSbM
[150] https://medium.com/ft-product-technology/inside-the-fts-ai-powered-story-finding-efe6d154f20b
[151] https://www.journalismai.info/research/2019-new-powers-new-responsibilities
[152] https://wan-ifra.org/insight/report-news-automation-the-rewards-risks-and-realities-of-machine-journalism/

153

Addendum E: Who to follow online

- Nicholas Diakopoulos [154] . Professor at Northwestern University and Director of the Computational Journalism Lab (CJL) and their newsletter, Generative AI in the Newsroom [155] .

- Ethan Mollick[156]. Excellent at translating generative AI into practical newsroom and storytelling workflows.

- Charlie Beckett [157]. Director at LSE's Polis; writes clearly about AI, power, trust, and journalism ethics.

- Rasmus Kleis Nielsen[158] . Focuses on how audiences, platforms, and AI reshape journalism economics and trust.

- Bettina Peters [159] . Strong on newsroom AI adoption, governance, and real-world constraints.

- Simon Rogers[160]. Deep experience in data-driven storytelling and automation at scale.

- Meredith Broussard [161]. Essential counterweight to hype; strong on why AI often fails journalists and communities.

- Anya Schiffrin [162]. Focuses on media sustainability, regulation, and AI's political consequences.

- Chris Moran [163]. One of the best public thinkers inside a major broadcaster on AI experimentation.

---

[153] https://www.mediasupport.org/publication/the-next-wave-of-disruption-emerging-market-media-use-of-artificial-intelligence-and-machine-learning
[154] https://www.linkedin.com/in/nicholas-diakopoulos-83848114/
[155] https://generative-ai-newsroom.com/
[156] https://www.linkedin.com/in/emollick

[157] https://www.linkedin.com/in/charlie-beckett-2743234/
[158] https://www.linkedin.com/in/rasmus-kleis-nielsen-5b97663/
[159] https://www.linkedin.com/in/bettina-peters-06a80b5/
[160] https://www.linkedin.com/in/simonrogersjournalist/

[161] https://www.linkedin.com/in/meredithbroussard/
[162] https://www.linkedin.com/in/anya-schiffrin-1459097/
[163] https://www.linkedin.com/in/chris-moran-6634a87/

- Emily Bell [164]. Columbia Journalism School; essential reading on platforms, AI, and media power.

[164] https://www.linkedin.com/in/emily-bell-7091663/