

Algorithmic Journalism and the “Black Box”

Dr. Rajkumari Lilapati Devi, *Associate Professor, Department of Mass Communication, Manipur University, Imphal, Manipur*

ABSTRACT

Applications of Artificial Intelligence in news industries use algorithms to select, prioritise, and generate news content. While these technologies enhance speed, efficiency, and audience engagement, they also introduce concerns regarding transparency, accountability, bias, and democratic communication. This article critically examines the relationship between algorithmic journalism and the “black box” phenomenon, referring to the opacity of algorithmic systems whose internal workings remain hidden from users and even from journalists. Drawing upon theories of political economy, platform studies, and critical algorithm studies, the paper explores how opaque algorithmic systems shape public discourse, influence journalistic practices, and redefine media power. The article argues that the black-box nature of algorithmic journalism challenges traditional journalistic norms of gatekeeping and editorial policy. It concludes by emphasising the need for algorithmic transparency and ethical governance to ensure democratic accountability in digital journalism.

Keywords: Algorithmic Journalism, Black Box, Algorithms, News Automation, Platform Power, Media Ethics

INTRODUCTION

Algorithms are sets of coded instructions designed to process information and perform specific tasks. Algorithmic journalism refers to the integration of computational technologies and automated systems into journalistic practices (Napoli, 2014). These systems are capable of collecting data, analysing patterns, generating narratives, and personalising news delivery with minimal human intervention. In journalism, algorithms are used in several ways: News Recommendation Systems, Automated News Writing, Audience Analytics, Content Moderation and Ranking. The integration of automated systems into newsrooms has transitioned journalism from a traditionally transparent “glass box” model to one defined by opaque, data-driven architectures. This shift complicates traditional ethical frameworks, as the proprietary “black-boxed” nature of advanced machine learning models often render the underlying decision-making processes inaccessible to both developers and the public (Waardenburg & Huysman, 2022). This opacity complicates the verification of algorithmic neutrality, as automated systems may inadvertently institutionalise social prejudices or obscure the lineage of data inputs used in natural language generation (Dörr & Hollnbuchner, 2017). Such systems frequently obscure critical nodes within the data, model, and inference

layers, thereby challenging the normative expectation of transparency in news production (Diakopoulos&Koliska, 2017).

The black box metaphor, is a concept rooted in systems theory that describes technological processes where internal functions remain hidden from public view. Within the context of modern media, these opaque systems use complex algorithms to manage information flows and social behavior without offering transparency to the users they impact. The sources identify three primary forms of opacity: intentional secrecy for corporate advantage, technical illiteracy among the general public, and the intrinsic complexity of machine learning models that even developers cannot fully explain. This lack of clarity creates a significant imbalance of power, as corporations control visibility and data while shielding their decision-making from oversight. Ultimately, the "black box problem" represents an intersection of technical difficulty and political control, making it nearly impossible for individuals to audit or challenge automated systems. Many algorithmic systems function as "black boxes," meaning that their internal logics, data-processing mechanisms, and decision-making criteria are inaccessible or incomprehensible to users (Pasquale, 2015).

Beyond these external accountability concerns, internal organizational opacity persists, as newsroom engineering teams often limit the circulation of technical insights among editorial staff, increasing the risk of systemic reporting errors (Cools &Koliska, 2024). This disconnect between technical deployment and journalistic understanding leaves newsroom practitioners reliant on guesswork, which fundamentally undermines their ability to critically evaluate and maintain agency over narrative output (Jones et al., 2022). Furthermore, this reliance on proprietary tools creates an ethical-political vacuum, as the speed of deployment often leads to the adoption of superficial "quick-fix" solutions that fail to address the underlying complexities of algorithmic governance (Arora & Sarkar, 2023). Consequently, journalists have begun to adopt investigative strategies that integrate technical hacking and coding skills to deconstruct these obscure systems, effectively turning the newsroom into a space for auditing digital power (Di Salvo, 2025). This emerging paradigm of algorithmic accountability journalism necessitates a move toward reverse engineering input-output relationships to delineate the logic encoded within these inaccessible systems. Such investigative methodologies aim to illuminate how these computational artifacts influence societal power structures and potentially reproduce existing biases (Diakopoulos, 2014)

Algorithmic Gatekeeping and Media Power

Algorithmic gatekeeping refers to the process by which algorithms control access to information, shape content visibility, and influence user engagement on digital platforms (Gillespie, 2018). Unlike traditional gatekeeping, which relied on human editors and institutional authorities to select and filter information, algorithmic gatekeeping operates through automated systems that prioritize, recommend, or suppress content based on complex computational criteria (Shoemaker & Vos, 2009). These criteria often include user behavior, engagement metrics, relevance scores, and platform-specific goals such as maximizing time spent on platforms or advertising revenue (Pasquale, 2015).

Media power in the context of algorithmic gatekeeping has undergone significant transformation. Algorithms act as powerful intermediaries that determine which voices are amplified or marginalized, thereby shaping public discourse and knowledge production (Couldry&Mejias, 2019). This shift raises critical concerns about transparency, accountability, and bias, as algorithms may reinforce existing inequalities by privileging certain forms of content or perspectives (Noble, 2018). Moreover, the opacity of algorithmic decision-making processes limits users' understanding of why certain information is presented or withheld (Pasquale, 2015).

The interplay between algorithmic gatekeeping and media power also influences democratic processes and cultural dynamics. By controlling the flow of information, algorithms can affect political mobilization, public opinion formation, and social polarization (Pariser, 2011). The concentration of algorithmic control within a few dominant technology companies further consolidates media power, raising concerns about monopolistic influence and the erosion of diverse media ecosystems (van Dijck, Poell, & de Waal, 2018).

To critically engage with algorithmic gatekeeping, it is essential to analyze the design, deployment, and impact of these systems while considering their ethical implications. Scholars advocate for greater transparency, algorithmic accountability, and enhanced user agency in digital media environments (Gillespie, 2018). This includes promoting algorithmic literacy and encouraging the development of alternative models that prioritize fairness and inclusivity over purely commercial interests (Couldry&Mejias, 2019).

The transformation of media power through algorithmic gatekeeping is profound. Algorithms function as intermediaries that mediate the flow of information, effectively determining which voices are amplified and which are marginalised (Beer, 2017). This dynamic reshapes public discourse and knowledge production by privileging certain narratives and topics over others. The opacity of algorithmic decision-making processes raises concerns about transparency and accountability, as users rarely understand the logic governing content selection (Pasquale, 2015).

Furthermore, prioritising sensational or emotionally charged content can exacerbate social polarisation by promoting echo chambers and filter bubbles, in which users are predominantly exposed to information that confirms pre-existing beliefs (Pariser, 2011). The concentration of algorithmic control among dominant technology companies contributes to reduced media pluralism and threatens the diversity of independent media ecosystems (van Dijck et al., 2018).

Mention may be made of the case of Facebook's News Feed algorithm during the 2016 U.S. Presidential Election, as it was criticised for amplifying misinformation and politically polarised content (Allcott& Gentzkow, 2017). YouTube's recommendation system has been found to promote increasingly extreme content by suggesting videos aligned with users' viewing histories (Tufekci, 2018). Twitter's shift from a chronological timeline to an algorithmically curated feed altered the visibility of tweets by prioritising relevance and engagement metrics (Gillespie, 2018). Instagram's Explore algorithm tends to reinforce

popularity biases by privileging influencers with large followings, thereby limiting the visibility of emerging or niche creators (Bishop, 2019).

Ethical Considerations of Algorithmic Gatekeeping

The ethical implications of algorithmic gatekeeping are multifaceted. First, the opacity of algorithmic systems undermines accountability because users cannot fully understand or challenge content prioritisation decisions (Pasquale, 2015). Search engines, news feeds, and recommendation systems constantly prioritise content according to criteria such as relevance, popularity, engagement, or profitability. However, these criteria may privilege dominant viewpoints while marginalising minority perspectives (Gillespie, 2014). It might also suppress dissenting voices, amplify dominant narratives, or shape public opinion in ways that are invisible. This raises important concerns about transparency, freedom of expression, and democratic participation in digital societies (Couldry&Mejias, 2019).

Algorithms used by social media platforms also often prioritise emotionally charged or sensational content because such material generates higher user engagement. This process can intensify political polarization and spread misinformation, especially during elections or public crises (Tufekci, 2015). Since algorithms classify, they categorise individuals based on demographics, interests, ideas or behavioural patterns. This classification system may reflect cultural prejudices or flawed assumptions embedded in training data (Noble, 2018). Automated moderation systems sometimes incorrectly flag minority dialects, activist speech, or marginalised identities as inappropriate content. This reinforces stereotypes or unfairly connects individuals with stigmatised categories (Beer, 2017). Such classification errors reveal how algorithmic systems can reproduce existing social inequalities and forms of discrimination (Benjamin, 2019). Advertising algorithms have been criticised for disproportionately associating certain racial or gender groups with low-paying jobs or negative stereotypes. These patterns demonstrate how algorithmic associations may reproduce historical inequalities embedded within data infrastructures (O’Neil, 2016).

Several policy responses have been proposed to mitigate these concerns. One important approach is mandating algorithmic transparency through disclosure requirements and external audits (Gillespie, 2018). Regulatory mechanisms such as algorithmic impact assessments and grievance redressal systems could also improve accountability (Suzor, 2019). Another policy direction involves promoting fairness and inclusivity in algorithm design by minimising bias and ensuring diverse representation (Noble, 2018). Enhancing algorithmic literacy among users can further empower individuals to critically navigate digital platforms and understand the effects of personalisation (Couldry&Mejias, 2019). The antitrust regulations aimed at dominant technology companies may reduce monopolistic control over digital information flows and encourage greater media pluralism (van Dijck et al., 2018). Transparent and accountable content moderation standards, combined with human oversight, can help balance freedom of expression with the need to reduce harmful content online.

Advocacy for greater transparency includes calls for explainable algorithms, open data practices, and mechanisms that allow users to understand and challenge content moderation decisions (Gillespie, 2018). Regulatory frameworks must also evolve to address the

challenges posed by algorithmic gatekeeping. Policies promoting accountability, competition, and democratic safeguards are increasingly necessary in the digital era (Suzor, 2019). At the same time, alternative algorithmic models that prioritise diversity, inclusivity, and ethical considerations over commercial objectives may help foster more equitable digital media environments

The integration of algorithmic systems into news production necessitates a fundamental recalibration of editorial autonomy, ensuring that computational efficiency remains subordinate to the normative requirements of public interest journalism. A proactive editorial management model will necessitate reconciling the inherent unpredictability of machine learning with the rigid ethical standards of the journalistic profession.

Methodological Counter Measures: Reverse Engineering

In response to the opacity of algorithmic systems, scholars and media practitioners have increasingly advocated for methods of algorithmic accountability reporting. To address these concerns, they have developed reverse engineering as a methodological strategy to investigate how algorithmic systems function (Diakopoulos, 2014). Reverse engineering involves systematically observing the relationship between algorithmic inputs and outputs in order to infer the logic embedded within the system. Rather than gaining direct access to source code, researchers examine patterns in algorithmic behaviour to identify biases, inconsistencies, or discriminatory outcomes. This approach is particularly important because algorithms increasingly influence social realities by prioritising information, classifying users, filtering content, and shaping public discourse (Beer, 2017).

In the Input–Output Analysis, there is a systematic classification of the various data fed into an algorithm and observing the resulting outputs. By comparing different outputs under controlled conditions, researchers can infer the rules, priorities, and assumptions embedded in the system (Diakopoulos, 2014).

In fully observable scenarios, an algorithm is accessible through an Application Programming Interface (API) or another public interface, large quantities of data can be collected and identify “deviant” or unexpected outcomes that may reveal hidden biases in the system. In many cases, search engine algorithms may rank certain political viewpoints higher than others, while recommendation systems may disproportionately amplify sensational or polarising content (Gillespie, 2014). A notable example is the investigation of autocomplete systems used by search engines. Researchers found inconsistencies in blocked search terms and discovered that certain politically sensitive or controversial phrases were selectively suppressed. Such findings demonstrated that search algorithms are not neutral mechanisms but systems shaped by institutional policies and cultural assumptions (Noble, 2018).

In many cases, algorithms are only partially observable because the inputs used by the system are hidden from researchers. Under such circumstances, journalists often rely on crowdsourcing to collect outputs from large numbers of users. Surveys, demographic information, or publicly available data can then be used as proxies for missing inputs. Statistical analysis helps researchers infer what variables the algorithm is prioritising or

“paying attention” to (Diakopoulos, 2014). For instance, social media platforms personalise content feeds based on user interactions, browsing histories, and demographic characteristics. Since many of these variables are hidden from public view, researchers often compare outputs across multiple accounts to identify patterns of personalisation or ideological filtering. These studies have contributed to debates about “filter bubbles” and algorithmic echo chambers in digital communication (Pariser, 2011).

Reverse engineering can also be used to investigate what Diakopoulos (2014) describes as the “atomic” decisions embedded within algorithmic systems. These are the basic operations through which algorithms organise and regulate information. Human values and institutional biases are often embedded within these decisions. Another important reverse engineering technique involves analysing false positives and false negatives in algorithmic decision-making systems. False positives occur when a system incorrectly identifies something as problematic, while false negatives occur when harmful content or behaviour goes undetected. The balance between these two errors reflects the priorities and value judgments of algorithm designers (Diakopoulos, 2014). For instance, a content moderation algorithm designed to minimise false negatives may aggressively remove large quantities of content in order to ensure that harmful material is not missed. However, this approach often increases false positives, resulting in the removal of legitimate speech. Such decisions reveal institutional biases that may privilege corporate interests, state security, or advertiser concerns over democratic freedoms and public accountability (Pasquale, 2015). Similarly, predictive policing systems have been criticised for generating disproportionate false positives against marginalised communities because they rely on historically biased policing data. Researchers argue that these systems risk reinforcing existing patterns of social inequality and surveillance (Benjamin, 2019).

Reverse engineering can expose algorithmic bias in practice. Investigations into search engine autocomplete systems revealed inconsistencies in blocked terms and selective censorship practices. Researchers found that politically sensitive or controversial topics were treated differently across geographic locations and institutional regulations. These findings highlighted the role of algorithms in shaping access to information and public discourse (Noble, 2018). Journalists investigating e-commerce platforms discovered that algorithms varied prices according to users’ geographic locations, browsing histories, and purchasing behaviours. Consumers in affluent areas were sometimes shown higher prices than users in less wealthy regions. These practices revealed how algorithmic systems can reproduce economic inequalities through personalised pricing strategies (O’Neil, 2016).

Reverse engineering techniques also expose how political campaigns used personal data, donation histories, and demographic profiles to deliver customised political advertisements. These targeted messaging strategies enabled campaigns to influence voters through highly personalised communication that was often invisible to the broader public (Tufekci, 2015). The controversy surrounding political data analytics firms demonstrated how algorithmic targeting could manipulate political behaviour, spread misinformation, and undermine democratic transparency. Scholars argue that such practices raise serious ethical concerns

regarding privacy, consent, and the concentration of communicative power within digital platforms (Couldry&Mejias, 2019).

Governance and the Future of Transparency

To mitigate the risks of the “black box,” news organizations are encouraged to adopt a Tripartite Governance Model. They are Transparency Protocols: Mandatory audience disclosure when AI is used in news production including explainable output logs and audit trails; Hybrid Workflows (“Human-in-the-Loop”): Maintaining human editorial oversight at critical checkpoints to provide interpretive depth and ethical reasoning that machines lack; Adaptive Regulation: Developing governance frameworks that evolve alongside technological advances and socio-political conditions (Helberger, Diakopoulos, &Klais, 2020).

Emerging technologies like MetaScribe suggest using immutable data storage systems such as blockchain to track data provenance, ensuring that the origins and transformations of any news item are reliably recorded and attributable. Blockchain technology can strengthen accountability in journalism by creating immutable records of provenance. MetaScribe integrates blockchain within its systems infrastructure so that once a piece of data is recorded, it cannot be altered without leaving a traceable record. Blockchain can document: Actor Identification, Process Tracking and Real-Time Tagging. Blockchain functions as a “notarization service,” ensuring that modifications to a news item are verified and attributable. This is particularly significant in an era of AI-generated “deepfake” media. By recording provenance at the point of creation, blockchain-backed systems provide greater transparency into the role of machines in content production.

This reconciliation necessitates a shift toward robust transparency mechanisms that provide clear insights into data provenance and reasoning, thereby enabling both practitioners and the public to trust the systems that translate raw information into journalistic narratives (Ali & Hassoun, 2019). Furthermore, media outlets must reconcile the practical information needs of news consumers—who demand visible, detailed source references—with the technical constraints of proprietary generative architectures (Venkatraj et al., 2025).

CONCLUSION

Algorithmic journalism represents a major transformation in contemporary media systems. Algorithms now influence nearly every aspect of news production, distribution, and consumption and audience engagement. While algorithmic systems offer efficiency, personalisation, and scalability, they also introduce significant concerns regarding transparency, accountability, bias, and media power.

As journalism becomes increasingly automated and data-driven, critical attention must focus on the ethical and political implications of algorithmic governance. Transparency, accountability, and public oversight are essential to ensuring that algorithmic journalism serves democratic interests rather than purely commercial objectives. Control over algorithms has become a form of informational power exercised largely by platform corporations

operating beyond traditional forms of public accountability. Future media policy must evolve to provide formal mechanisms for auditing these automated decision-making processes, ensuring they remain strictly aligned with established editorial values rather than solely market-driven objectives. The algorithmic accountability will likely necessitate a redefinition of the journalist's role from a primary producer to a rigorous supervisor of automated processes. In this role, the journalist becomes the final arbiter of news-worthiness, tasked with guiding, verifying, and refining outputs to ensure they remain grounded in the nuanced context.

The editorial independence becomes the primary challenge in an era characterised by increasing structural dependencies on opaque technological platforms. To mitigate these risks, media organisations must develop procedural frameworks that clearly delineate the allocation of responsibility between business-driven automation and traditional editorial oversight. Implementing these safeguards will be vital to preventing the homogenization of information and the structural dependence on external platforms, which currently threaten platform diversity. Ultimately, institutional oversight will depend on newsrooms' ability to move beyond off-the-shelf implementation toward a more critical, knowledge-based engagement with platform-governed technologies. Such a pivot demands that the industry treats AI not as an immutable, neutral utility, but as an active participant that reshapes editorial hand in the media landscape. This step goes far beyond technological application to unearth the socio-technical power structures and institutional responsibility.

REFERENCES

1. Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
2. Ali, W., & Hassoun, M. (2019). Artificial intelligence and automated journalism: Contemporary challenges and new opportunities. *International Journal of Media, Journalism and Mass Communications*, 5(1), 40–49. <https://doi.org/10.20431/2454-9479.0501004>
3. Arora, C., & Sarkar, D. (2023). Auditing artificial intelligence as a new layer of mediation: Introduction of a new black box to address another black box. *Hipertext.net*, 26, 65–68.
4. Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
5. Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.
6. Bishop, S. (2019). Managing visibility on YouTube through algorithmic gossip. *New Media & Society*, 21(11–12), 2589–2606. <https://doi.org/10.1177/1461444819854731>
7. Couldry, N., & Mejias, U. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
8. Cools, H., & Koliska, M. (2024). News automation and algorithmic transparency in the newsroom: The case of *The Washington Post*. *Journalism Studies*, 25(6), 662–680. <https://doi.org/10.1080/1461670X.2024.2326636>
9. Diakopoulos, N. (2014). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415. <https://doi.org/10.1080/21670811.2014.976411>

10. Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. *Digital Journalism*, 5(7), 809–828. <https://doi.org/10.1080/21670811.2016.1208053>
11. Di Salvo, P. (2025). Dealing with the black box: European journalists and the threats of spyware. *Digital Journalism*, 13(10), 1715–1734. <https://doi.org/10.1080/21670811.2024.2378122>
12. Dörr, K. N., & Hollnbuchner, K. (2017). Ethical challenges of algorithmic journalism. *Digital Journalism*, 5(4), 404–419.
13. Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media technologies: Essays on communication, materiality, and society* (pp. 167–194). MIT Press.
14. Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
15. Helberger, N., Diakopoulos, N., & Klais, Y. (2020). Debating the role of news recommender systems in democracy. *Digital Journalism*, 8(8), 997–1016. <https://doi.org/10.1080/21670811.2020.1760981>
16. Jones, B., Jones, R., & Luger, E. (2022). AI ‘Everywhere and Nowhere’: Addressing the AI intelligibility problem in public service journalism. *Digital Journalism*, 10(10), 1731–1755. <https://doi.org/10.1080/21670811.2022.2145328>
17. Napoli, P. M. (2014). Automated media: An institutional theory perspective on algorithmic media production and consumption. *Communication Theory*, 24(3), 340–360. <https://doi.org/10.1111/comt.12039>
18. Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
19. O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.
20. Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin Press.
21. Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
22. Shoemaker, P. J., & Vos, T. P. (2009). *Gatekeeping theory*. Routledge.
23. Suzor, N. P. (2019). *Lawless: The secret rules that govern our digital lives*. Cambridge University Press.
24. Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, 13(2), 203–218.
25. Tufekci, Z. (2018, March 10). YouTube, the great radicalizer. *The New York Times*.
26. van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
27. Venkatraj, K. P., Morosoli, S., Cools, H., Naudts, L., de Vreese, C., Helberger, N., Cesar, P., & El Ali, A. (2025). Understanding AI disclosure needs for news production and journalism. In S. Sorce, P. Elagroudy, & M. Khamis (Eds.), *Proceedings of the 24th International Conference on Mobile and Ubiquitous Multimedia* (pp. 202–208). Association for Computing Machinery. <https://doi.org/10.1145/3771882.3771899>
28. Waardenburg, L., & Huysman, M. (2022). From coexistence to co-creation: Blurring boundaries in the age of AI. *Information and Organization*, 32(4), Article 100432. <https://doi.org/10.1016/j.infoandorg.2022.100432>