ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0011

Journalism and Communication in the Era of Big Data: A Review of Research

Guangxu Hao

Sichuan University of Culture and Arts, Mianyang 621000, China

Abstract

Big data has evolved from a purely technical tool to a force to reshape the structure of journalism. The review maps the lines of research in the four areas of news production, distribution, trust in audiences and ethical governance. Three main results emerged: in the production process, computer-assisted reports, data news, the computation of news lines, databases and algorithms redefined the distribution of work in the newsroom. In its distribution, plasterization has elevated social media and video networks to news organizations ' own channels as the main gateway to news; Despite individualized recommendations to improve matching efficiency, filtering foams, reduced audience control and other structural ethical issues such as value drift have emerged. From the perspective of the audience, the rapid expansion of the coverage of the news, even when trust and attention are declining, is a widespread exposure and low credibility. There are currently interregional differences, narrow approaches and slow technological advances. Future research should give priority to the establishment of standardized data systems across platforms, the auditing of algorithms and the examination of how news is consumed once genetic AI is part of the process.

Keywords

Big Data; Data Journalism; Algorithmic Recommendation; Platformization; Computational Communication; News Trust.

1. Introduction

Big data is not just a technology. As described by Lewis and Westlund, the term refers to a social, cultural and technological phenomenon - huge digital data, new analytical methods and narratives of data-driven insights, and increasing criticism of the impact of big data practices on society [1]. When this multi-faceted phenomenon enters the press, what changes are not only the technology available to journalists and the perception of the press, how it is produced, and how it is organized, and how it maintains agreements with social values and values.

Consider the scale table. According to the projections of the IDC global data layer, approximately 274 Zetabits are expected to be generated by 2026 and over 718 Zetabits by 2030, and global storage is expected to increase by only 26 Zetabits from 13 Zetabits during this period. The difference in size between the generation and retention of data is the material basis that any news dissemination review now has to face. As of December 2025, there were 1,125 million Internet users in China; According to CNNIC, China's 56th Statistical Report on Internet Development, as of June 2025 there were 1,068 million short video users, representing 95.1 per cent of all Internet users, and a total of 626 million microatric viewers. These are an overview of the empirical background, with the audience shifting mostly to mobile video, and news organizations must gain the attention of third-party platforms.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0011

Based on the four axes of production, assignment, trust and governance, this review establishes the primary dynamics and debate on the ground, and then highlights the limitations that exist and the way forward.

2. The Reconstruction of News Production: From Assisted Reporting to Computational Journalism

2.1 Three Quantitative Forms, and the Lines between Them

When quantitative methods are now fully integrated into professional journalism, “data journalism” in its broad sense can no longer capture actual differences. Codington separates computer-assisted reporting (CAR), data news from computational news, and proposes four partial types: emphasis on professional expertise and network engagement, transparency and non-transparency, reliance on large data and target sampling, and positive perceptions of passive audiences [2]. Distinction is important because all three forms have a “data-talking” shell but reflect unique news values and career visions.

2.2 Epistemological Conflict: From Disclosing Public Issues to Producing Publicness

Further tension emerged from empirical research carried out in the Chicago press room by Parasi and D ' Aguilar. Since the mid-2000s, United States and British news organizations have begun hiring programmers to develop data-driven news projects in their newsrooms. These “producer-journalists” have experience from open-source communities and open government advocacy campaigns that challenge some of the ideas rooted in the American CAR tradition (i.e. that data can help shape the political agenda by exposing public issues).[3] The difference, therefore, is between the use of data and who they are and in what way they generate openness.

2.3 Automated Writing and the Redefinition of Labor and Authority

According to Carlson's “Romantic Reporter”, automated writing is not only a tool for efficiency, but also a change in news work, in the form of text-formulation and in the generation of press authority.[5] Puntland has broadened this perspective to include all changes in the production of intelligent content, considering that smart technology has triggered a “new content revolution” and that the role of authors has evolved from human beings to human beings, cooperation of machines, platforms, content formats, distribution channels and user functions [6].

Pushing further, the fourth lineage-computational journalism-treats computation not as an instrument applied to journalism but as a constitutive condition of it, folding algorithmic selection, platform metrics, and machine learning into the very definition of what counts as news. Diakopoulos argues that this shift brings with it a distinctive set of values-accuracy, objectivity, and accountability-that must be deliberately designed into computational systems rather than assumed to emerge from them [4]. The consequence is a redistribution of epistemic authority: the newsroom's claim to know is increasingly mediated by infrastructures it neither owns nor fully understands, and the journalist's judgment must be articulated alongside, and sometimes against, the logics of the systems that produce and rank content. It is here that the production and distribution of news become difficult to separate, since the same algorithmic apparatus that generates stories also determines their visibility, and the same metrics that inform editorial decisions also discipline them. This entanglement sets the stage for the review's second axis, which turns from how news is made to how it reaches, or fails to reach, its audiences.

3. The Turn in Distribution: Platformization and Algorithmic Recommendation

3.1 Platformization: Third-party Entry Points Overtake Owned Channels

Reuters's 2026 Digital News reported that, including 48 markets and nearly 100,000 interviewees, 77 per cent of the population viewed online news videos every week. The use of AI chats for news has increased from 7 per cent a year ago to 10 per cent, and the number of children under 35 has risen to

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0011

16 per cent. Twenty-seven per cent received some news on individual creators or influencers with a focus on news, and 46 per cent on any type of creator.

In China, this change is shown in video, with 1,068 million short video users and 626 million microtrama users. An increasing number of news agencies have arranged their content for the platform rather than announcing on their own channels that distribution control has shifted from the newsroom to the platform ' s recommended engine.

3.2 The Structural Ethical Risks of Algorithmic Recommendation

Chan Changfong and Sven have technically solved this problem by examining three recommended algorithm structural deficiencies: content-based recommendations often create filter bubbles, cooperative filtering prevents viewers from controlling their reading options, and time-based popular proposals may create a wave of digital “yellow news”. They also believe that, in addition to the deviations in the input of data, these structural deficiencies mean that moral defects cannot be corrected by technical means alone - because news is a special product, and a more comprehensive human value is needed to reconcile the value of algorithms from an external perspective [7].

3.3 Treating the Algorithm as an Object of Investigation

The second line operates “as opposed to the distribution of algorithms”: the algorithm itself becomes narrative. Journalists can review algorithms as dynamic computational structures through reverse engineering, interviews, data scratching and black box testing, as implied by the "calculatory accountability" concept of Diacoplos [4]. This has transformed the role of the news from the use of data to the monitoring of data systems, creating new public responsibilities for data journalists.

Where, then, does the field go from here? The three turns traced above-in production, distribution, and epistemology-are not separate stories but a single institutional transformation unfolding at different speeds. Data journalism's early promise was that numbers would open the state to scrutiny; platformization and automation have since made the newsroom itself the site where algorithmic power must be interrogated. This raises the stakes of the questions that follow: if the authority to define what counts as news now migrates toward recommendation systems and generative models, journalism's public mandate depends less on its capacity to publish than on its capacity to render these systems legible and contestable. The sections that follow examine the shortcomings of the existing research landscape and chart the directions-conceptual, methodological, and normative-along which the field must move if it is to keep pace with the objects it studies.

4. Audiences and Trust: The Paradox of High Reach and Low Trust

The same Reuters report records a second, sobering set of figures. Global trust in news has fallen to 37 percent, the lowest since measurement began in 2015, with declines in 29 of 48 markets. Since 2021, the share describing themselves as extremely or very interested in news has fallen by an average of 13 percentage points. Twenty-five percent are now "casual or passive" users who engage with news about once a week and express little interest, up from 16 percent in 2021. Concern about fake news rose four points to 62 percent. Across twenty tracked countries, the share paying for online news remains stuck at 17 percent.

One contrast deserves emphasis: despite extreme fragmentation of distribution channels, 45 percent of respondents still prefer news that does not take sides. The public has not abandoned its expectation of professional news norms; it simply satisfies that expectation less and less often through professional channels.

Read through Lewis and Westlund's four lenses, this is precisely the point at which journalism's economics and its ethics come under pressure simultaneously. Platforms intercept audiences and dilute the commercial value of legacy institutions; because audiences encounter news through third-party channels they trust less, aggregate trust is pulled down structurally.

The trust deficit is not simply a matter of audiences abandoning journalism; it is a structural consequence of the platform turn described above. When recommendation systems mediate the

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0011

relationship between news organizations and their publics, trust is no longer negotiated in the newsroom but in the opaque architectures of ranking and personalization. Lewis and Westlund's framework suggests that this displacement has an epistemological dimension: audiences cannot easily verify the provenance of the news they encounter, nor can they hold anyone accountable for its selection. The 45 percent who still prefer unbiased news represent a latent demand that professional journalism has failed to meet through its current distribution strategies, not a public that has lost interest in the norms journalism claims to uphold.

This paradox of high reach and low trust also carries implications for how the field conceptualizes its audience. If trust is produced relationally-through repeated interactions in which audiences can observe and evaluate journalistic practice-then the migration of news consumption to platforms that obscure these practices undermines the very conditions under which trust can be built. The decline from 16 to 25 percent of "casual or passive" users suggests a widening gap between the routine presence of news in daily life and the active engagement that trust formation requires. What appears as audience disengagement may in fact be an artifact of an information environment that offers reach without reciprocity, visibility without verification.

5. Ethics and Governance: From Privacy Protection to Algorithmic Accountability

Borges-Rey's semi-structured interviews in British mainstream newsrooms found that most professional journalists still display hesitancy toward numbers and computational literacy, which directly limits their capacity to investigate the power structures that data sustains [6]. This echoes the "expertise" lens in Lewis and Westlund: big data requires not that journalists be able to calculate, but that newsrooms possess the organizational capacity to interrogate algorithms and trace the provenance of data .

At the level of governance, discussion has moved from early concern with individual privacy toward algorithmic transparency, explainability, and platform responsibility. China's approach emphasizes inclusive, prudent, and agile governance of new technologies: as of April 2026, 868 large-model services had completed regulatory filing, and generative AI has been brought within a framework of standardized development.

This regulatory infrastructure, however, is largely external to the newsroom. Filing requirements and transparency obligations structure how firms develop and deploy models, but they do not by themselves equip journalists with the means to interrogate the systems that now shape news selection, ranking, and generation. The gap between governance capacity and newsroom capacity is thus the central problem the field has yet to address: regulation can mandate disclosure, but it cannot supply the interpretive skills, organizational routines, and institutional incentives through which disclosure becomes accountability. A filed model is not an examined model, and transparency documents prepared for regulators rarely answer the questions journalists would need to ask-about training data provenance, about the criteria embedded in ranking decisions, about the failure modes that only surface under sustained adversarial testing. Closing this gap requires treating algorithmic accountability not as a norm to be declared but as a craft to be built, with its own methods, standards, and professional rewards.

6. Shortcomings and Directions

Three gaps stand out in the literature. The first is the imbalance in geography and context. Empirical work concentrates heavily on the Anglophone world and Western Europe; systematic evidence from the Global South and from Chinese-language contexts remains thin, and the cross-cultural transferability of many conclusions is untested. The second is methodological narrowness. Existing findings rest largely on interviews and case studies, with little comparable large-scale data across platforms or periods and few designs capable of identifying causal mechanisms. As a result, core questions such as whether algorithms truly cause polarization remain stuck at the level of correlation.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202611_12(11).0011

The third is theory lagging behind iteration. Generative AI and chatbots have entered news consumption faster than conceptual frameworks and norms have been updated. The Report's observation that chatbot users are markedly less likely to click through to original news sources still lacks an adequate theoretical account.

Four directions follow. Build cross-platform, comparable infrastructure for audience behavior data so that cross-national replication becomes possible. Advance algorithmic auditing and transparency research, turning "algorithmic accountability" from a claim into operable and reproducible method. Study the norms and division of labor in human–machine collaborative newsrooms, clarifying responsibility for automated content. And attend to the mediating mechanisms through which AI-mediated reading shapes trust, memory, and the formation of public issues-answering what the public actually remembers when news is relayed by a machine.

7. Conclusion

Big data have shifted the news from “talking about what happened” to “investigating what should be said in mass traces”. The same transition took place in four interrelated places: productive rewriting of epidemiology, platformization of distribution, low confidence in the audience and algorithm accountability in governance. The question for researchers is not whether big data will change news, but how big data will change news, who will be affected and who will be responsible for adjusting. The solution does not seem to be to make the press more proficient in the use of data, but rather to ensure that it can assess its own data skills.

Acknowledgments

This work was supported by the 2024 University-Level Scientific Research Project of Sichuan University of Culture and Arts (General Project), entitled “Research on the Development of Journalism and Communication in the Context of Big Data” (Project No. CWYB202471).

References

[1] Lewis, S. C., & Westlund, O. (2015). Big data and journalism: Epistemology, expertise, economics, and ethics. *Digital Journalism*, 3(3), 447–466. https://doi.org/10.1080/21670811.2014.976402

[2] 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

[3] Parasie, S., & Dagiral, E. (2013). Data-driven journalism and the public good: “Computer-assisted-reporters” and “programmer-journalists” in Chicago. *New Media & Society*, 15(6), 853–871. https://doi.org/10.1177/1461444812462418

[4] Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

[5] Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, 3(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

[6] Peng, L. (2018). A new revolution of content in the intelligent era. *Chinese Journal of Journalism & Communication*, 40(6), 88–109. (In Chinese)

[7] Chen, C., & Shi, W. (2018). A technical reading and value analysis of personalized news recommendation algorithms. *China Editor*, (10), 9–14. (In Chinese)