http://doi.org/10.62405/osi.2026.02.03

VERIFICATION OF SCIENTIFIC INFORMATION IN DIGITAL JOURNALISM AMID THE DEVELOPMENT OF OPEN DATA AND GENERATIVE ARTIFICIAL INTELLIGENCE

Danyil Zemlianov,
V. N. Karazin Kharkiv National University, Kharkiv

Abstract. *Scientific information has become much more accessible to digital journalists through open-access publications, preprints, repositories and research datasets. This access gives reporters more opportunities to work with primary evidence, while also requiring a clearer understanding of publication status, version history and the limits of available data. Generative artificial intelligence adds another layer to this process. It can speed up search, translation and text processing, yet it can also introduce fabricated references, factual inaccuracies and misleading simplifications. The purpose of this article is to develop a practical model for verifying scientific information in digital journalism with attention to open data, FAIR principles and the use of generative AI. The study relies on literature analysis, comparison, systematisation and theoretical generalisation. Particular attention is paid to source provenance, peer-review status, metadata, persistent identifiers, publication versions and the link between a public claim and its evidentiary basis. The analysis resulted in a five-level verification model covering source identification, assessment of publication status, examination of evidence, external cross-checking and review of AI use. FAIR principles are treated as a practical framework for finding, tracing and responsibly reusing research data. The proposed approach can support newsroom protocols, science reporting, fact-checking and data journalism where accuracy depends on a clear path from a published claim to its original source.*

Ключові слова: *open science, open data, digital journalism, scientific information, verification, FAIR, generative artificial intelligence, data journalism, fact-checking.*

INTRODUCTION

Digital journalism has gained direct access to sources that until recently remained largely within the academic environment. Today, a journalist can open the full text of an article, find a preprint, access a dataset, verify a DOI, review version history and examine materials stored in a repository. This way of working creates more opportunities for independent verification, but it also requires closer attention to the status of each document. Open science creates conditions in which the origin of a claim and part of its evidentiary basis can be traced directly. UNESCO links this model with transparency, critical scrutiny, verifiability and the reuse of scientific results (UNESCO, 2021). In Ukraine, this direction is embedded in the National Plan for Open Science, which gives particular attention to open access, research data and integration into European research infrastructures (Cabinet of Ministers of Ukraine, 2022).

For a newsroom, the mere fact that a source is openly accessible says little about its scientific status. A peer-reviewed article, a preprint, a presentation, a repository record, a standalone

© Zemlianov Danyil, 2026

dataset and a social-media summary may all appear next to one another in search results. They have different origins, pass through different review procedures and provide different levels of information about research methods. This is where one of the main difficulties for journalists arises. A publication available without registration or payment may look convincing even though its conclusions are still under discussion. A review of journalists' use of open scientific materials confirms that assessing preprints and open-access publications often requires additional verification and an understanding of context (Fleerackers et al., 2024). An error at this stage can quickly move from an internal editorial decision into a public statement.

Generative systems are having an increasingly strong influence on editorial practice. They are used to identify topics, process lengthy documents, translate material, transcribe interviews and provide preliminary structure for a story. Ukrainian researchers describe the growing use of such tools in the media and connect their application with requirements for transparency, human oversight and professional responsibility (Nadtochii & Prudius, 2024; Yaroshenko, 2024). H. Cools and N. Diakopoulos, analysing journalists' experience, identified a wide range of practical scenarios for the use of generative AI. Alongside convenience, respondents mentioned factual errors, difficulty in establishing the origin of individual claims and the risk of shifting some editorial decisions to the system (Cools & Diakopoulos, 2026).

These risks are especially significant in science journalism, where an error in the title of a study, a numerical value or a bibliographic record can alter the meaning of the entire story. W. H. Walters and E. I. Wilder analysed 636 bibliographic references generated by early versions of ChatGPT and found both fabricated sources and inaccuracies in descriptions of real publications (Walters & Wilder, 2023). For this reason, generative AI is more appropriately treated as an auxiliary tool for search and text processing. The final basis for a factual claim should be a primary source that the journalist can open and verify. European Commission guidelines on the use of generative AI in research likewise emphasise human responsibility for outputs, verification of references and transparent disclosure when such tools play a substantial role (European Commission, 2026).

Under these conditions, open scientific resources and editorial verification effectively converge within a single workflow. A journalist needs to establish where the data came from, what status the publication has, whether other versions exist and how closely the stated claim corresponds to the evidence. Generative AI adds another question to this process: at what stage did the system influence the search, summary or interpretation of information? This approach shifts attention from simple access to a source towards the ability to trace the entire path from a scientific statement to the editorial text.

In this article, verification refers to the newsroom process of tracing a scientific claim to an identifiable source, establishing the status and version of that source, and checking whether the available evidence supports the wording used in the story. Fact-checking is used in the narrower editorial sense of testing the accuracy of a public claim against accessible sources. Scientific expertise refers to disciplinary evaluation of research design, methods, statistics and interpretation by a qualified specialist. The proposed model belongs to editorial verification. It can include fact-checking and expert consultation, but it does not replace peer review or independent scientific assessment.

The aim of the study is to develop a model for verifying scientific information in digital journalism that takes open data and generative artificial intelligence into account. To achieve this aim, the study analyses elements of open science that are useful for editorial verification, compares different types of scientific materials and identifies risks that arise when generative systems are used. Particular attention is paid to the accessibility, traceability and evidentiary reliability of scientific information, as well as to a sequence of actions that helps move from a secondary account to the primary source and its evidence base.

LITERATURE REVIEW

Open science is gradually changing the conventional model of scholarly communication. External users now gain access not only to the final article but also to a broader set of materials, including data, code, protocols, preprints, methodological descriptions and repository records. T. Yaroshenko views open access as part of a wider transition towards a model in which openness covers publications, data and the digital infrastructure of science (Yaroshenko, 2021). In Ukraine, this process is also connected with the work of universities and libraries. They maintain repositories, support researchers in disseminating results and gradually develop services through which scientific materials become available to wider audiences (Yaroshenko et al., 2022). This shift matters for the media because a journalist gains more entry points into a single study and can compare the final publication with materials that preceded it.

For journalists, openness of a source and its evidentiary strength are different characteristics. Full-text access makes methods, tables and limitations easier to examine, while scientific status depends on peer review, the sample, statistical approaches and the publication stage. This distinction became particularly visible during the COVID-19 pandemic, when preprints entered public discussion very quickly. A. Fleerackers and co-authors show that journalists can serve as an important link between open science and society, although working with open publications requires careful assessment of the quality and status of results (Fleerackers et al., 2024). Before interpreting a finding, a newsroom should therefore establish exactly what type of scientific material it is dealing with.

The FAIR principles provide a separate framework for working with research data. They concern the findability, accessibility, interoperability and reusability of data. The methodological recommendations of the Ministry of Education and Science of Ukraine associate FAIR with the organisation of electronic data in a way that facilitates search, reuse and interaction among different types of information. The same document emphasises the importance of persistent identifiers and repositories (Ministry of Education and Science of Ukraine, 2024). The European EOSC strategy approaches these principles through the quality of metadata, the use of identifiers and licences, and compatible digital infrastructures (EOSC Association, 2024).

In journalistic work, FAIR is useful primarily because it helps establish the provenance of a digital object. When a dataset has an identified author or responsible institution, a clear description, an update date, a version, a licence and a DOI or another persistent identifier, it becomes much easier for an editor to understand its context. Such documentation provides only part of the information needed to judge research quality, because methodological assessment still requires separate attention to the sample, data collection and analysis. A well-documented dataset may still contain sampling bias or be interpreted too broadly. For this reason, FAIR is treated in this study as an infrastructural basis for traceability and repeat verification. Reviews of FAIR practices also show that the practical value of these principles depends on the quality of metadata, standards, tools and user preparation (Guillot et al., 2023).

In Ukraine, open-science infrastructure is developing unevenly. The 2022 National Plan established a general framework for open access and research-data management, yet institutional practice still varies. O. Rachynska analysed 351 Ukrainian institutions and identified noticeable differences in the availability of repositories, open-access policies and documents devoted to FAIR and data management (Rachynska, 2025). For journalists, this means working with sources that differ considerably in the quality of their documentation. In one case, metadata, version information and a stable identifier are available, while in another the provenance of the material has to be reconstructed from additional indicators.

Research on the use of AI in journalism focuses mainly on the automation of particular editorial operations and on responsibility for the final result. O. I. Yaroshenko describes the potential of such systems for analysing documents and public data, fact-checking, translation and technical processing of media content (Yaroshenko, 2024). O. Nadtochii and D. Prudius draw attention to

Ukrainian experience with AI and connect its use with ethical assessment and the control of accuracy (Nadtochii & Prudius, 2024). A. M. Dijkstra and co-authors consider the quality of science journalism through integrity, rigour, source transparency, accessibility and social relevance. Journalists interviewed in their study see value in AI for routine tasks, but they also point to risks for rigour and transparency in working with sources (Dijkstra et al., 2024).

The problem posed by generative models extends far beyond fabricated references. A system may shift emphasis, simplify a causal relationship, reproduce hidden bias or present a cautious scientific conclusion as almost definitive. X. Fang and co-authors identified bias in news content produced by large language models (Fang et al., 2024). S. Farquhar and co-authors examined false answers through measures of semantic uncertainty and demonstrated the systemic nature of this problem (Farquhar et al., 2024). In science communication, generative systems can make complex material easier to understand, yet the same simplification mechanism can increase the risk of misinformation when a newsroom loses the connection with the primary source (Alvarez et al., 2024; Biyela et al., 2024).

Existing verification frameworks already cover many operations reflected in Levels 1–4. SIFT focuses on investigating the source, finding better coverage and tracing claims to their original context (Caulfield & Wineburg, 2023). The Verification Handbook addresses digital-source provenance (Silverman, 2020), while the Trust Project indicators emphasise references, methods, expertise and corrections (The Trust Project, 2026). The proposed model therefore does not claim novelty for general source checking. Its added contribution is the integration of publication-status and version checks, selected FAIR elements for data traceability, and a separate AI audit requiring machine-generated claims and citations to be traced to verifiable scientific sources.

The research gap concerns integration rather than the absence of verification procedures. Open-science studies focus on access, infrastructure and data management, while AI research emphasises automation, ethics and trust. General verification frameworks support source checking, corroboration and transparency, but do not combine the publication lifecycle of scientific outputs, FAIR-oriented data traceability and generative-AI provenance risks in one sequence. The proposed model addresses that combined newsroom situation.

MATERIALS AND METHODS

The study is theoretical and analytical and is based on scholarly publications, regulatory documents and methodological materials on open science, research data, FAIR, digital journalism and generative artificial intelligence. The source base includes Ukrainian and international studies, as well as documents of UNESCO, the European Commission and the EOSC Association.

The study used literature analysis, comparison, systematisation and theoretical generalisation. These methods were applied to compare peer-reviewed publications, preprints, datasets and AI-generated materials, as well as to identify risks related to source provenance, publication status, versioning, metadata and interpretation. The proposed model was also compared with SIFT, the Verification Handbook and the Trust Project indicators to distinguish general verification procedures from elements specific to scientific information, FAIR and AI use.

Theoretical generalisation was used to develop the editorial verification model, which combines source identification, publication-status assessment, evidence analysis, data traceability and AI control. Two Ukrainian science news reports were examined to illustrate its practical application at Levels 1 and 2. These examples are illustrative and do not constitute an empirical validation of the model.

RESULTS AND DISCUSSION

Open science allows journalists to work with scientific material in considerably greater depth than a press release or short summary permits. The full text of an article provides access to methods,

limitations and explanations of the results. A repository record adds information about the author, date, version and conditions of use, while an open dataset shows the data on which a conclusion is based. A persistent identifier helps distinguish a particular object from similar records and trace its history. UNESCO links the development of open science with transparency, reproducibility and the public use of knowledge (UNESCO, 2023). For a newsroom, these features provide a practical way to verify the provenance of a claim before publication.

Open access to a material should be distinguished from an assessment of its quality. A preprint may have a DOI, a detailed methodological description and a full text even while external peer review is still in progress. A well-prepared dataset may also contain sampling or data-collection limitations that substantially affect the conclusion. Even a peer-reviewed article may later receive a correction, a new version or a retraction notice. In editorial work, it is useful to distinguish accessibility, traceability and evidentiary reliability. Accessibility shows whether the material itself can be obtained, traceability makes it possible to establish its provenance and version, and evidentiary reliability concerns the relationship between methods, data and the stated conclusion.

The type of source also determines the verification procedure. For a peer-reviewed article, the editor should pay attention to the journal, DOI, publication date, current status and any corrections. For a preprint, the key issues are the version, peer-review status and the existence of a later journal publication. When working with an open dataset, attention shifts to metadata, provenance, the collection method, update date and licence. Material generated with generative AI can be used for search, translation or preliminary structuring, but every substantive factual claim must ultimately be traced back to a specific primary source. These criteria are summarised in Table 1.

The FAIR principles can be used as a quick reference point during the initial assessment of research data. Findable refers to the ability to identify an object unambiguously through metadata or a persistent identifier. Accessible describes the method and conditions of access. Interoperable indicates whether the format and structure of the data allow them to be compared with other datasets. Reusable concerns the completeness of the description, provenance and licensing conditions. Ukrainian methodological recommendations also require the location where research data are stored, a unique identifier and conditions of use to be specified (Ministry of Education and Science of Ukraine, 2024).

Table 1

Types of scientific information and key features of editorial verification

Object type	What is checked first	Main risk	Minimum editorial action
Peer-reviewed article	DOI, journal, authors, date, version, corrections/ retractions	treating peer review mechanically as a guarantee of truth	read the methods and limitations and check the current publication status
Preprint	server, authors, date, version number, existence of a later journal publication	presenting a preliminary result as an established scientific fact	state clearly that the material has not yet been peer reviewed and check later versions
Open dataset	owner, metadata, DOI/ PID, collection method, licence, update date	misinterpreting the structure or limits of the data	compare the data with the methodology and the primary publication
AI-generated summary	titles of primary sources, accuracy of citations, completeness of context	fabricated sources, hallucinations, overconfident simplification	do not cite AI as evidence, verify every substantive claim against the primary source

Source: *developed by the author based on European Commission (2026), Fleerackers et al. (2024), Walters and Wilder (2023), and Dijkstra et al. (2024).*

For a newsroom, these characteristics have a very practical meaning. A stable identifier helps confirm that the researcher and the journalist are referring to the same digital object. Metadata explain the meaning of variables, units of measurement, the date of collection and the provenance of the information. A description of methods and licensing terms becomes important when a newsroom plans to recalculate, compare or visualise the data independently. The more fully a dataset is documented, the easier it is to repeat individual steps and understand the limits of its use. The adaptation of FAIR to journalistic practice is presented in Table 2.

FAIR primarily describes the organisation of a digital object, whereas assessment of scientific quality requires a separate analysis of methodology. Sample representativeness, the appropriateness of the statistical model and the justification of conclusions remain substantive questions. In the proposed approach, FAIR therefore occupies an intermediate position between source identification and analysis of the evidence base. For journalists, this framework is useful because it helps establish the provenance of a material, the conditions of access and the possibility of its appropriate reuse.

FAIR has clear limits in this adaptation. It was developed for research-data stewardship and machine-actionable infrastructure, not as a newsroom quality standard (GO FAIR Foundation, 2026). A DOI/PID and basic metadata support findability and traceability but do not prove full FAIR compliance, methodological quality or the truth of a scientific claim. Table 2 therefore uses selected FAIR elements only as editorial prompts, while the broader EOSC framework remains more comprehensive (EOSC Association, 2024).

The speed of editorial work has a particularly visible influence on the use of preprints. Scientific material may become openly available long before the conventional publication cycle is complete, and the media can access it almost immediately. For topics of high public interest, this creates a strong incentive to publish findings as quickly as possible. Research on journalists' work with open, non-peer-reviewed materials shows that assessing such sources often requires additional expertise (Fleerackers et al., 2024). Editors should therefore consider where the material was posted, which version they are viewing and what happened to it later. This is especially important when a newsroom returns to a topic after several weeks or months and needs to determine whether the initial conclusion remained unchanged after peer review.

Table 2

Adaptation of the FAIR principles to the needs of digital journalism

FAIR principle	Meaning for data	Journalistic question	Practical indicator
Findable	the data can be found unambiguously	Can I identify the exact dataset or version cited by the author?	DOI/PID, complete metadata, repository name
Accessible	the method and conditions of access are specified	Do I have access to the evidence base, and are the restrictions clear?	open file or clearly described access regime
Interoperable	the data are provided in a form suitable for interaction	Can I correctly compare these data with other datasets?	standardised formats, variable dictionaries, structured metadata
Reusable	the data are sufficiently described for reuse	Is it clear how, when and by whom the data were obtained and what may be done with them?	methodology, provenance, licence, version history

Source: *developed by the author based on EOSC Association (2024), Ministry of Education and Science of Ukraine (2024), and Rachynska (2025).*

Note. *Table 2 is a simplified editorial adaptation. DOI/PID and metadata support FAIR-oriented traceability but do not demonstrate scientific quality or full FAIR compliance.*

Before using a preprint, it is advisable to check the date of first posting, the number of the current version, information about peer review or journal publication, changes between versions and the response of the professional community. If a peer-reviewed version later appears, the newsroom should work with that version and compare whether the key conclusions changed. A substantial change also matters for the audience, especially when the preliminary result has already circulated widely. This approach allows the life cycle of a scientific claim to be checked rather than evaluating only a single file at a particular moment. It also helps journalists avoid repeating a conclusion from an early version after the authors have already revised it in response to reviewers' comments.

Versioning becomes even more important when generative AI is used. A system may summarise an older fragment that entered an index earlier, even though the authors later issued a correction or published a revised version. An AI response should therefore be treated as an initial navigation point. If the system names a study, the journalist should verify the DOI, authors, year, title and current status. If a numerical value appears in the response, the editor should return to the article or dataset and locate the table, figure or calculation from which the number was taken.

Generative AI is not treated as a single technical category. A standalone chatbot without live source access carries higher provenance risk because current status and citations may come from static model knowledge. A search or retrieval-augmented assistant can ground answers in retrieved documents and expose links, but retrieval omissions and source-summary mismatches remain possible. Specialised scientific assistants linked to scholarly indexes may improve source resolution, yet publication status, version and evidence still require checking. Retrieval-augmented generation can reduce some limits of static LLM knowledge without removing the need for verification (Han et al., 2024).

The level of risk associated with generative AI depends on the role the system plays in the workflow. Translation, transcription, preliminary document structuring or generating a list of search terms have a relatively limited effect on substantive content. The situation changes when AI selects scientific sources, generates a bibliography, interprets statistical results or establishes causal relationships. Such uses require much more careful verification. European guidelines emphasise that responsibility for the result remains with the human user and that references and summaries produced by a system must be checked (European Commission, 2026). For newsrooms, this principle means clearly distinguishing the auxiliary role of AI from its substantive role.

The reason for such caution is clearly visible in empirical research. W. H. Walters and E. I. Wilder found both entirely fabricated bibliographic records and errors in genuine references (Walters & Wilder, 2023). S. Farquhar and co-authors showed that a language model can produce a false answer in convincing language, which means that assessing its reliability requires additional mechanisms (Farquhar et al., 2024). Science communication creates an additional difficulty because specialised content often has to be explained in accessible language. During this simplification, statistical uncertainty, the limits of applicability of a result or the distinction between correlation and causation can easily be lost.

The quality of a science story depends to a large extent on how transparently a newsroom works with sources and context. A. M. Dijkstra and co-authors identify integrity, rigour, clear source attribution and accessibility for the audience among the key characteristics of quality science journalism (Dijkstra et al., 2024). Generative AI can help make complex material easier to understand, but the editor still needs to know where every substantive claim comes from. Traceability is what makes it possible to distinguish a convenient machine-generated summary from information that can be defended before an editor and the reader.

In practice, it is useful to distinguish the degree to which a generative system intervenes in the content. When AI helps formulate a search query, explains a term, translates a passage or organises known text, verification is largely a matter of checking the result against the original. A more

complex level begins when the system itself compares studies, proposes a conclusion or interprets statistics. In such cases, the journalist should return to primary materials, methods and data and, where necessary, consult an independent specialist. A single general disclosure that AI was used therefore provides little information. For editorial policy, it is more useful to record the specific function performed by the system.

The analysis made it possible to bring the main verification steps together into a single sequence. The process begins by establishing the author, place of publication or storage, date, version, DOI or another persistent identifier. The editor then identifies the institution or repository responsible for the record. Standardised metadata make this stage considerably easier because they allow the material to be linked unambiguously to a specific digital object. Ukrainian recommendations on research-data management likewise treat a persistent identifier as an important means of identifying such an object (Ministry of Education and Science of Ukraine, 2024).

The next step concerns the scientific status of the material. A peer-reviewed article, preprint, conference paper, report, protocol and dataset pass through different procedures before publication, so the wording used by a newsroom should also differ. The journalist checks whether a newer version has appeared, whether peer review has been completed and whether corrections or a retraction have been issued. This information determines the level of certainty with which the result can be presented to the audience.

Level 2 should use scholarly-infrastructure tools rather than the publication page alone. Crossref Crossmark exposes corrections, retractions and updates (Crossref, 2020), and the Retraction Watch Database is searchable through Crossref (Crossref, 2025). PubPeer can reveal post-publication discussion, although comments are signals for further examination rather than scientific verdicts (PubPeer Foundation, 2026). DataCite and OpenAlex help resolve research objects and metadata, ORCID disambiguates researchers, and ROR identifies research organisations.

Once the status has been established, attention shifts to the evidence base. This requires reading the description of the sample and methods, the main tables and figures, and the limitations identified by the authors themselves. If open data are available, the editor checks their connection with the publication, the structure of the variables, the collection date and the units of measurement. FAIR helps assess how readily these data can be found and reused. The substantive question remains central: does the headline claim actually follow from what was measured? In science news, exaggeration often occurs at precisely this stage, when a local finding is presented as a universal conclusion or a statistical association is described in causal language.

An individual result acquires meaning only within a broader scientific context. A newsroom should therefore turn to other studies, systematic reviews, positions of professional organisations and comments from specialists who are independent of the authors of the original work. This helps determine whether a new finding is consistent with the accumulated evidence or represents an isolated result. Expert commentary complements the primary source and helps define the limits of its interpretation more precisely. For topics with major public consequences, it is also useful to check whether professional guidelines or consensus documents exist that can provide a wider context for the new study.

The final element of the model concerns the use of generative AI. Bibliographic information suggested by a system should be checked against scholarly databases and publishers' websites. Numerical values should be verified in the article or dataset, translations of important passages compared with the original, and machine-generated summaries checked against the study's methods, results and limitations. If AI has substantially influenced the analysis or the text, a newsroom should have a transparent rule governing such use. This approach corresponds to the principles of human responsibility and transparency set out in the European Commission guidelines (European Commission, 2026). The structure of the model is presented in Table 3.

Table 3

Five-level model for verifying scientific information in digital journalism

Level	Main question	Key actions	Condition for moving forward	Action if criterion is not met
1. Identification	Does the source exist and has its provenance been established?	DOI/PID, authors, institution, repository, date, version	the source is unambiguously identified	locate the primary record. If it remains unresolved, attribute the claim only to the secondary source or omit it
2. Status	What scientific status does the material have?	peer review, preprint, version history, correction/retraction, Crossmark, Retraction Watch, PubPeer	the status is clear and will be described correctly in the story	pause or qualify the claim, label the status explicitly and search for updates or a later reviewed version
3. Evidence	Do the methods and data support the stated conclusion?	sample, method, main results, limitations, dataset	the claim stays within the limits of the evidence base	narrow or rewrite the claim, remove unsupported figures or causal wording, consult a specialist if needed
4. Cross-checking	How does the result relate to other evidence?	other studies, reviews, independent experts	the context and degree of scientific certainty are established	seek additional studies or an external expert, state uncertainty, or postpone the stronger conclusion
5. AI audit	Has the AI system introduced any unverified elements?	check citations, figures, translations and summaries, record the role of AI	every substantive fact has an independently verifiable primary source	remove the AI-derived element or replace it with information verified directly against the primary source

Source: *developed by the author.*

The model uses an explicit stop rule. Failure at a level stops the unqualified claim, not necessarily the entire story. Under a minimum deadline mode, the newsroom completes source identification, status and evidence checks, one external corroboration and AI tracing when AI contributed factual content. If a criterion fails, the editor finds the missing evidence, narrows the wording, consults a specialist or states the verification limit. The extended mode adds structured status searches, version comparison, deeper review of methods and data, multiple sources and expert consultation. A core claim without a traceable source or adequate evidence is removed or publication is postponed.

Two Ukrainian media cases illustrate the stop rules. A ZN.UA item on conversational swarm intelligence linked readers to Phys.org, while the underlying conference paper with its DOI was not linked (Duman, 2025; Rosenberg et al., 2025). This is a Level 1 issue. The minimum mode requires locating the primary paper before using its numerical results as scientific evidence. Otherwise, the claim remains attributed to secondary reporting or is omitted. A ZN.UA item on LLM-based cryptography linked to arXiv but did not explicitly label the work as a preprint awaiting peer review (Moskalenko, 2025; Gligoroski et al., 2025). This is a Level 2 issue. The minimum action is to label the preprint and moderate certainty. Extended verification checks later reviewed versions, corrections or post-publication discussion. These cases illustrate application rather than empirical effectiveness.

The proposed model takes into account features of the scientific cycle that ordinary fact-checking covers only partly. A newsroom needs to establish that a source actually exists, but this alone is insufficient for accurate reporting on science. It is also necessary to know the publication status,

version history, availability of evidence and the limits within which the authors draw their conclusions. At the same time, journalists need only verify those elements that directly affect the media claim, rather than repeat the experiment or conduct a full statistical audit. Their task is to ensure that the level of certainty in the media text corresponds to the level of evidence in the study itself. This approach preserves the boundary between journalistic and scientific expertise while making editorial decisions more transparent.

For Ukrainian journalism, such a model is particularly useful because data journalism is developing alongside an unevenly mature open-science infrastructure. A. Mykolaienko shows that Ukrainian media actively use state registers and open data in investigations and other socially significant reporting (Mykolaienko, 2025). Scientific data require different preparation because research design, variables, statistical uncertainty, ethical limitations and scientific context are especially important. Experience with administrative open data already provides a strong foundation that can be supplemented with skills in scientific verification.

Methodological recommendations on research-data management are gradually forming a useful regulatory basis for external users as well. The document states that information about access to data should accompany publications and recommends that datasets be deposited in institutional, national, disciplinary or interdisciplinary repositories. It also emphasises the importance of open formats, persistent identifiers and conditions of use (Ministry of Education and Science of Ukraine, 2024). For journalists, this provides a clear reference point: a well-documented scientific publication should make it possible to establish where the data are stored and under what conditions they can be accessed.

In practice, the level of such documentation differs across Ukrainian institutions. O. Rachynska found that repositories are more common than formalised data-management policies and FAIR-related documents, while the use of persistent identifiers remains uneven across some repositories (Rachynska, 2025). The absence of an open dataset therefore says little by itself about the quality of a study. In such a situation, a journalist can verify the publication, methodology and authors, consult an external specialist and clearly indicate the limits of the verification that was possible.

The National Plan for Open Science creates conditions for a gradual alignment of these practices through the development of open access and research infrastructure and closer convergence with European approaches (Cabinet of Ministers of Ukraine, 2022). For the media, this has a direct practical effect. Repositories, standardised metadata and open datasets create more opportunities to verify how a scientific result moves into the public sphere. In such a system, quality science journalism becomes one of the users of open-science infrastructure.

In editorial practice, the proposed approach can serve as the basis for an internal checklist for science stories. Where possible, an online publication should link to the primary article, preprint or dataset. The status of the source should be established before the headline and lead are written, because this is where wording often conveys an unjustifiably high degree of certainty. When a newsroom works directly with open data, it is useful to retain information about filtering, cleaning, table merges and the formulas used. Such documentation simplifies editorial review and makes it possible to reconstruct the logic of a calculation.

Newsrooms also need specific rules for generative AI that reflect real scenarios of use. Such tools are already entering different stages of news production (Cools & Diakopoulos, 2026). Transcription, translation and text structuring involve a lower level of substantive risk than source selection, interpretation of statistics, assessment of causal relationships or generation of quotations. The more strongly the system influences content, the more thoroughly the result should be checked against primary sources.

AI should also be used alongside several independent sources. A generalised text produced by a language model can easily be mistaken for a ready-made review, even though it is generated

differently from a systematic search of the scientific literature. I. Klein-Avraham and co-authors emphasise the importance of evaluating the content of the response, the characteristics of the technology itself, the way it accesses sources and the possibility of tracing the provenance of information (Klein-Avraham et al., 2024). In this sense, AI-assisted search is useful as an auxiliary navigation channel alongside scholarly databases and repositories. When an editor sees several independent primary sources, it becomes easier to understand the limits of generalisation and less tempting to rely on a single machine interpretation.

The model is designed primarily for digital scientific materials that are available for editorial verification. Closed commercial databases, confidential datasets and studies with restricted access require different procedures. Complex statistical methods may also require consultation with a specialist. The practical value of the proposed approach lies elsewhere: it helps organise the path from a scientific source to a journalistic claim and keeps that path clear enough to be verified again.

CONCLUSIONS

The development of open science has given digital journalism broader access to primary materials and research data. At the same time, the logic of verification itself has changed. Accurate reporting requires the provenance of a source, its version, scientific status and connection with the evidence base to be established. At this stage, the FAIR principles are useful as a guide to finding, identifying and reusing data, while methodological assessment defines the limits of the journalistic conclusion.

Generative AI expands the possibilities for working with large volumes of text, translation and search, but it also brings part of the risk directly into the editorial process. The most sensitive issues remain bibliographic errors, loss of context through simplification and inaccurate interpretation of results. A machine-generated response should therefore lead the journalist back to the primary source and serve as the beginning of verification. The greater the system's influence on the content, the more important manual checking of facts, figures and citations becomes. For a newsroom, this can be expressed as a simple practical rule: every claim with substantial factual weight should have a clear and verified provenance.

The article develops a verification model that combines source identification, assessment of publication status, analysis of the evidence base, external cross-checking and oversight of the role of AI. This sequence takes account of versioning in scientific materials and the specific features of preprints and open data. It can be used as the basis of an editorial protocol for fact-checking, data journalism and reporting on science.

The model is a normative conceptual proposal derived from literature comparison and illustrative cases, not an empirically validated editorial protocol. The two media examples demonstrate how its decision rules operate but do not establish effectiveness, reliability or error reduction. These properties require separate newsroom testing.

For Ukrainian media, the practical importance of this approach grows alongside the development of repositories, open-access policies and research-data management practices. A subsequent stage of research should examine real Ukrainian science stories and test the model in editorial work. This would make it possible to identify which verification steps are most often omitted and which criteria actually influence the accuracy of media reporting.

REFERENCES

Alvarez, A., Caliskan, A., Crockett, M. J., Ho, S. S., Messeri, L., & West, J. (2024). Science communication with generative AI. *Nature Human Behaviour, 8*, 625–627. https://doi.org/10.1038/s41562-024-01846-3

Biyela, S., Dihal, K., Gero, K. I., Ippolito, D., Menczer, F., Schäfer, M. S., & Yokoyama, H. M. (2024). Generative AI and science communication in the physical sciences. *Nature Reviews Physics, 6*, 162–165. https://doi.org/10.1038/s42254-024-00691-7

Cabinet of Ministers of Ukraine. (2022, October 8). *On approval of the National Plan for Open Science* (Order No. 892-r) [in Ukrainian]. Verkhovna Rada of Ukraine. Retrieved August 28, 2026. https://zakon.rada.gov.ua/laws/show/892-2022-%D1%80

Caulfield, M., & Wineburg, S. (2023). *Verified: How to think straight, get duped less, and make better decisions about what to believe online*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226829845.001.0001

Cools, H., & Diakopoulos, N. (2026). Uses of generative AI in the newsroom: Mapping journalists' perceptions of perils and possibilities. *Journalism Practice, 20*(3), 878–896. https://doi.org/10.1080/17512786.2024.2394558

Crossref. (2020). *Crossmark*. Retrieved. https://www.crossref.org/services/crossmark/

Crossref. (2025). *Retraction Watch*. https://www.crossref.org/documentation/retrieve-metadata/retraction-watch/

Dijkstra, A. M., de Jong, A., & Boscolo, M. (2024). Quality of science journalism in the age of artificial intelligence explored with a mixed methodology. *PLOS ONE, 19*(6), e0303367. https://doi.org/10.1371/journal.pone.0303367

Duman, V. (2025, May 2). Developers test conversational swarm intelligence: The platform links people into a "superbrain". *ZN.UA*. [in Ukrainian]. https://zn.ua/ukr/TECHNOLOGIES/rozrobniki-viprobovujut-rozmovnij-rojovij-intelekt.html

EOSC Association. (2024). *Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)* (version 1.3). Zenodo. https://doi.org/10.5281/zenodo.17582648

European Commission. (2026). *Living guidelines on the responsible use of generative AI in research* (3rd ed.). Directorate-General for Research and Innovation. https://research-and-innovation.ec.europa.eu/document/download/2b6cf7e5-36ac-41cb-aab5-0d32050143dc_en

Fang, X., Che, S., Mao, M., Zhang, H., Zhao, M., & Zhao, X. (2024). Bias of AI-generated content: An examination of news produced by large language models. *Scientific Reports, 14*, 5224. https://doi.org/10.1038/s41598-024-55686-2

Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. *Nature, 630*, 625–630. https://doi.org/10.1038/s41586-024-07421-0

Fleerackers, A., Chtena, N., Pinfield, S., Alperin, J. P., Barata, G., Oliveira, M., & Peters, I. (2024). Making science public: A review of journalists' use of Open Access research. *F1000Research, 12*, 512. https://doi.org/10.12688/f1000research.133710.2

Gligoroski, D., Raikwar, M., & Jha, S. K. (2025). An LLM framework for cryptography over chat channels. *ArXiv*, 2504.08871. https://arxiv.org/abs/2504.08871

GO FAIR Foundation. (2026). *The FAIR Guiding Principles*. https://www.gofair.foundation/fair-principles

Guillot, P., Bøgsted, M., & Vesteghem, C. (2023). FAIR sharing of health data: A systematic review of applicable solutions. *Health and Technology, 13*, 869–882. https://doi.org/10.1007/s12553-023-00789-5

Han, B., Susnjak, T., & Mathrani, A. (2024). Automating systematic literature reviews with retrieval-augmented generation: A comprehensive overview. *Applied Sciences, 14*(19), 9103. https://doi.org/10.3390/app14199103

Klein-Avraham, I., Greussing, E., Taddicken, M., Dabran-Zivan, S., Jonas, E., & Baram-Tsabari, A. (2024). How to make sense of generative AI as a science communication researcher? A concep-

tual framework in the context of critical engagement with scientific information. *Journal of Science Communication, 23*(6), A05. https://doi.org/10.22323/2.23060205

Ministry of Education and Science of Ukraine. (2024). *Methodological recommendations on the management of scientific (research) data* [in Ukrainian]. https://mon.gov.ua/static-objects/mon/sites/1/nauka/2024/12/31/metod-rekomendatsiyi-shchodo-upravlinnya-naukovymy-danymy-31-12-2024.pdf

Moskalenko, Y. (2025, April 29). Scientists taught artificial intelligence to encrypt secret messages *ZN.UA*. [in Ukrainian]. https://zn.ua/ukr/TECHNOLOGIES/ucheni-navchili-shtuchnij-intelekt-zashifrovuvati-sekretni-povidomlennja.html

Mykolaienko, A. (2025). Data journalism in the Ukrainian media space: Trends, challenges, prospects. *Ukrainian Information Space, 1*(15), 49–60. [in Ukrainian]. https://doi.org/10.31866/2616-7948.1(15).2025.335344

Nadtochii, O., & Prudius, D. (2024). Media practice of using artificial intelligence technologies in Ukraine *Integrated Communications, 1*(17), 129–138. [in Ukrainian]. https://doi.org/10.28925/2524-2644.2024.1717

PubPeer Foundation. (2026). *Frequently asked questions*. https://pubpeer.com/static/faq

Rachynska, O. (2025). Content analysis of open access and open science policies in universities and research institutions of Ukraine *Open Science and Innovation, 2*, 19–32. [in Ukrainian]. https://doi.org/10.62405/osi.2025.02.02

Rosenberg, L., Schumann, H., Dishop, C., Willcox, G., Woolley, A., & Mani, G. (2025). Large-scale group brainstorming and deliberation using swarm intelligence and generative AI. In *Proceedings of the 27th International Conference on Enterprise Information Systems, 1,* 867–872. https://doi.org/10.5220/0013379800003929

Silverman, C. (Ed.). (2020). *Verification handbook: For disinformation and media manipulation*. European Journalism Centre. https://datajournalism.com/read/handbook/verification-3

The Trust Project. (2026). *The 8 Trust Indicators*. https://thetrustproject.org/trust-indicators/

UNESCO. (2021). *UNESCO Recommendation on Open Science*. https://doi.org/10.54677/MNMH8546

UNESCO. (2023). *Open Science Outlook 1: Status and trends around the world*. https://doi.org/10.54677/GIIC6829

Walters, W. H., & Wilder, E. I. (2023). Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports, 13*, 14045. https://doi.org/10.1038/s41598-023-41032-5

Yaroshenko, O. I. (2024). Artificial intelligence in journalism: The future of media under the influence of new technologies. *Scientific Notes of the Institute of Journalism, 2*(85), 139–156. [in Ukrainian]. https://doi.org/10.17721/2522-1272.2024.85.10

Yaroshenko, T. (2021). Open access, open science, open data: How it was and where we are going (to the 20th anniversary of the Budapest Open Access Initiative). *Ukrainian Journal on Library and Information Science, 8*, 10–26. [in Ukrainian]. https://doi.org/10.31866/2616-7654.8.2021.247582

Yaroshenko, T., Serbin, O., & Yaroshenko, O. (2022). Open science: The role of universities and libraries in current changes in scholarly communication. *Digital Platform: Information Technologies in Sociocultural Sphere, 5*(2), 277–292. [in Ukrainian]. https://doi.org/10.31866/2617-796X.5.2.2022.270132

ВЕРИФІКАЦІЯ НАУКОВОЇ ІНФОРМАЦІЇ В ЦИФРОВІЙ ЖУРНАЛІСТИЦІ В УМОВАХ РОЗВИТКУ ВІДКРИТИХ ДАНИХ І ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ

Анотація. *Робота з науковою інформацією в цифровій журналістиці помітно змінилася після поширення відкритих публікацій, репозитаріїв, препринтів і дослідницьких даних. Журналіст отримав прямий доступ до значно ширшого кола першоджерел, а разом із цим зросла потреба розрізняти статус матеріалів, версії публікацій і рівень їх перевірки. Генеративний штучний інтелект додав ще один рівень складності, оскільки такі системи швидко опрацьовують великі обсяги тексту, перекладають і структурують інформацію, проте здатні створювати помилкові посилання, спрощувати зміст і зміщувати акценти. Мета статті полягає в розробленні моделі верифікації наукової інформації для цифрової журналістики з урахуванням можливостей відкритих даних, принципів FAIR та використання генеративного ШІ. У роботі застосовано аналіз наукової літератури, порівняння, систематизацію й теоретичне узагальнення. Увагу зосереджено на походженні джерела, статусі рецензування, версійності, метаданих, постійних ідентифікаторах і зв'язку між публічним твердженням та його доказовою основою. За результатами аналізу сформовано п'ятирівневу модель редакційної перевірки, яка охоплює ідентифікацію джерела, встановлення статусу матеріалу, аналіз доказів, зовнішнє зіставлення та контроль використання ШІ. Принципи FAIR розглянуто як практичний орієнтир для пошуку, простежуваності та повторного використання даних. Практичне значення роботи пов'язане з можливістю застосовувати запропонований підхід у науковій журналістиці, фактчекінгу, журналістиці даних і редакційних протоколах роботи з дослідницькими матеріалами.*

Ключові слова: *відкрита наука, відкриті дані, цифрова журналістика, наукова інформація, верифікація даних, принципи FAIR, генеративний штучний інтелект, журналістика даних, фактчекінг.*

Декларації

Декларація про персональні та конфіденційні дані.
Дослідження не передбачало використання персональних чи конфіденційних даних. Висновок етичного комітету не передбачено.

Декларація про конфлікт інтересів.
Автор заявляє про відсутність конфлікту інтересів.

Декларація про джерело фінансування.
Автор заявляє про відсутність джерела фінансування дослідження.

Декларація про використання інструментів штучного інтелекту.
Автор заявив про використання інструменту ШІ GPT-4o та DeepL для орфографічної перевірки,стилістичного редагування та деяких моментів з перекладом. Автор перевірив та відредагував результат і несе повну відповідальність за зміст, достовірність і новизну рукопису.

Заява про доступність даних: *набори даних, згенеровані та/або проаналізовані під час поточного дослідження, доступні в авторів за обґрунтованим запитом.*

Відкритий доступ: *стаття розповсюджується на умовах ліцензії Creative Commons «Зазначення авторства» 4.0 Міжнародна (CC BY 4.0):* https://creativecommons.org/licenses/by/4.0/deed.uk

Авторські права: *авторські права на цю статтю зберігаються за автором, а права першої публікації надаються журналу «Відкрита наука та інновації».*

INFORMATION ABOUT THE AUTHOR

Zemlianov Danyil — PhD Student at the at the Department of Media Communications, Educational and Scientific Institute of Sociology and Media Communications, V.N. Karazin Kharkiv National University, Kharkiv, Ukraine; e-mail: dan.z.2000.ua@gmail.com; ORCID: 0009-0008-7643-5648

ІНФОРМАЦІЯ ПРО АВТОРА

Землянов Даниїл — аспірант кафедри медіакомунікацій, ННІ соціології та медіакомунікацій, Харківський національний університет імені В. Н. Каразіна, Харків, Україна; e-mail: dan.z.2000.ua@gmail.com; ORCID: 0009-0008-7643-5648

Надійшла до редакції 25.08.2026
Прийнята до друку 25.10.2026
Дата публікації 26.11.2026