Universal Library of Innovative Research and Studies Book

ISSN: 3065-0003

Open Access | PP: 15-25

DOI: https://doi.org/10.70315/uloap.ulirs.2022.003

A Hybrid Fact-Checking Model: A Methodology for Integrating AI-Based Tools into the Editorial Workflow of a News Portal

Sprinchinat Kateryna

Journalist at the Regional Television/Radio Company Chernivtsi Promin, Chernivtsi, Ukraine.

Abstract

Amid the exponential growth in the volume and velocity of misinformation, countering contemporary information threats effectively is outside the structural capacity of customary editorial fact-checking models. Manual verification remains the gold standard for quality, yet it lacks the requisite scalability for real-time operation. This lack creates critical vulnerabilities for both media and society. This methodology presents the author's Hybrid Fact-Checking Model, a thorough sociotechnical framework that integrates artificial intelligence (AI)- based tools into newsroom workflows. It aims for faster verification, broader coverage, and greater accuracy while preserving journalistic ethics and assuring complete human supervision. The methodological novelty is found in a formalizing of three key components: (1) human, machine interaction protocols that operate for a regulation of collaboration between journalists and AI assistants; (2) a task allocation matrix that clearly demarcates zones of responsibility between automated systems that monitor and collect primary data and humans who analyze context, appraise ethics, and adjudicate finally; and (3) a risk management and effectiveness evaluation system that manages risk, evaluates effectiveness, and includes practical methods for minimization of AI hallucinations and algorithmic bias, as well as a set of key performance indicators (KPIs) for a hybrid newsroom. The methodology targets editors and media leaders, serving as a deployment-ready guide to technological modernization that aims to strengthen the competitiveness and authority of news outlets in today's information environment.

Keywords: *Hybrid Fact-Checking, AI-Assisted Verification, Misinformation Detection, Human-in-the-Loop, Algorithmic Bias Management, Newsroom* KPI.

INTRODUCTION

The contemporary information ecosystem is now experiencing a permanent crisis due to the unprecedented rise in the volume and velocity of misinformation. The threat's global recognition is evidenced in bibliometric analyses, which indicate growth in scholarly publications on this topic, primarily after 2019 (Wang et al., 2022). Digital platforms and social networks, designed to democratize information, have become powerful vectors distorting it. Since it is optimized for maximal engagement, the architecture of these platforms fosters echo chambers and filter bubbles. This acceleration throughout the viral spread of false narratives, frequently outpacing any attempts at refutation, is a result (Muhammed & Mathew, 2022).

Journalism, as well as fact-checking, faces a fundamental challenge under such conditions. Its main tool is part of the challenge. Professional fact-checking is a cornerstone of quality journalism, grounded in expert evaluation, the pursuit of primary sources, and meticulous manual analysis. Yet the process is inherently slow and resource-intensive (Allen et al., 2021). As researchers note, while new content is continuously published and disseminated online, it becomes increasingly complex for journalists to verify it all promptly, even with the best tools available (Caled & Silva, 2022).

The problem is not merely resource scarcity but a structural mismatch of speeds. The production process of quality journalism, with its verification time, lags hopelessly behind the dissemination of misinformation, which operates in real-time. This kind of asymmetry enables such false information to reach a global audience. Before a correction can be published, it harms reputations, public opinion, and democratic institutions. The inability, as a result of customary fact-checking models, to scale against the speed and volume of misinformation is a systemic vulnerability that requires a radical rethinking of verification approaches, rather than simply an operational difficulty.

Awareness about the structural limits of manual fact-checking means inexorably transforming it technologically. Advanced technologies especially artificial intelligence (AI) and Natural Language Processing (NLP)'s integration is a deliberate imperative not just an opportunity for news media survival and relevance. AI technologies are able to analyse enormous volumes of textual as well as multimedia data within real-time, whilst they recognise trends in misinformation, they detect misleading claims that exist too, also they optimise verification workflows so fact-checkers cope with the huge volume of information circulating online.

This alteration indicates a move away from complete manual

checks toward a hybrid human-involved model structure. Technologies augment the journalist's capabilities instead of replacing them inside this model framework. In such a model, AI acts as a powerful assistant that continuously monitors the information space also automatically detects potentially false claims. Also, it gathers evidence and structures it mostly for routine yet labor-intensive tasks. It frees time and cognitive resources up for tasks requiring critical thinking, deep analysis, ethical evaluation, and judgment.

Accordingly, fact-checking transforms technologically without renouncing journalistic standards for automation. Instead, transformation reinforces all of those standards through some clever tools. It transitions from a reactive model where fact-checkers belatedly respond to falsehoods already spread to a proactive one where automated systems detect then reduce early-stage threats. The successful execution of this transition is pivotal for news organizations not only to withstand the information war but also to lead it, thereby reinforcing their authority as reliable sources of verified information.

The purpose of this methodology is to present a reproducible, scientifically grounded model for integrating AI-based tools into a news portal's editorial process, increasing the speed, coverage, and accuracy of verification while preserving fundamental journalistic standards and complete human control.

The scientific novelty lies in the formalization along with development of the author's Hybrid Fact-Checking Model. This plan is a logical integrated structure, not simply some practical suggestions. It addresses a key dilemma within contemporary journalism: the need for scalability conflicts with the pursuit of epistemological rigor. Customary fact-checking is epistemologically strong yet not scalable. Automation which is full is scalable yet brittle epistemologically and prone to biases and errors and inability for analysis of context. This model resolves the contradiction that exists with a hybrid epistemology for journalism, combining human judgment plus machine speed with accountability proposed.

Three interlinked components define the model's novelty:

1. Human–machine interaction protocols. Unlike ad hoc practices of AI use, the methodology proposes standardized, reproducible operational protocols that govern each stage of collaboration since they monitor automatically, escalate suspicious materials, verify via AI assistance, and issue verdicts.
2. Task allocation matrix. The methodology introduces an authorial instrument. This tool is a grid that plainly demarcates duty accountability areas. Tasks that are fully automated, and tasks requiring control of mandatory human-in-the-loop type, as well as tasks that remain the journalist's exclusive prerogative, like ethical judgments, sarcasm detection, and final reliability assessment, are specified.
3. Risk management system. The framework includes a proactive system to identify, monitor, and mitigate inherent AI risks such as hallucination and systemic algorithmic bias. This system draws on leading practices and standards in the responsible use of AI.

Thus, the methodology offers not merely a technological solution but a comprehensive managerial model that ensures a strategic, operational, and ethical transition of a newsroom to a higher level of technological maturity.

CHAPTER 1. AUDIT AND DESIGN OF THE HYBRID NEWSROOM: THE PREPARATORY STAGE

Diagnosing Existing Processes: A Methodology for Analyzing the Current Editorial Fact-Checking Cycle

Before introducing any AI technologies, it is vital to conduct a detailed review of current editorial workflows. The goal of this stage is to systematically identify operational bottlenecks, routine repetitive tasks, and areas in which human resources are used inefficiently, also to describe the current workflow. Such an approach is one that isolates concrete tasks that are suitable for automation. This ensures that the technology addresses real problems, and that the technology does not create any new ones.

The methodology for analyzing the current editorial fact-checking cycle includes the following steps:

1. Process mapping. The entire life cycle of verifying a single claim must be visualized in detail, from detection (e.g., via social media monitoring, reader submissions, politician statements) to publication and dissemination of the debunk. Mapping should encompass all stages, including identifying the claim, searching for and collecting evidence, verifying sources, consulting experts, drafting, editing, and publication.
2. Task and actor identification. For each mapped stage, specify the concrete tasks performed and who is responsible (e.g., junior journalist, senior fact-checker, editor, data specialist).
3. Time-and-motion analysis. Estimate the time and effort spent on each task. Pay special attention to recurring, time-consuming tasks such as manual feed monitoring, keyword searching across large corpora, or transcribing audio/video.
4. Bottleneck detection and automation candidates. Based on the data, identify key problems, such as process delays (e.g., long waits for expert feedback), tasks with high human error risk (e.g., missing salient items during manual monitoring), or purely mechanical tasks that do not require creative or critical thinking.

To systematize this analysis, a diagnostic workflow, as shown in Figure 1, is proposed. It enables editors to evaluate each step in the fact-checking cycle and make informed decisions on the advisability of automation.

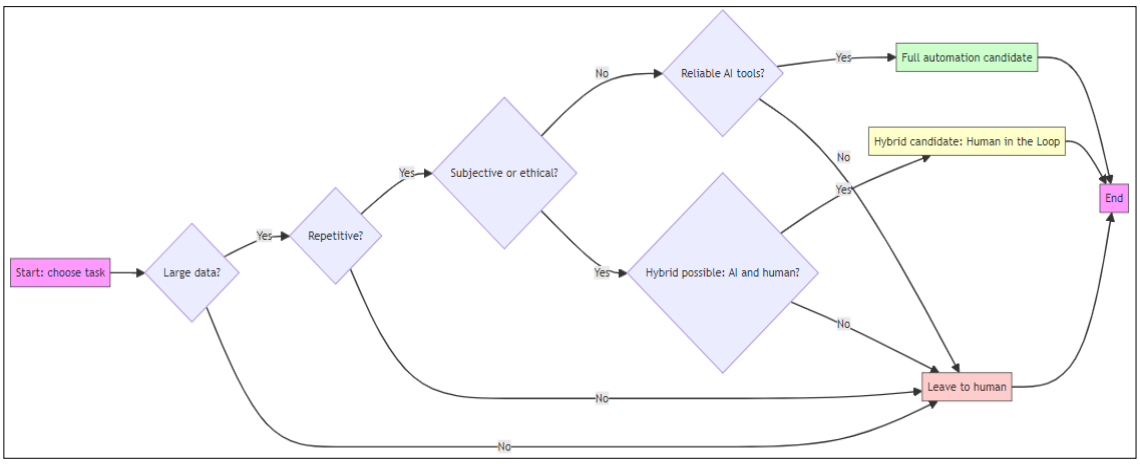

Figure 1. Diagnostic workflow for identifying tasks to be automated in the editorial fact-checking cycle

Conducting such an audit enables a shift from abstract notions of using AI to a concrete, data-driven modernization plan. It ensures that AI investments are directed toward the newsroom's most pressing pain points, thereby increasing overall productivity and allowing journalists to focus on their core mission: conducting in-depth, high-quality analysis of information.

Criteria for Selecting AI Tools: Developing a Software Evaluation Matrix

Selecting the right AI tool is a critical decision that determines the success of the entire hybrid model. The market offers numerous solutions, from social media monitoring systems to image verification platforms and text analysis suites. However, evaluation must transcend marketing claims and instead rest on rigorous, multidimensional criteria tailored to the newsroom's specific needs and values.

To systematize vendor selection, a proposed evaluation matrix is presented in Table 1.

Table 1. Matrix for evaluating AI tools for fact-checking

Evaluation Criterion	Weight (1–5)	Tool A (Monitoring)	Tool B (Text Verification)	Tool C (Image Analysis)
Functionality				
Accuracy and Reproducibility	5	4 / High accuracy in keyword detection, but average in virality assessment.	5 / Consistently extracts facts, but requires citation verification.	3 / Works well with reverse search, but ELA analysis is unstable.
Data Sources and Coverage	4	5 / Covers all key social networks and news aggregators.	4 / Relies on high-quality academic and news databases.	4 / Uses an extensive image database, including archival sources.
Ethics and Law				
Transparency and Explainability	5	3 / Ranking algorithm is a black box.	5 / Always provides source links for each extracted fact.	2 / Does not explain which features are used to detect manipulation.
Bias Management	5	3 / Risk of amplifying popular topics at the expense of niche ones. The developer provides an unbiased report.	4 / Claims to reduce bias, but requires independent testing.	3 / May misinterpret images involving minorities.
Security and Privacy	4	4 / GDPR compliant, data stored on EU servers.	5 / Offers on-premise solution for maximum security.	4 / Cloud-based solution with reliable encryption.
Technology and Operations				
Scalability and Integration	3	5 / Easily integrates with Slack and Trello via API.	4 / API available for CMS integration, but needs refinement.	3 / Works as a standalone app, no API.
Ease of Use and Support	3	5 / Intuitive interface, excellent documentation.	3 / Requires training for practical query usage.	5 / Straightforward and user-friendly interface.
Cost	2	3 / High subscription cost, but justified by functionality.	4 / Moderate cost, flexible pricing plans.	5 / Free basic version, with a paid tier offering advanced features.
Final Weighted Score		**4.03**	**4.38**	**3.31**

This instrument enables the newsroom not only to compare products against a unified standard but also to set priorities by weighting each criterion according to its importance for specific editorial tasks. Key criteria include features that are intertwined. In a functional sense, being both accurate and reproducible matters. It measures accuracy via precision, recall and F1 scores for classification/detection tasks. For identical inputs, a system must output the same to be reproducible. Data sources, as well as coverage, determine the quality, breadth, and relevance of the data. For fact-checking, authoritative and also diverse sources are indeed necessary, and this is precisely on what the tool is trained to operate then. Ethically as well as legally, transparency and explainability matter: one must trace inference logic, understand training data, and obtain citations to supporting sources; black boxes risk compromising journalism greatly. To manage bias, you must build in mechanisms that detect as well as mitigate systemic, statistical, and human biases, and ensure the tool does not perpetuate stereotypes or discriminate against individuals. To secure and protect privacy, one must comply with data protection laws, such as the GDPR, and adhere to strong protocols that safeguard newsroom confidences and sources. Integration via APIs into the editorial systems, along with scalability, is indeed technically necessary, as the tool must process growing data volumes. Usability encompasses a gentle learning curve, an intuitive interface, and strong documentation. Service is also a type of support that is offered. A cost assessment should finally incorporate subscription, implementation and maintenance costs, weighed up against efficiency gains and the benefits expected.

Such a matrix converts software selection from a subjective to an objective and manageable process. This enables a tactically sound decision toward the long-term success of the hybrid fact-checking model.

Architecture of the Hybrid Workflow: Designing a New Interaction Model

Successful AI integration requires nothing less than bolting on a new tool it fundamentally re-engineers the editorial workflow. The new architecture should be built upon the principle of synergy, in which machine strengths like speed as well as data processing offset limitations of humans. Instead, human strengths such as ethics and critical reasoning guide and supervise machine outputs.

In the hybrid model, AI assumes the primary role of screening and data collection. Its core task is to sift massive information streams, detect potentially unreliable claims, assemble preliminary evidence, and route the most consequential cases to humans. This shifts the journalist's paradigm of work: instead of starting the day with a cold topic search, they receive a filtered, prioritized queue, each item accompanied by an AI-generated preliminary dossier.

For effective operation, within the newsroom, two specialized new roles exist.

AI Operator is a journalist skilled technically or a data expert. An AI Operator is responsible for configuring, monitoring, and optimizing AI systems. System configuration touching keywords, sources as well as topics must be monitored within the duties. Furthermore, fixing problems, introducing enhancements, analyzing to identify mistakes, and refining a model for improved precision using internal newsroom data are duties that can be handled through vendor liaison. The AI operator bridges that divide between technology and editorial practice for it ensures the correct and efficient operation on the machine side.

Verifier, that classic journalist and fact-checker, finds a role that gets more analytical and focused. When they receive an AI signal concerning a suspicious claim, the verifier conducts deep analysis as well as critically assesses AI-collected evidence; they perform contextual evaluation, accounting for subtlety, sarcasm, irony, and cultural context beyond machine reach; reassess source reliability, seek additional independent corroboration, and then ultimately make the final, distinctly human, determination concerning truthfulness.

A typical day for a fact-checker under the hybrid model might be structured as follows. In the morning, the verifier opens a specialized dashboard, not generic news feeds, which displays 10–15 of the hottest claims surfaced by AI over the recent hours, sorted by virality and potential harm. Each claim includes a data packet, which comprises a link to the primary source, a curated set of related publications, results of preliminary checks against a database of previously debunked falsehoods, and a list of relevant sources for further verification. The verifier chooses the highest-priority task and begins not with search but with analysis, shifting from information gathering to synthesis and critique. This change increases both the speed and depth of verification.

CHAPTER 2. THE HYBRID FACT-CHECKING METHODOLOGY: OPERATIONAL PROTOCOLS AND STANDARDS

Protocol No. 1: Automated Monitoring and Escalation

This protocol outlines a step-by-step algorithm for configuring and operating a continuous information-space scanning system to detect potential misinformation promptly and route it to the on-duty fact-checker.

Step 1: Configure the monitoring system (AI Operator).

First, define sources. The AI operator compiles and continuously updates monitoring lists: social-media accounts (politicians, public figures, groups with high misinformation propagation), news sites, blogs and forums, official sources (parliamentary transcripts, agency press releases) for tracking claims needing verification, and audience submission channels (e.g., a WhatsApp chatbot as used by Maldita.es).

Second, configure classifiers and filters. Set thematic classifiers (e.g., health, politics, economy) and keyword

sets tied to current and sensitive topics (e.g., vaccination, elections, climate change) to enable automatic categorization of incoming information.

Step 2: Automatic claim detection (AI).

In real-time, the system scans the configured sources. Using NLP models, it extracts concrete factual claims from text, audio (with automatic transcription), and video, claims amenable to verification (e.g., Unemployment in country N increased by 5% in the last quarter). Statements of opinion or subjectivity are ignored.

Step 3: Automatic prioritization (AI).

Each detected claim receives a check-worthiness score computed from several factors:

Virality (speed and breadth of spread; reposts, likes, comments), source influence (authority or popularity of the originating account/site), potential harm (membership in topics where misinformation may inflict significant societal damage, e.g., health, safety), and novelty (whether it is new or a reprise of a known falsehood). The system compares new claims against a database of previously verified facts; if a match or near-duplicate of an already debunked item is found, it flags and prioritizes accordingly.

Step 4: Escalation and task formation for the verifier (AI).

High-scoring claims are automatically delivered to the on-duty fact-checker's dashboard. For each, the system generates a task card that includes the claim text, a direct link to the primary source, context (e.g., the author's prior posts on the topic), propagation metrics, and links to related materials from the newsroom archive (if matches are found).

This process, depicted in Figure 2, transforms chaotic information streams into an ordered task queue, ensuring journalistic attention concentrates on the most salient and potentially dangerous cases.

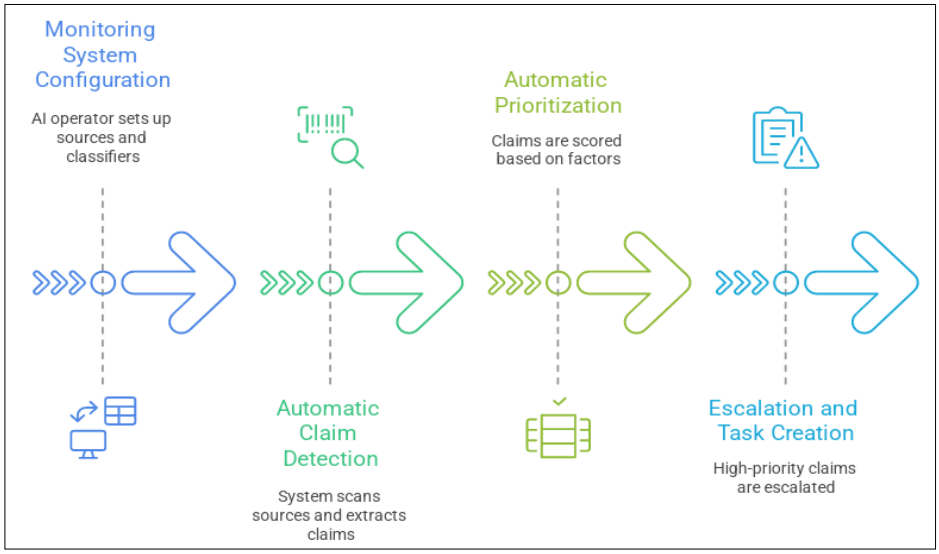

Figure 2. Automated monitoring and escalation protocol

Protocol No. 2: AI-Assisted Verification

Once a task has been created and prioritized via Protocol No. 1, the stage of joint work between journalist and AI begins. This protocol regulates that process to maximize efficiency while preserving full human control.

Consider step-by-step verification of a single falsehood.

Step 1: Task intake and preliminary analysis (Verifier). The verifier receives a notification of a new high-priority task on their dashboard and reviews the AI-generated task card, which includes the claim, source, context, and preliminary dissemination data.

Step 2: AI-assisted evidence collection and analysis (Verifier–AI interaction). The verifier uses an AI assistant to gather statistics, geolocations, dates, names, plus all key factual entities from the source. The AI is then instructed by them to search these entities across the open web along with predefined reliable sources (government databases, scholarly publications, the newsroom archive). The AI gathers and deduplicates materials, and it returns concise summaries for each source. This highlights information that supports or refutes the claim as well as this enables rapid situational assessment since people do not need to read dozens of full texts. The verifier utilizes AI tools for examining images or videos, which facilitate technical analysis. The tools incorporate a reverse image search for identification of the source and context, metadata analysis for verification of the camera, capture date, and further details, as well as Error Level Analysis (ELA) for detection of potential digital manipulation. AI gives all the technical data, yet humans interpret it.

Step 3: Critical human evaluation alongside verification (Verifier-exclusive). This is by far the most important phase. AI outputs require cross-checking. Examining the originals is vital, and the journalist must consult primary sources instead of relying on AI-generated summaries. Automation

is prevented by not relying too heavily on it. The journalist judges the degree to which a source is reliable, as algorithms might retrieve materials at times, but not adequately characterize their authority or reveal bias; professionals must still assess. AI systems often interpret sarcasm, irony, metaphors as well as cultural references literally or overlook same, so they must uncover context with latent meaning. The reporter has to pinpoint artificial intelligence fabrications. Any factual assertion that AI provides must be treated as a hypothesis, and it requires independent confirmation.

Step 4: Verdict and Publication (Verifier-Exclusive)- Synthesizing all vetted data, the journalist renders a final verdict (True, False, Manipulation, or No Verdict). They draft a detailed debunk explaining the verification path, presenting evidence, and supplying context. The article undergoes standard editorial cycles (editing and copyediting) and is then published.

This iterative interaction, where the journalist repeatedly queries the AI and critically evaluates its outputs, is illustrated in Figure 3.

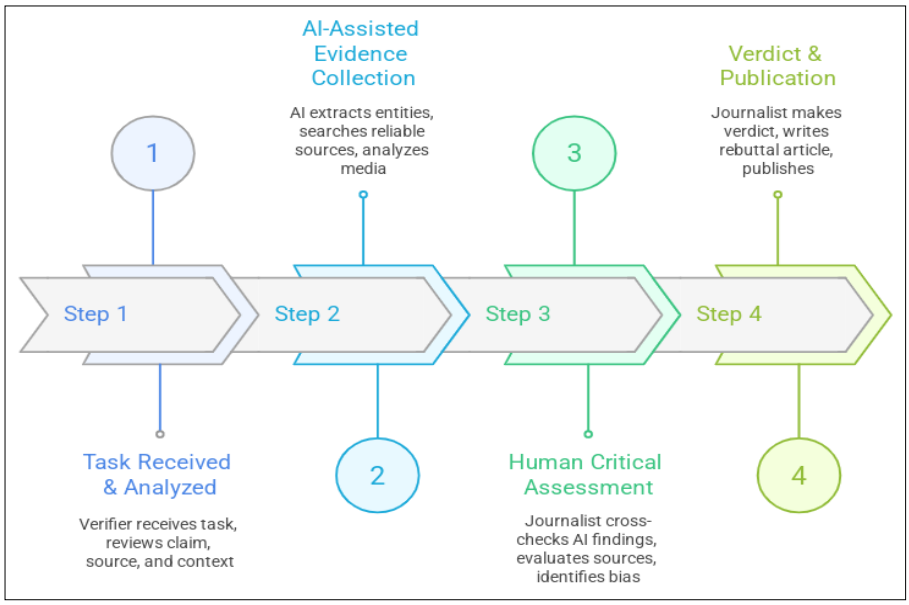

Figure 3. AI-assisted verification cycle

This protocol avoids both the bottleneck of manual search and the risks associated with relying solely on automation, yielding a balanced and reliable verification system.

Verifying Information: An Analysis of the Pizzagate Case

The principles of the hybrid model are universal and apply not only to multimedia content but also to textual disinformation, which often underpins complex conspiracy theories. The 2016 Pizzagate case is a paradigmatic example of how a wholly fabricated narrative, originating on anonymous forums, can precipitate real-world violence (Imamura, 2017). That conspiracy falsely alleged that Comet Ping Pong, a pizzeria in Washington, D.C., was the center of a child sexual-exploitation ring in which then-presidential candidate Hillary Clinton was purportedly involved. Rumors disseminated via anonymous forums and social media culminated on December 4, 2016, when a man from North Carolina entered the restaurant armed with a rifle to investigate these claims.

Consider how a hybrid model could be applied to the analysis and refutation of this disinformation campaign.

Automated monitoring (Protocol No. 1). An AI system configured to track anomalous activity would be capable of detecting the nascent stages of the campaign. Algorithms would register a precipitous increase in mentions of the hashtag #pizzagate and associated keywords on platforms such as 4chan, Reddit, and Twitter. The system would prioritize this narrative as high-risk because of its rapid diffusion and toxic content (explicit accusations of serious crimes) and would escalate it to a human verifier.

AI-assisted verification (Protocol No. 2). Upon assignment, the verifier would employ an AI assistant to analyze the message stream and automatically extract discrete, verifiable assertions. Examples of such extracted claims might include: "A child-sex exploitation network operates in the basement of the Comet Ping Pong pizzeria," and "Hillary Clinton and John Podesta are implicated in this network."

The next step would be an evidentiary search for these claims. The AI assistant could rapidly scan news archives, court records, police reports, and other authoritative sources. In the case of Pizzagate, this search would return no corroborating evidence because no verifiable support existed. The complete absence of confirmation in reputable sources constitutes a strong indicator of the claims' spurious character.

AI-driven network-analysis tools would also allow visualization of how the narrative propagated from anonymous forums to a broader audience via influential

social-media accounts and alternative media outlets. Such analysis clarifies the campaign's propagation mechanics and identifies key amplification nodes.

On the basis of the total lack of evidence and the provenance analysis of the narrative, the verifier would render a definitive verdict of False and produce a detailed refutation that explains how the conspiracy theory was fabricated and disseminated.

This case illustrates how a hybrid model can effectively deconstruct even complex, text-based disinformation campaigns by progressing from automated detection to in-depth human analysis and the issuance of a well-founded verdict.

Task Allocation Matrix

A key element of the Hybrid Fact-Checking methodology is the Task Allocation Matrix. This tool is not merely advisory, but a strict organizational and ethical standard that formalizes the division of labor between AI and humans. Its purpose is to maximize efficiency by automating routine operations while safeguarding the core of journalistic practice, critical judgment, contextual analysis, and ethical responsibility from improper automation.

The matrix prevents two principal errors in AI adoption: over-reliance (delegating tasks that require human nuance to AI, risking serious factual and ethical errors) and underutilization (retaining mechanical, labor-intensive tasks for humans that AI can execute faster and at scale, thereby nullifying the technological advantage).

The proposed matrix, shown in Table 2, partitions all fact-checking cycle tasks into three categories, clearly defining who holds ultimate responsibility.

Table 2. Human-AI task distribution matrix in hybrid fact-checking

Category	Key tasks (short)	Executor	Note
Fully automated (AI-only)	Monitor many sources; auto-transcribe audio/video; detect duplicates; check DB of known fakes; simple virality score	**AI**	Fast, high-volume jobs - human only for setup/oversight
Human-in-the-Loop	Prioritize claims; collect evidence; summarize docs; extract names/dates/numbers; basic image checks	**AI + Human**	AI preps data; human verifies and validates
Human-only	Final truth verdict; judge source reliability & bias; read sarcasm/context; ethical decisions; contact sources; write final text; fix AI errors	**Human**	Requires judgment, ethics, and communication - cannot be automated

Embedding this matrix in the editorial charter sets up a system that is transparent as well as accountable. It acts in the capacity of a practical guide for each staff member. When they use AI tools, it clarifies the boundaries around capability and responsibility. Furthermore, it is a key instrument in risk management, and it codifies the principle that a qualified journalist must participate within and approve any meaningful analytical or ethical decision. Therefore, technological modernisation strengthens the fundamental values that journalism has instead of weakening them.

CHAPTER 3. RISK MANAGEMENT AND PERFORMANCE EVALUATION IN THE HYBRID MODEL

Methods for Risk Minimization

Introducing AI into editorial processes, despite its apparent benefits, entails significant risks that demand proactive management. The most serious concerns are AI hallucinations, systemic algorithmic biases, and the potential for external manipulation. This chapter presents research-informed, practical methods to minimize these threats.

Combating AI Hallucinations

Hallucinations occur when generative AI models produce outputs that appear credible yet are factually incorrect or entirely fabricated. This is surely an existential threat to any newsroom. The newsroom's reputation absolutely hinges upon accuracy. Consider mitigation methods.

1. Technological approach: Retrieval augmented generation (RAG). Do implement RAG systems instead of letting AI respond from a general as well as opaque world model. This architecture compels the AI to ground its outputs solely in information retrieved from a predefined, trusted knowledge base for example the newsroom archive, scholarly journals, official government sources. Before the system generates an answer, it locates relevant fragments within this base also uses them as context, substantially reducing fabrication and improving how citable sources are.
2. Procedural approach: Mandatory verification protocol. The editorial charter must codify the rule that any factual assertion generated or suggested by AI cannot be used in publication unless a journalist has independently verified it through at least two authoritative primary sources. This zero-trust principle toward AI outputs is foundational. The journalist's decision workflow when encountering a potential hallucination is depicted in Figure 4.
3. Feedback and fine-tuning. Provide an easy mechanism for journalists to flag hallucinations. The AI operator should use this data for regular model fine-tuning, allowing the system to learn from errors and reduce their frequency over time.

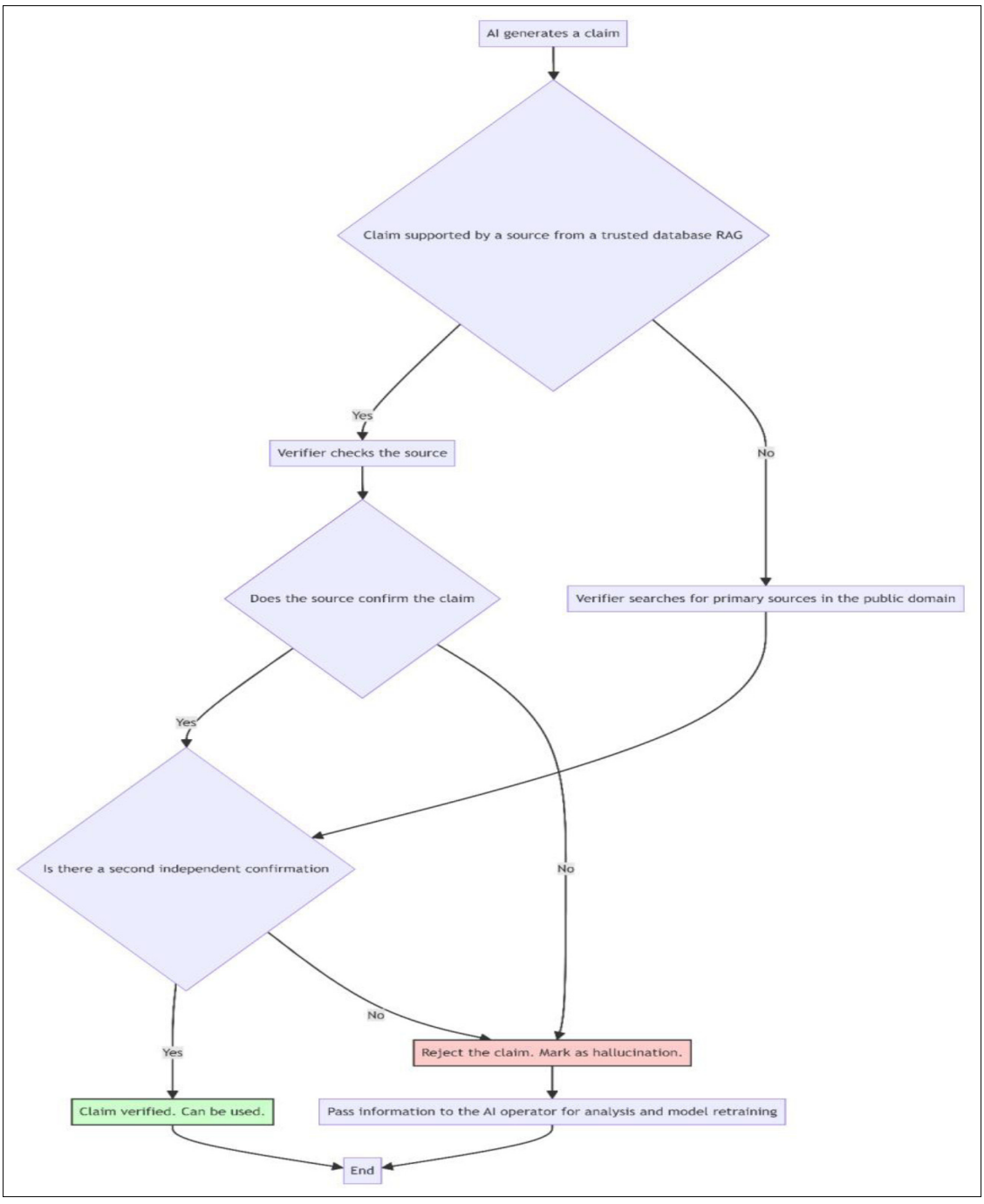

Figure 4. Journalist's protocol for verifying an AI-generated claim

Managing Systemic Biases

Algorithmic bias is a systematic error within an AI model here that results in outcomes unfair or inaccurate for certain people or for groups. It can arise due to skewed training data, the algorithm itself, or human decisions made during development (Schwartz et al., 2022). Indeed, bias in fact-checking can emerge within the topics or sources the system considers essential, potentially distorting the information agenda.

Mitigation measures include orienting in the direction of frameworks, such as the IEEE P7003 Standard for Algorithmic Bias Considerations, which offers a methodology for identifying, analyzing, and indeed mitigating unintended and unjustified biases in algorithmic systems (Koene et al., 2018). Teams should also conduct regular audits of both training data and models, including representativeness checks as well as performance testing across subgroups, involve technical specialists along with journalists, editors, legal experts, and community representatives where possible in selection, implementation, and oversight because team diversity helps surface blind spots and biases, maintain an internal registry of AI tools as well as their purposes as well as known limitations including potential bias risks, and develop a bias impact statement for each new tool. The bias-management process should be continuous, as shown in Figure 5.

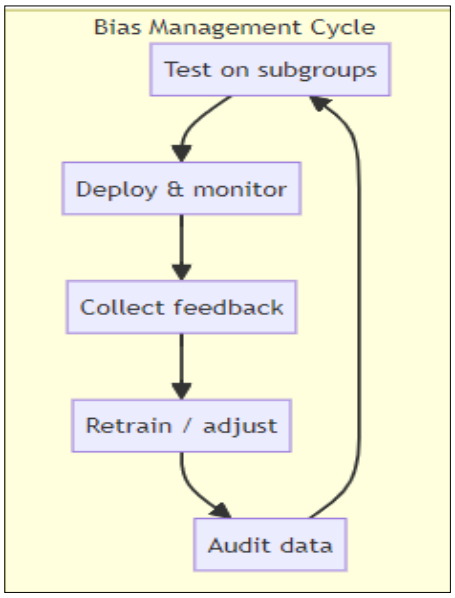

Figure 5. Algorithmic bias management lifecycle

These methods cannot eliminate risk (which is impossible), but they can establish a robust control system that ensures AI use aligns with journalism's high ethical standards.

KPI System for Hybrid Fact-Checking

Evaluating the effectiveness of the hybrid model requires more than measuring speed and output volume. A balanced system of Key Performance Indicators (KPIs) must reflect both operational performance and the quality, accuracy, and real-world impact of fact-checking work on the information environment. Traditional output-oriented KPIs should be supplemented by metrics capturing verification quality and audience impact.

The KPI system for a hybrid newsroom should include three main metric categories presented in Table 3.

Table 3. Key performance indicators (KPIs) for hybrid fact-checking

KPI Category	Indicator	Formula / Measurement Method
Efficiency and Speed	Mean Time to Debunk (MTTD)	Average time (in hours) between the detection of a fake by the monitoring system and the publication of a debunking article.
Verifications per Fact-Checker (VpF)	Total number of published verifications during a given period divided by the number of staff fact-checkers.	Enhance journalist productivity by automating routine tasks.
Quality and Accuracy	Pre-Publication Prevention Rate (PPR)	(Number of inaccurate claims identified and corrected by the AI system before publication) / (Total number of verified claims). Internal metric.
Inter-Checker Agreement (ICA)	Percentage of matching verdicts issued by two different verifiers on the same set of claims (measured via blind testing).	Ensure consistency and objectivity in verdicts, and reduce subjectivity.
AI Fact Extraction Accuracy	Precision, Recall, and F1-score for the task of automatic fact extraction from text, measured on a test dataset.	Monitor and improve the quality of core AI models.
Impact and Reach	Dissemination Reduction Ratio (DRR)	Comparison of dissemination speed (e.g., retweets per hour) of a fake claim before and after the debunking is published.
Fact-Check Citation Index (FCI)	Number of references to a published fact-check in other reputable media outlets and official sources over a defined period.	Measure the authority and recognition of the editorial team's work in the media community.
Audience Engagement Score (AES)	Composite metric including views, reading time, shares, and comments on debunking articles.	Evaluate the effectiveness of fact-checking content in attracting and retaining audience attention.

For clarity in tracking, KPI dynamics should be visualized in real-time by way of an internal dashboard. The implementation of a multidimensional KPI system is intended to ascertain whether the newsroom's output has demonstrably improved following the adoption of AI. It shifts focus away from mere output counts to a thorough appraisal of speed, accuracy, and real-world impact. The hybrid model's active refinement as well as calculated decisions have an objective basis since it provides that focus.

Transparency Protocol

In an era of mounting skepticism toward both media and technology, transparency in AI use is not merely an ethical requirement but a key factor in maintaining and strengthening audience trust. A lack of clear communication can erode trust, even when technology is applied for good.

This protocol offers concrete recommendations for informing audiences and building honest, open relationships. Newsrooms should develop and publicly publish an AI use policy, accessible under "About Us" or "Editorial Policy," that clearly explains the principles and practices of AI in journalistic work. Each organization should articulate a philosophy of AI use that emphasizes AI as a tool that augments journalists rather than replaces them, and affirm that final responsibility for published content rests with individuals.

The policy must clearly enumerate the domains for AI application like social media monitoring, interview transcription, and large-scale data analysis to clarify acceptable practices. It also must demarcate limits of use. AI should never be employed for activities such as it writing news articles from scratch, it creating images without any labeling, or when it is issuing the final fact-checking verdict.

The policy should specify supervision and checking systems, and it should feature methods of human monitoring and responsible positions that ensure correctness and morals in AI-supported resources. In order to reflect any substantial role of AI, the newsroom should implement clear consistent labels in publications with contextual disclosures placed directly within the material, rather than relegated to a general policy page.

A useful label answers how this fits with editorial norms, how humans took part, why it was used and what the AI did. Format, as well as the audience, will affect the wording. Visual design must be noticeable yet remain unobtrusive, as it informs the readers without impairing their comprehension.

AI should not author articles, and journals should not publish pieces under the name of Artificial Intelligence or under a fictitious name. For every piece, a specific journalist or an editorial team must be held responsible. Newsrooms must explain policies as well as practices regularly by means of articles, webinars, and Q&A sessions involving the editors, also they must proactively engage audiences instead of awaiting questions.

Educational materials that elucidate the principles, benefits, and risks of AI in the journalistic domain constitute a valuable supplement, enhancing institutional transparency and fostering audience media literacy. The introduction of the associated protocol may, however, give rise to potential risks. To strengthen public trust, appropriate mitigation measures are implemented; practices are guided by adherence to rigorous journalistic standards, integrity, and accountability.

CONCLUSION

This methodology presents the Hybrid Fact-Checking Model as a systemic and optimal approach for modern media confronting unprecedented challenges in the information environment. The analysis shows that the customary manual fact-checking process, quite limited in scale and in speed, and also fully automated systems, error-prone and incapable in any deep contextual analysis, cannot alone effectively counter modern misinformation. Through forging a strong synergy between AI speed with analytic capabilities as well as journalists' irreplaceable cognitive and ethical competencies, the proposed hybrid model resolves this fundamental contradiction.

The main finding is that successful AI integration is not technological, but about organisation and strategy. It involves more than acquiring software; it requires rethinking editorial processes, introducing new roles, developing strict operational protocols, and establishing strong risk management. The core elements do constitute the basis of the methodology. They translate the abstract notion about human-machine collaboration into concrete, reproducible, and governable workflows, which include Automated Monitoring and Escalation, AI-Assisted Verification, and the Task Allocation Matrix. The model asserts that AI is a powerful instrument to extend a journalist's cognitive reach, but it is never a substitute for critical judgment and ethical responsibility.

The practical significance lies in offering a deployment-ready guide for media leaders, editors, and product managers. Rather than theoretical musings, it supplies a comprehensive, structured action plan for newsroom modernization. The methodology enables media organizations to transition from a fragmented, unsystematic use of isolated AI tools to the development of a coherent, efficient, and safe hybrid verification system.

Thus, the methodology is a strategic asset enabling news organizations not merely to survive in an information war but to lead, effectively fulfilling their public mission. The Hybrid Fact-Checking Model is well-suited to the current state of technology and information threats; however, the field is evolving, opening new horizons for research and practice.

Potential directions in which further work goes include:

1. Full automation narrows tasks. It can be explored for complete automation in verification for specific claim

types if it is grounded in structured, trustworthy data (e.g., it was checked that statistical indicators against official databases). Advanced models must do more than information extraction alone. The semantics must align across various sources as well.

2. AI tool industry standards along with certification. For journalism, the AI market's continued growth shall require more unified industry standards regarding quality, transparency and also ethics. Newsrooms could receive some help for the selection of tools both reliable and safe through establishing an independent system for certification.
3. Studies on the long-term audience impact. Further sociological study is necessary to explore how the use of AI shapes audience views and trust over time. More media psychology research must explore audience views and trust as transparency protocols evolve with time.
4. Integration of multimodal systems. AI models for analyzing text, images, audio, and video, collectively (including deepfake detection), will develop into much more powerful and versatile verification systems.
5. Personalization of fact-checking. Investigating AI-driven personalization of debunking and educational content for different audience segments, potentially increasing the effectiveness of counter-misinformation efforts at the individual level.

Advancing along these lines will refine the hybrid model, making it an even more powerful instrument for journalism in defense of an informed, democratic society.

REFERENCES

1. Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science Advances*, *7*(36). https://doi.org/10.1126/sciadv.abf4393
2. Caled, D., & Silva, M. J. (2022). Digital media and misinformation: An outlook on multidisciplinary strategies against manipulation. *Journal of Computational Social Science*, *5*(1), 123–159. https://doi.org/10.1007/s42001-021-00118-8
3. Koene, A., Dowthwaite, L., & Seth, S. (2018). IEEE P7003™ standard for algorithmic bias considerations. *Proceedings of the International Workshop on Software Fairness,* 38–41. https://doi.org/10.1145/3194770.3194773
4. Muhammed, S. T., & Mathew, S. K. (2022). The Disaster of misinformation: a Review of Research in Social Media. *International Journal of Data Science and Analytics*, *13*(4), 271–285. https://doi.org/10.1007/s41060-022-00311-6
5. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). Towards a Standard for Identifying and Managing Bias in Artificial Intelligence. *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, *1270*. https://doi.org/10.6028/nist.sp.1270
6. Wang, S., Su, F., Ye, L., & Jing, Y. (2022). Disinformation: A Bibliometric Review. *International Journal of Environmental Research and Public Health*, *19*(24), 16849. https://doi.org/10.3390/ijerph192416849

Citation: Sprinchinat Kateryna, "A Hybrid Fact-Checking Model: A Methodology for Integrating AI-Based Tools into the Editorial Workflow of a News Portal", Universal Library of Innovative Research and Studies, 2022; 15-25. DOI: https://doi.org/10.70315/uloap.ulirs.2022.003.