

AI Applications in Investigative Journalism

CNTI's Global AI & Journalism Research Group Briefing #4 | June 2026

This is the fourth in a series of reports from the AI and Journalism Research Working Group convened by the Center for News, Technology & Innovation (CNTI). The working group currently consists of 21 cross-industry members from around the world, bringing research, journalism and technology expertise to the discussions.

The goal of the working group is to offer succinct summaries of global research on specific topics at the intersection of journalism and AI. Each quarter, the working group will synthesize the state of research across two to three topics for journalism practitioners, researchers and industry leaders around the world, focusing on actionable recommendations for journalism — not other fields that are concerned with AI.

In each report, we lay out the general findings of the research to date, suggested considerations and/or actions for practitioners and areas where more or new research is needed. This report was prepared by the research and professional staff of CNTI in partnership with several external contributors who collectively authored this briefing. If you have ideas or research findings that are important for CNTI and the working group to include, please email them to info@cnti.org.

What do we mean by “AI”?

This report uses the OECD definition: *“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.”*

Wherever possible, we try to use specific terms rather than “AI” to avoid conflation or confusion. Journalism has been adopting forms of automation for more than 50 years,¹ but widespread use of the term “AI” is more recent — and may include both newer technologies and those that have been in use for quite some time.

Investigative journalism plays a critical watchdog role in holding power to account, and AI tools — ranging from automated document parsing to Large Language Models — are both enhancing and complicating that role. While AI systems can make investigations more efficient and scalable, they are not uniformly adopted due to constraints such as data quality, technical expertise, source confidentiality, economic resources and global inequalities. At the same time, AI companies and technologies themselves are becoming a subject of investigation, requiring journalists to scrutinize opaque algorithmic systems and powerful technology actors. Understanding these dynamics is essential for practitioners, researchers and industry leaders seeking to responsibly integrate new tools into investigative workflows while preserving journalistic integrity, independence and impact.

CNTI's AI and Journalism Research Working Group examined how artificial intelligence (AI) is reshaping investigative journalism, drawing on a review of 44 recent academic and industry studies. This briefing synthesizes global research to assess both the opportunities and structural challenges AI introduces into investigative reporting.

Research we reviewed suggests

- AI tools are transforming investigative journalism foremost by expanding its capacity to handle large-scale data and investigate complex systems. They also redefine what counts as evidence, what is investigable and how journalistic knowledge is produced.
- AI tools do not replace core journalistic functions, which remain deeply human and context-dependent.
- Adoption of AI tools is shaped by economic constraints, skill gaps, language barriers and global inequalities, as well as by newsroom hierarchies, organizational culture, digital divides and external infrastructures. Each investigation is somewhat idiosyncratic, which means that customized AI tools provide more value than off-the-shelf solutions.
- Using AI tools in investigative journalism requires collaborative efforts that bring together people from diverse backgrounds and different areas of expertise.
- AI itself opens a major new area of investigative journalism around holding institutions accountable for decisions made by AI systems.

Investigative journalism has a distinct position within the news ecosystem, characterized — at least under ideal conditions — by extended timelines,² exceptionally rigorous verification standards and the uncovering of hidden or suppressed information often with moral or legal implications.³ In academic research, investigative journalism is described as “*activist, reformer, and exposé*”⁴ that is inherently adversarial, whose goal is to delineate crimes, identify victims and abusers, and advocate for social justice. Unlike routine news, which often recycles available information, investigative reporting begins with a hypothesis or a “*denunciation*” that must be confirmed or diluted through meticulous research and the collection of evidence, including on-the-ground reporting, confidential-source work and analysis of public records.⁵

These characteristics highlight why investigative journalism presents a uniquely complex environment for adopting new forms of automation. While AI can augment journalists’ capabilities through large-scale data processing, pattern detection, document handling and information verification, it cannot (as of now) replace the core epistemic and professional functions of investigative reporting. Investigative work often involves standalone cases and thus unique and project-specific methods of investigation: in-person “*shoe leather*” reporting, “*dirty*” or “*incomplete*” data with “*messy*” formats, hidden or withheld data,⁶ multisource verification, and high legal and ethical stakes that require human oversight given the need for high accuracy and contextual judgment.⁷ Current AI systems rely on existing data and patterns and therefore “*can neither produce nor verify*”⁸ genuinely new information uncovered through investigative work that is not yet available in digital form.

As a result, investigative journalists tend to view AI less as a substitute for human labor and more as a tool that supports time-consuming tasks to allow reporters to focus on uniquely human aspects of the craft, such as source-building, persuasion and accountability reporting.

Altogether, the field of investigative journalism is currently navigating this tension between opportunities and limitations.

From Data and Computational Journalism to AI-Driven Investigations

A central consensus across the literature is that AI systems in investigative journalism should be understood within the broader trajectory of data and computational journalism. Investigative journalism grounds technological practices like computational journalism within its core watchdog function and normative mission.⁹ In fact, investigative journalism has historically been at the forefront of adopting computational methods. Early computer-assisted reporting (CAR) in the 1960s used social science methods, spreadsheets and statistical analysis for long-term projects,¹⁰ practices that evolved into modern AI-driven techniques, such as machine learning and large-scale data processing.¹¹ As information increasingly shifted to digital formats, investigative journalists turned to computational tools to manage growing volumes of data, exemplified by large-scale projects like the Panama Papers. That project took more than a year, involved millions of documents and required techniques such as optical character recognition (OCR) and algorithmic sorting, as well as large-scale indexing and search capabilities that would otherwise be impossible for human teams to perform.¹²

These technology-based approaches enabled new pathways to story discovery through pattern recognition, anomaly detection¹³ and large-scale data linkage,¹⁴ while also expanding verification practices through methods like open-source intelligence (OSINT), including satellite analysis and digital forensics.¹⁵ Computational journalism in this context applies “*computational thinking*,” meaning the abstraction of investigative tasks into granular, computable elements that can be processed algorithmically.¹⁶

Normative goals — namely accountability, exposure and reform — shape the role of AI systems in investigative journalism, enhancing the watchdog function and efficiency. In this sense, AI tools and data-driven methods are not ends in themselves but tools that support journalism’s public-interest mission by enabling the analysis of large and complex datasets, uncovering hidden patterns and strengthening transparency.

Use Cases

The range of applications of AI systems in investigative journalism represents targeted interventions across specific stages of the investigative process.

Finding Needles in Haystacks

Machine learning (ML) techniques can reduce massive datasets into actionable leads. For example, Swiss broadcaster SRF trained a supervised classifier on labeled data of authentic and fraudulent Instagram profiles to expose how influencers frequently buy fake engagement.

At the front end, AI tools are particularly valuable for establishing the scale of problems and “*finding needles in haystacks*.”¹⁷ Two studies surveying the uses of AI in investigative journalism¹⁸

show that ML is used to reduce massive datasets into actionable leads through data cleaning, record linkage and text-as-data techniques such as sentiment analysis, entity extraction, topic modeling and similarity detection (e.g., Locality-Sensitive Hashing). Another practice-based study with eight investigative journalists in Germany¹⁹ showed that journalists see particular value in tools that could automate repetitive web-based research tasks, such as monitoring multiple websites, scraping unstructured local-government records, tracking changes in online content over time, extracting information from social media or leaked materials, and converting collected material into searchable or analyzable formats.

Examples

AI tools have several practical uses in investigative journalism. For instance, a supervised classifier was developed by the Swiss broadcaster SRF. By training the model on labeled data of authentic and fraudulent Instagram profiles, researchers were able to analyze millions of accounts, exposing how influencers frequently buy fake engagement. In the United States, BuzzFeed News developed ML models to detect anomalous flight patterns, such as repetitive circling over specific locations, to uncover the scale of covert “spy plane” surveillance, while the Atlanta Journal-Constitution used ML-assisted filtering to reduce more than 100,000 regulatory documents to about 6,000 likely cases of sexual misconduct by doctors, exposing systemic regulatory failures. Similarly, reporters at Grist used predictive modeling in the “Waves of Abandonment” project to forecast which oil wells in Texas were likely to be abandoned, enabling forward-looking accountability reporting on environmental and financial risks.

These workflows depend on foundational processes of extraction, matching and cleaning — described by data journalists interviewed in a multi-country (U.S., Canada, the U.K. and India) mixed-methods, practice-based study²⁰ as “MacGyvering” messy data. Using tools such as OpenRefine, Google Pinpoint, Python scripts, fuzzy matching and clustering, journalists integrate fragmented and inconsistent datasets. For example, these tools were used in investigations where ML clustering and fuzzy matching linked government records with leaked documents to reveal procurement collusion networks in East African municipalities.²¹

AI systems also extend investigative capacity into multimodal and spatial domains through computer vision (CV) and remote sensing. In “satellite journalism,” reporters combine CV models with geospatial data to monitor physical changes that are otherwise difficult to observe.²²

Examples

The visual investigations team at *The New York Times* used object-detection techniques on high-resolution satellite imagery to identify craters and blast signatures consistent with 2,000-pound bombs in southern Gaza. In Southeast Asia, journalists collaborated with scientists using platforms like Google Earth Engine and Sentinel Hub to detect deforestation hotspots in protected areas, layering satellite detections with land-use registries and corporate concession maps to trace responsibility.²³ Another example, examined in an article that reviews applications of AI systems in newsrooms using the JournalismAI database,²⁴ is a project in Ukraine where a news outlet called Texty applied CV models to satellite imagery to identify “lunar-like” landscapes caused by illegal amber mining.

Algorithmic Accountability

Journalists can use AI tools to test opaque systems and platforms at scale. For example, [this report](#) found that Spain's social security agency used opaque AI systems to make high-stakes sick-leave decisions affecting millions of people despite poor accuracy, limited transparency, and little public accountability of the so-called "AI Doctors."

AI tools play a central role in algorithmic accountability and investigative work in constrained environments. While this briefing does not include a comprehensive look at investigative journalism about AI systems, it does include cases where journalists use AI tools to investigate opaque platform systems. In a practice-based study conducted in a Dutch context²⁵ they describe using AI tools to audit and reverse-engineer opaque platform systems as "fighting fire with fire." In collaboration with De Groene Amsterdammer, these researchers created sock-puppet accounts and used a vision-language model to simulate user behavior on TikTok, demonstrating that the platform could infer user interests within seconds and begin amplifying eating-disorder content within minutes. In a separate audit of a Google dataset, the same researchers used the ML tool Homepage2vec to classify 2.5 million Dutch-language websites included in large language model training data and uncovered significant proportions of copyrighted journalism, conspiracy content and even leaked personal data. Complementing these efforts, [ProPublica's "Machine Bias" investigation](#) revealed systemic racial disparities in algorithms used in the criminal justice system.

Reconstruction & Network Mapping*

In addition to performing algorithmic accountability through audits, AI tools also support modeling and reconstruction in transnational investigations. For example, [Forensic Architecture](#) trained computer vision models to identify tear gas canisters across millions of online images, using 3D modeling and synthetic image generation to reconstruct the global deployment of the weapons. In another practice-based study—a collaboration of data scientists, AI experts, and journalists with the Norwegian Association for Investigative Journalism—graph databases and AI techniques were used to map financial and regulatory networks in Norway's petroleum sector.²⁶

Subterfuge and Source Protection

In regions with limited press freedom, the use of generative AI becomes closely tied to practices of subterfuge and source protection. A focus group study of 34 investigative journalists across five East African countries showed that AI tools are enabling journalists to investigate indirectly through data analysis, simulation and anonymization while minimizing exposure.²⁷

Innovative Storytelling

Finally, AI tools offer new storytelling formats — such as automated text generation, personalization and versioning using Natural Language Generation (NLG) and generative AI — expanding how investigative findings are produced and communicated to diverse audiences.²⁸

* Corrections: The "Reconstruction and Networking Mapping" section was revised after publication (on July 14, 2026) because one of the examples used computational methods that may not meet our definition of AI. An additional minor wording change was made for clarification.

AI Limits and Constraints in Investigative Journalism

The constraints and limitations of AI systems in investigative journalism fall under five interrelated domains: data availability and quality; professional identity and dependence on external actors; technical and epistemic limitations; linguistic barriers; and economic factors.

Data Availability and Quality

One of the most significant limitations of AI tools in investigative journalism is the lack of accessible, reliable and structured data, foundational to both AI systems and investigative work. As discussed, investigative journalism often depends on data that are hidden, incomplete, messy or intentionally withheld, which requires journalists to “request, negotiate, scrape, or purchase” datasets before their analysis can even begin.²⁹ Even when data is available, it can be messy (e.g., spread across multiple sources, poorly formatted, inconsistent over time, lacking metadata), further complicating automation.³⁰ These constraints reflect a broader “infrastructural gap,” in which investigative capacity is systematically limited by factors such as proprietary roadblocks and the absence of accessible public databases. Because AI systems rely heavily on consistently structured data, the risk of inaccurate or misleading outputs increases, making full automation not only unreliable but also impossible.³¹

Professional Identity and Dependence on External Actors

AI systems introduce concerns about the erosion of journalistic autonomy and expertise. Journalists across all subfields emphasize that their “*irreplaceable human qualities*” (e.g., judgment, fieldwork, interviewing, source-building) remain central to their work despite increasing automation of other

areas of work.³² At the same time, reliance on external technology companies, nongovernmental organizations and universities for AI tools and infrastructure can create newsroom dependencies and raise concerns about influence, information security, copyright, privacy and data protection.³³

A comparable dynamic already exists in non-AI investigative practices. For example, during the U.S.-Israel war with Iran, news organizations relied heavily on commercial satellite imagery to verify strikes and assess damage in Iran, but access was restricted when Planet Labs limited imagery distribution at the request of the U.S. government, citing operational security concerns. This shift to a managed, case-by-case release system sharply reduced journalists’ ability to independently document events and diminished the transparency previously enabled by open commercial imagery, illustrating that newsroom capabilities depend on the policies and geopolitical constraints of external data providers.

This increased dependence on external actors is part of a broader structural transformation in journalism. The “platformization” of news — understood as the rise of platforms as the dominant infrastructural and economic model of the social web³⁴ — applies to the entire news ecosystem. As news organizations increasingly depend on third-party platforms for production tools, distribution channels and revenue streams, their autonomy is affected by systems they do not control.³⁵ More broadly, studies stress that “*technology in journalism is not neutral*,”³⁶ as its effects depend on institutional contexts and power relations, meaning AI tools can reshape — not just support — journalistic practice.

Technical and Epistemic Limitations

Despite the promise of natural-language tool interactions,³⁷ implementing AI tools in investigative journalism requires substantial technical expertise, along with collaboration across disciplines and roles. AI tools are sometimes complex and difficult to integrate into newsroom workflows, especially when they require customization or advanced technical knowledge. In an interview study with 25 national, regional and local Dutch news media organizations,³⁸ the authors identified a significant “*language barrier*” between investigative journalists and IT professionals that serves as a major impediment to AI adoption. This barrier manifests in several ways. First, there is a lack of shared vocabulary. Some journalists find it difficult to articulate their specific editorial needs to technical experts because they lack a foundational vocabulary for AI techniques and applications. Second, interviewees noted that IT professionals “look through different glasses” and often assume a level of technical knowledge that journalists do not have, making communication difficult. Third, there is a disconnect between the “work sprints” of IT departments, which may plan activities months in advance, and the fast-paced or unpredictable deadlines of investigative news desks. Fourth, IT professionals are often based outside the newsroom, which further limits the opportunity for the spontaneous, interdisciplinary collaboration needed to co-create AI tools.

To overcome this challenge, research offers a few recommendations. One is to involve “*boundary spanners*”³⁹ — individuals who operate at the intersection of journalism and technology and can translate between the two groups. These boundary spanners need not be additional staff; existing personnel can take on these roles and commit to approaching

technology from this perspective by bringing together colleagues with distinct expertise. Another recommendation is to take a collaborative approach requiring active participation from all stakeholders, including, at times, those outside journalism.⁴⁰

In addition to communication barriers between journalists and IT professionals, AI adoption is shaped by other organizational and professional dynamics. Sometimes AI uptake is driven by enthusiasm for new technologies without strategy, which some journalists and scholars call “*Shiny Things Syndrome*.”⁴¹ As shown in an interview-based study in the Dutch news media context,⁴² AI discussions sometimes remain at the managerial level and do not translate to everyday newsroom workflows. As a result, many investigative journalists position themselves as “*reluctant adopters*,” particularly when faced with technical language barriers and the need to collaborate with specialized teams such as data scientists or IT professionals. These professional dynamics are not specific to AI adoption; across a range of methods, earlier research on technology adoption in journalism also shows widespread curiosity about new technologies alongside resistance to top-down technology mandates and concerns about dependency on third-party companies.⁴³

These challenges are closely tied to the broader issue of insufficient AI and data literacy in journalism. Journalists struggle with limited access to reliable learning resources, difficulty identifying appropriate expertise, general reluctance and the challenge of keeping pace with rapidly evolving technologies.⁴⁴ In response, scholars have proposed frameworks such as the Accuracy-Fairness-Transparency (AFT) framework, which emerged from a synthesis of prior research

on AI systems in journalism. AFT applies core journalistic ethics (accuracy, fairness and transparency) to ensure high-quality, trustworthy data in AI-driven journalism. AFT can improve data practices and serve as a pathway for building AI literacy and embedding journalistic values into technological workflows.⁴⁵

Additionally, the opacity of AI tools limits journalists' understanding of their outputs and how and why they are processed.⁴⁶ This, in turn, complicates accountability and the epistemic traceability of evidence, especially in the context of generative AI.⁴⁷ Scholars distinguish between explainability — the ability to explain what a model does and what it produces — and interpretability, which refers to understanding the specific mechanics of how the model reached that result.⁴⁸ Because of these concerns, many investigative teams purposefully opt for “decision tree” or “random forest” algorithms over more powerful neural network or deep learning approaches because they prefer tools that produce human-readable partitions of data that allow for better scrutiny.⁴⁹ In the specific context of generative AI, the traceability of evidence is further undermined by hallucinations, where models predict plausible but factually incorrect data points. This lack of reliability has led some newsrooms to abandon high-performing AI anomaly detection tools.⁵⁰

Finally, many tools remain project-specific, limiting their reuse and reducing long-term efficiency gains. These constraints mean that AI adoption is not simply a matter of availability but of sustained organizational capacity, training and collaboration.⁵¹

Linguistic Barriers

AI systems in investigative journalism create another type of language barrier, particularly in non-English and Global South contexts, because the underlying technologies are primarily developed, trained and optimized for English. This results in clear technical performance gaps: while English-based models are highly-developed, even widely used languages like French remain comparatively underdeveloped and “clumsy,”⁵² and AI tools often fail to function effectively in many non-Anglophone languages.⁵³ As a result, journalists (including non-investigative journalists) in these contexts face limitations in using AI tools for core tasks like data mining, transcription and automated analysis.⁵⁴ These constraints are compounded by a broader “Eurocentric” information lens, where AI systems — shaped by Western design and data — struggle to interpret local cultural nuances and are viewed by some as tools that extract information for Western purposes rather than supporting local knowledge production. Because they are not trained on localized linguistic and cultural data, such systems remain fundamentally exclusionary in their ability to represent non-Western perspectives.⁵⁵

These dynamics extend into professional and educational domains, where English-centric AI systems reinforce existing hierarchies and forms of exclusion. An interview and focus group study of 34 investigative journalists across five East African countries shows that in some newsrooms, competence is tied to English proficiency rather than investigative skill, leading to the loss of talented journalists who are unable to communicate their work effectively in English. AI tools can intensify this divide by amplifying the capabilities of those already fluent in dominant languages and technologies, while

further marginalizing those without these skills.⁵⁶ A related mixed-methods study of journalism education across four South African universities found similar resistance to AI tools that lack support for Indigenous language contexts, prompting calls to “*decolonize*” language models by training them on local languages and contexts. Without such efforts, investigative AI tools risk continuing to privilege English-speaking environments and limiting locally grounded journalism.⁵⁷ For example, Latam-GPT, developed by Chile’s National Center for Artificial Intelligence (CENIA), aims to build an open-source AI model trained on Latin American languages and contexts (see this working group’s November 2025 briefing on translation and transcription for further discussion of linguistic strengths and limitations).

Economic Factors

From an economic perspective, investigative journalism faces some disadvantages that limit the broader use of AI. For one, investigative reporting relies heavily on human effort. Unlike advertising or other automated communication fields, it is much harder to use technology to speed up the deep, manual work required for investigations. Furthermore, while the social importance of investigative journalism remains constant, its production costs increase relative to other sectors, often branding it as “costly” by comparison.⁵⁸ This creates a condition

in which AI tools accelerate and facilitate some tasks, such as data processing, but fail to reduce the overall labor. Moreover, the limited productivity gains from AI systems in investigative journalism, relative to the significant gains in sectors like advertising and public relations, create a cost disadvantage.⁵⁹ This “cost disease” makes it economically difficult for newsrooms to fund competitive wages for highly skilled human labor, increasing the risk of talent drain as journalists move to better-paying communication sectors with overlapping skill sets.⁶⁰

As discussed earlier, these constraints are compounded by the project-specific nature of AI systems in investigative journalism.⁶¹ As a result, AI adoption tends to be viable only for a narrow subset of investigations where data is central, automation is technically feasible and simpler computational methods or traditional reporting cannot achieve comparable results. These financial barriers are even more pronounced for smaller and local newsrooms, which are less able to absorb the high costs of AI development or even to customize off-the-shelf tools — sometimes paired with infrastructural issues, such as unreliable internet connectivity⁶² — and therefore face greater barriers to adoption, reinforcing existing inequalities in the distribution of AI capabilities across the industry.⁶³

Global Perspectives

Working group member Gregory Gondwe writes:

Research from Africa and other parts of the Global South suggests that the adoption of AI in investigative journalism is influenced by the practical realities of reporting within resource-constrained environments, uneven digital infrastructures, and varying levels of press freedom. A 2025 Thomson Reuters Foundation survey of journalists across more than 70 countries found widespread use of AI tools but limited institutional governance, with many newsrooms lacking formal policies or training structures.⁶⁴ Across these contexts, journalists adapt AI systems to address localized challenges associated with incomplete datasets, language barriers, weak information infrastructures, and political constraints.

One example comes from East Africa, where investigative journalists use AI-assisted tools to support both reporting and personal security. In a study of journalists from the Democratic Republic of Congo, Ethiopia, Rwanda, Tanzania, and Uganda, participants described using AI-supported anonymization, encrypted communications, automated monitoring tools, and other forms of digital subterfuge while investigating corruption, human rights abuses, and state misconduct.⁶⁵ Journalists

viewed these technologies as tools that help sustain investigative work under conditions of surveillance and political pressure. The study also found that technological adaptation occurs on both sides of the information environment, with journalists deploying protective tools while governments expand their own AI-enabled surveillance capabilities.

“A second example highlights the role of computational tools in addressing weaknesses within public information infrastructures. Journalists investigating procurement corruption across several African countries have combined fuzzy matching, clustering techniques, and extensive manual verification to connect fragmented government records with leaked documents and supplier databases. These investigations demonstrate how successful AI-supported reporting depends on the interaction between computational analysis and human judgment. Patterns discovered through automated techniques still require contextual interpretation, source development, and rigorous verification before they become publishable evidence.⁶⁶

Where More Research would be Helpful

- **Cross-industry comparison:** More research is needed to compare how AI tools are adopted across journalistic subfields to better understand their unique applications and limitations in investigative contexts.
- **Generalizable global research with investigative reporters:** Much of the literature analyzed in this briefing used practice-based research techniques, bringing together investigative journalists, academics, data journalists and data scientists, to achieve a deeper understanding of investigative journalism as it is practiced today. While these provide deep insights, there is a clear opportunity for future studies to adopt broader quantitative approaches, such as large-scale surveys, to capture a more comprehensive global outlook on how investigative journalists use and perceive AI systems in their professional practice.
- **Global South and context-sensitive AI:** Most existing research on AI applications for investigative journalism focuses on the United States and Europe. Future work should focus more on developing and evaluating AI tools that are tailored to non-Western, non-English and culturally specific contexts in investigative journalism.
- **Emerging and evolving nature of AI:** Given the rapid evolution of AI technologies and the limited number of existing studies, ongoing empirical research is needed to track their long-term impacts on investigative journalism.
- **Local newsrooms:** More research should examine how under-resourced newsrooms adopt, adapt to or are excluded from AI-driven investigative practices.
- **Sustainable funding models:** Future studies should explore viable and independent funding models that support AI innovation in investigative journalism without increasing reliance on external actors or compromising editorial autonomy.

The following cases highlight that, across all regions, the consensus remains that AI tools serve as a force multiplier rather than a replacement for human labor. They enable journalists to examine data creatively and pursue stories that would otherwise be buried under the sheer scale of modern information.

Data sifting and lead generation

- **AP Local News AI:** In collaboration with AppliedXL, the Associated Press developed AP Local News AI. This tool monitors federal agency data to identify how national regulations impact specific local communities, effectively “anticipating” stories before they break.
- **Datashare:** It is a free, open-source desktop software developed by the International Consortium of Investigative Journalists (ICIJ) that enables users to index, search and analyze large volumes of documents (e.g., PDFs, emails, images) locally or on a server.
- **Nubia AI:** an AI-powered data journalism platform designed to transform complex, messy datasets, such as PDFs, spreadsheets and satellite imagery, into structured data and publishable, newsroom-grade stories. Developed by Dataphyte Nigeria in 2022, it is designed to help journalists, researchers and civic organizations produce data-driven stories quickly and accurately.

Fact-checking and verification

- **StatCheck:** Developed by Inria (National Institute for Research in Digital Science and Technology) and RadioFrance, StatCheck automates the verification of statistical claims by cross-referencing articles against the INSEE and Eurostat databases.
- **Dubawa:** To combat regional misinformation in Western Africa, the Dubawa project launched an AI-powered chatbot and audio platform specifically designed to debunk viral falsehoods.

Content production and contextualization

- **Odin:** The Colombian outlet Cuestión Pública developed Odin, an AI system that connects past investigations to breaking news. Using retrieval-augmented generation (RAG), it searches proprietary databases and generates draft content grounded in verified reporting. The tool combines BERT for data indexing with GPT-4 for drafting, reducing production time from hours to minutes while maintaining editorial control.

Resources and Practical Guidelines

While academic research on AI in investigative journalism remains relatively limited, a growing body of practitioner-oriented resources offers guidance on how these tools are being used in newsrooms. These sources are particularly valuable because they translate technical possibilities into replicable workflows, editorial strategies and ethical considerations. Key examples include:

- The *Global Investigative Journalism Network's* guide, *"Tips for Using AI as a Reporting Tool to Uncover Wrongdoing,"* explains how AI can help structure large datasets, generate leads and build searchable systems, while emphasizing that tools should align with clearly defined reporting needs.
- GIJN's *Leveraging AI and Technology to Investigate Power* highlights the importance of practical toolkits, low-code solutions, collaborative infrastructures and human-led investigative processes.
- GIJN's *7 Best Practices for More Impactful Investigative Reporting*, which stresses that core investigative practices — including patience, source development and narrative construction — remain central despite technological change.
- Analysis from the *Center for International Media Assistance*, showing how document-query systems and AI-assisted databases can reduce the time needed to analyze large document collections while still requiring verification and editorial oversight.

Scope of included works

The reviewed studies span an international landscape, covering more than 20 countries and regions across Europe, North America, Africa and Asia and several cross-regional contexts such as Eastern Africa, Southeast Asia and Western Europe. Methodologically, the literature reflects a strong dominance of qualitative and mixed-methods approaches. Several studies also incorporated literature reviews, prototype development, usability testing and data-driven analytical techniques such as topic modeling, demonstrating substantial methodological diversity within the field.

Becker, K.B. (2023). New game, new rules: An investigation into editorial guidelines for dealing with artificial intelligence in the newsroom. *Journalism Research*, 6(2). <https://journalistik.online/en/paper-en/new-game-new-rules/>

Boukhssas, M. K. (February 15, 2026). Methodological approaches to investigative journalism and their impact on enhancing its quality. Al Jazeera Centre for Studies. <https://studies.aljazeera.net/en/analyses/methodological-approaches-investigative-journalism-and-their-impact-enhancing-its-quality>

Bradshaw, P. (2025). Investigative journalism and AI. In N. Macfarlane, B. Longo-Flint, & J. Price (Eds.), *Insights on Investigative Journalism* (1st ed., pp. 155–170). Routledge. <https://doi.org/10.4324/9781003478157-14>

Broussard, M. (2015). Artificial intelligence for investigative reporting: Using an expert system to enhance journalists' ability to discover original public affairs stories. *Digital Journalism*, 3(6), 814–831. <https://doi.org/10.1080/21670811.2014.985497>

Broussard, M., Diakopoulos, N., Guzman, A. L., Abebe, R., Dupagne, M., & Chuan, C.-H. (2019). Artificial intelligence and journalism. *Journalism & Mass Communication Quarterly*, 96(3), 673–695. <http://doi.org/10.1177/1077699019859901>

Čakš, P., Kožuh, I., & Žilič Fišer, S. (2025). Journalistic roles and community expectations: External and internal challenges for local Investigative journalism. *Journalism Practice*, 1–19. <https://doi.org/10.1080/17512786.2025.2582166>

Chijioke, A. (2021, November 8). 7 best practices for more impactful investigative reporting. *Global Investigative Journalism Network*. <https://gijn.org/stories/7-best-practices-for-more-impactful-investigative-reporting/>

Chimoio, M. (2026, February 17). Following the digital trail: How East African journalists investigate corruption through procurement data. *Global Investigative Journalism Network*. <https://gijn.org/stories/investigating-corruption-procurement-data-africa/>

CNTI. (2025, November 20). AI transcription and translation in journalism. *Center for News, Technology, and Innovation*. <https://cnti.org/reports/ai-transcription-and-translation-in-journalism/>

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating data journalism, computational journalism, and computer-assisted reporting. *Digital Journalism*, 3(3), 331–348. <https://doi.org/10.1080/21670811.2014.976400>

Dalgali, A, & Crowston, K. (2020). Algorithmic journalism and its impacts on work. *C+J 2020 Symposium*. <https://par.nsf.gov/biblio/10301595>.

da Silva, J. R. (2023). Connect-the-dots: Artificial intelligence and automation in investigative journalism. PhD thesis University of Porto. <https://repositorio-aberto.up.pt/bitstream/10216/153970/2/648196.pdf>

De Cooker, J., Arets, D., Brugman, M., Wernaart, B., Van Vliet, M., & Van Leeuwen, T. (2025). Lost in Translation? How Structural, Individual, and Professional Factors Hinder AI Adoption in Investigative Journalism. *Journalism Studies*, 1–17. <https://doi.org/10.1080/1461670X.2025.2595290>

de-Lima-Santos, M. F., & Ceron, W. (2022). Artificial intelligence in news media: Current perceptions and future outlook. *Journalism and Media*, 3(1), Article 1. <https://doi.org/10.3390/journalmedia3010002>

de-Lima-Santos, M.F., Yeung, W.N., & Dodds, T. (2024). Guiding the way: A comprehensive examination of AI guidelines in global media. *AI & SOCIETY*, 40(4), 2585–2603. <https://link.springer.com/article/10.1007/s00146-024-01973-5>

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Diakopoulos, N. (2024). Data journalism: The emergence of computational journalism at Georgia Tech, 2006–2008. In J. Pavlik (Ed.) *Milestones in Digital Journalism* (1st ed., pp. 1–16). Routledge.

Dierickx, L., Opdahl, A. L., Khan, S. A., Lindén, C.-G., & Guerrero Rojas, D. C. (2024). A data-centric approach for ethical and trustworthy AI in journalism. *Ethics and Information Technology*, 26(4), 64. <https://doi.org/10.1007/s10676-024-09801-6>

Fridman, M., Krøvel, R., & Palumbo, F. (2025). How (not to) run an AI project in investigative journalism. *Journalism Practice*, 19(6), 1362–1379. <https://doi.org/10.1080/17512786.2023.2253797>

Ganguly, M. (2022). *The future of investigative journalism in the age of automation, open-source intelligence (OSINT) and artificial intelligence (AI)*. PhD thesis, University of Westminster, Westminster School of Media and Communication. <https://doi.org/10.34737/vx128>

Gondwe, G. (2025). Investigative journalism and AI-driven subterfuge in countries with limited press freedom in East Africa. *Journalism Practice*, 0(0), 1–20. <https://doi.org/10.1080/17512786.2025.2541254>

Grimme, M. (2021). Factors Influencing the Rejection of Automated Journalism: A Systematic Literature Review. *Nordic Journal of Media Management*, 2(1), 3–21. <https://doi.org/10.5278/njmm.2597-0445.6826>

Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social media & Society*, 1(2).

Hofeditz, L., Jung, A.-K., Mirbabaie, M., & Stieglitz, S. (2025). Ethical guidelines for the application of generative AI in German journalism. *Digital Society*, 4(1), 4. <https://doi.org/10.1007/s44206-024-00151-w>

Hollanek, T., Peters, D., Drage, E., & Hernandez, R. (2025). AI, journalism, and critical AI literacy: Exploring journalists' perspectives on AI and responsible reporting. *AI & SOCIETY*, 40(8), 6393–6405. <https://doi.org/10.1007/s00146-025-02407-6>

Kasica, S., Berret, C., & Munzner, T. (2023). Dirty data in the newsroom: Comparing data preparation in journalism and data Science. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, 1–18. <https://doi.org/10.1145/3544548.3581271>

Katches, M. (2014, November 30). 33 tips to keep daunting investigative reporting projects on track. *Reveal*. <https://revealnews.org/article/33-tips-to-keep-daunting-investigative-reporting-projects-on-track/>

Küng, L. (2013). Innovation, technology and organisational change. In T. Storsul & A. H. Krumsvik (Eds.), *Media innovations: A multidisciplinary study of change* (pp. 9-12). Nordicom.

Lischka, J. A., Mollen, A., Kristensen, L. M., Kunert, J., Braun, A. F., Koitie, P. K., Rolandsson, T., Schaetz, N., & Kammer, A. (2025). What is data? A conceptual and empirical inquiry of the facets of data in news organizations. SocArXiv. https://doi.org/10.31235/osf.io/2t6br_v2

Manolescu, M., & Vaudano, I. (2024, November 6). AI and the media: A (r)evolution in investigative reporting? *Polytechnique insights*. <https://www.polytechnique-insights.com/en/columns/science/ai-and-the-media-a-revolution-in-investigative-reporting/>

Mari, W. (2024). The Pre-History of News-Industry Discourse Around Artificial Intelligence. *Emerging Media*, 2(3), 499–522. <https://doi.org/10.1177/27523543241279577>

Mohamedy, H., & Wandwi, G. (2025). How digital tools are transforming investigative journalism in practice. *Journalism Practice*, 0(0), 1–27. <https://doi.org/10.1080/17512786.2025.2605212>

Møller, L. A., van Dalen, A., & Skovsgaard, M. (2025). A little of that human touch: How regular journalists redefine their expertise in the face of artificial intelligence. *Journalism Studies*, 26(1), 84–100. <https://doi.org/10.1080/1461670X.2024.2412212>

Ncube, L., Mofokeng, R. W., Chibuwe, A., Munoriyarwa, A., & Murangi, A. K (2025). 'Mind the gap': Artificial intelligence and journalism training in Southern African journalism schools. *Media Practice and Education*, 0(0), 1–17. <https://doi.org/10.1080/25741136.2025.2464483>

Örnebring, H. (2010). Technology and journalism-as-labour: Historical perspectives. *Journalism*, 11(1), 57–74. <https://doi.org/10.1177/1464884909350644>

Poell, T., Nieborg, D. B., & Duffy, B. E. (2023). Spaces of negotiation: Analyzing platform power in the news industry. *Digital Journalism*, 11(8), 1391–1409. <https://doi.org/10.1080/21670811.2022.2103011>

Posetti, J. (2018). *Time to step away from the 'bright, shiny things'? Towards a sustainable model of journalism innovation in an era of perpetual change*. Reuters Institute.

Radcliffe, D. (2025). Journalism in the AI era: Opportunities and challenges in the Global South and emerging economies. *Thomson Reuters Foundation*. <https://www.trust.org/resource/ai-revolution-journalists-global-south/>

Simon, F. M. (2022). Uneasy bedfellows: AI in the news, platform companies and the issue of journalistic autonomy. *Digital Journalism*, 10(10), 1832–1854. <https://doi.org/10.1080/21670811.2022.2063150>

Stray, J. (2019). Making artificial intelligence work for investigative journalism. *Digital Journalism*, 7(8), 1076–1097. <https://doi.org/10.1080/21670811.2019.1630289>

Thurman, N., Lewis, S. C., & Kunert, J. (2019). Algorithms, automation, and news. *Digital Journalism*, 7(8), 980992. <https://doi.org/10.1080/21670811.2019.1685395>

Thurman, N., Thäsler-Kordonouri, S., & Fletcher, R. (2025). AI adoption by UK journalists and their newsrooms: surveying applications, approaches, and attitudes. <https://doi.org/10.60625/risj-ea11-q402>

Umeora, C. C. (2025). The impact of AI on investigative journalism: Opportunities and challenges for Nigerian media professionals. *International Journal of Innovative Information Systems and Technology Research*, 13(1), 108-117.

Veerbeek, J. (2025). Fighting fire with fire: Journalistic investigations of artificial intelligence using artificial intelligence techniques. *Journalism Practice*, 1-19. <https://doi.org/10.1080/17512786.2025.2479499>

Wellbrock, C. M. (2024). Generative AI and investigative journalism—An application of Baumol's Cost Disease across mass communication sectors. *Emerging Media*, 2(4), 663-672. <https://doi.org/10.1177/27523543241299631>

1. See Mari, 2024
2. Broussard, 2015; Chijioke, 2021; Čakš et al., 2025; da Silva, 2023; Ganguly, 2022; Katches, 2014
3. Boukhssas, 2026; da Silva, 2023
4. De Cooker et al., 2025
5. da Silva, 2023
6. Wellbrock, 2024
7. Fridman et al., 2025
8. Wellbrock, 2024
9. da Silva, 2023
10. Coddington, 2015
11. Lischka et al., 2025
12. Diakopoulos, 2024
13. Diakopoulos, 2024
14. Coddington, 2015
15. da Silva, 2023
16. Diakopoulos, 2024
17. Bradshaw, 2025; Cifliku & Heuer, 2025
18. Bradshaw, 2025; Stray, 2019
19. Cifliku & Heuer, 2025
20. Kasica et al., 2023
21. Kasica et al., 2023; Mohamedy & Wandwi, 2025
22. Bradshaw, 2025
23. Mohamedy & Wandwi, 2025
24. de-Lima-Santos & Ceron, 2022
25. Veerbeek, 2025
26. Fridman et al., 2023
27. Gondwe, 2025
28. Bradshaw, 2025
29. Wellbrock, 2024
30. Kasica et al., 2023; Cifliku & Heuer, 2025
31. Wellbrock, 2024
32. Møller et al., 2025
33. Bradshaw, 2025
34. Helmond, 2015
35. Poell et al., 2022; Simon, 2022
36. Mohamedy & Wandwi, 2025
37. Dalgali & Crowston, 2020
38. De Cooker et al., 2025
39. De Cooker et al., 2025
40. Fridman et al., 2025; Mohamedy & Wandwi, 2025
41. Posetti 2018; see also Broussard et al. 2019; Kueng 2017; Thurman et al., 2019
42. De Cooker et al., 2025
43. For example, see Örnebring, 2010; Grimme, 2021. See also Møller et al. 2025 for AI specifically.
44. De Cooker et al., 2025; Hollanek et al., 2025; Thurman et al., 2025
45. Dierickx et al., 2024
46. Bradshaw, 2025; Mohamedy & Wandwi, 2025
47. Hofeditz et al., 2025
48. Becker, 2023; de-Lima-Santos et al., 2024
49. Bradshaw, 2025
50. Bradshaw, 2025
51. Stray, 2019
52. Manolescu & Vaudano, 2024
53. CNTI, 2025
54. Ncube et al., 2025
55. Ncube et al., 2025
56. Gondwe, 2025
57. Ncube et al., 2025
58. Wellbrock, 2024
59. Wellbrock, 2024
60. Stray, 2019
61. Stray, 2019; Verbeek, 2025
62. Umeora, 2025
63. de-Lima-Santos & Ceron, 2022
64. Radcliffe, 2025
65. Gondwe, 2025
66. Chimoio, 2026

Akintunde Babatunde	Executive Director, Centre for Journalism Innovation and Development
Claudia Báez	Associate Consultant, Fathm
Jay Barchas-Lichtenstein	Senior Research Manager, Center for News, Technology & Innovation
Madhav Chinnappa	Independent Media Consultant
Utsav Gandhi	PhD Student, University of Illinois Chicago
Gregory Gondwe	Assistant Professor of Journalism & Emerging Media Technologies, California State University, San Bernardino
K.V. Kurmanath	Senior Journalist and Academic
Amy Mitchell	Executive Director, Center for News, Technology & Innovation
Chris Moran	Head of Editorial Innovation, Guardian News & Media
Sophie Morosoli	Postdoctoral Researcher, AI, Media & Democracy Lab, University of Amsterdam
Afrooz Mosallaei	Research Associate, Center for News, Technology & Innovation
Gary Mundy	Director of Research, Policy and Impact, Thomson Foundation
Oluwapelumi Oginni	Project Manager, AI Initiatives, Centre for Journalism Innovation and Development
Joshua Olufemi	Executive Director, Dataphyte Foundation
Oluseyi Olufemi	Nigeria Country Director, Dataphyte
Esteban Ponce de León	Resident Fellow, Digital Forensic Research Lab (DFRLab) at the Atlantic Council
Amy Ross Arguedas	Research Fellow, Reuters Institute for the Study of Journalism
Zara Schroeder	Researcher, Research ICT Africa
Felix M. Simon	Research Fellow in AI and News, Reuters Institute for the Study of Journalism Research Associate, Oxford Internet Institute, University of Oxford
Sabina Tomkins	Assistant Professor of Information, School of Information and Faculty Associate, Center for Political Studies, Institute for Social Research, University of Michigan
Jaemark Tordecilla	Independent Media Advisor, Philippines