[un]reliable science

Policy Brief

Governing the Afterlife of Retracted Science

Retracted Scientific Articles Must Not Continue to Shape Public and Policy Discourse

KEY MESSAGES

- **Retracted scientific articles continue to be cited in media, policy documents, and digital information systems.**
- **Retractions correct the scholarly record but do not prevent the continued societal circulation of unreliable research.**
- **Social media and algorithmic systems amplify the visibility of retracted studies.**
- **Large Language Models (LLMs) systematically fail to recognise retractions and reproduce unreliable evidence.**
- **There is a structural gap between scientific self-correction and public knowledge circulation.**

PROBLEM DEFINITION

Retractions are a cornerstone of scientific self-correction. They are issued when serious errors, methodological flaws, ethical violations, or misconduct that invalidate the reliability of its findings are being discovered in published research. Within the scientific publication system, retractions serve a clear epistemic function: they signal that a study should no longer be used as a trustworthy basis for further research or decision-making.

However, empirical evidence demonstrates a growing disconnect between **scientific correction mechanisms** and **public knowledge circulation**. Although retractions successfully amend the scholarly record, they are ineffective at containing the diffusion of unreliable scientific claims within wider communicative ecosystems.

Retracted articles continue to be cited positively in academic literature, referenced in policy documents, and circulated in mainstream and social media, long after their withdrawal. This persistence is particularly pronounced for high-profile articles that received significant public attention prior to retraction. Even when retractions are formally issued, they are often poorly communicated, inconsistently labelled, or technically invisible outside specialist publishing infrastructures.

The problem is exacerbated by contemporary information dynamics. Recent evidence shows that this gap is further widened by the **increasing use of Large Language Models (LLMs) for information seeking, synthesis, and knowledge mediation**. Empirical testing demonstrates that ChatGPT 4o mini systematically fails to recognise retractions or expressions of concern associated with scientific articles. When asked to assess the quality of retracted studies, the model frequently rates them as “excellent” or “world leading” and does not flag their invalidated status. Moreover, when prompted to evaluate factual claims derived from retracted articles, LLMs affirm such claims as true or partially true in a majority of cases. These claims have been found to apply to one version of ChatGPT and when it is used through its programming interface and does not access the web to lookup information. Other versions or interfaces have not been tested.

[1] **Barbour V, Bloom T, Lin J, et al.:** Amending Published Articles: Time To Rethink Retractions And Corrections? bioRxiv. 2017; 118356

[2] **Thelwall, M., & Lehtisaari, M., Katsirea, I., Holmberg, K., & Zheng, E.-T.** (2025). Does ChatGPT ignore article retractions and other reliability concerns? Learned Publishing. 38(4), e2018. https://doi.org/10.1002/leap.2018

This creates a new layer of risk: unreliable science is not only preserved in archives or legacy media coverage but is actively reintroduced into contemporary discourse through AI-mediated systems that carry an aura of neutrality and authority.

Crucially, this phenomenon does not primarily stem from malicious intent. Many retractions result from honest error rather than fraud, and scientific dissent remains a legitimate and necessary component of epistemic progress. The challenge lies in **the structural inability of current systems to distinguish reliably between valid scientific disagreement and knowledge that has been formally invalidated**. Without such differentiation, retracted research continues to function socially as credible evidence.

EVIDENCE SUMMARY

- **Highly visible retracted articles continue to receive media and social media attention.**
- **Retractions lead only to a limited reduction in positive post-retraction citations.**
- **ChatGPT does not reliably identify retracted articles and frequently rates them as high-quality research.**
- **In two thirds of cases, LLMs affirm claims originating from retracted papers as "likely true" or "consistent with research".**
- **This contributes to the continued dissemination of unreliable science beyond the academic domain.**

POLICY IMPLICATIONS

The continued societal influence of retracted research reveals a systemic weakness at the intersection of science governance, media systems, and digital infrastructures. Retractions operate effectively as an internal corrective mechanism within academic publishing, but they are poorly aligned with the realities of modern information ecosystems.

From a policy perspective, this represents a form of **critical misinformation originating within the scientific field itself**. Unlike external disinformation campaigns, this type of misinformation derives its authority precisely from the norms and credibility of science. Its impact is therefore harder to detect and potentially more damaging, as it undermines trust in evidence-based decision-making while remaining formally anchored in "scientific" sources.

THE PERSISTENCE OF RETRACTED RESEARCH IN PUBLIC AND POLICY DISCOURSE HAS SEVERAL IMPLICATIONS:

First, it challenges the assumption that scientific self-correction automatically safeguards society. **Retractions correct the record, but they do not reverse the societal effects of earlier dissemination**. In policy contexts, where scientific findings are frequently cited to justify regulatory choices or public health interventions, reliance on invalidated research can distort decision-making processes.

Second, it exposes a regulatory blind spot in digital knowledge infrastructures. **AI systems and algorithmic intermediaries increasingly mediate access to scientific knowledge, yet they are not systematically equipped to recognise or signal retracted content**. This allows unreliable science to be continuously reproduced, often stripped of contextual qualifiers or uncertainty.

Third, the issue complicates debates around scientific freedom and dissent. **While minority or heterodox scientific views must be protected, retracted research occupies a different category: it has been formally identified as unreliable within the scientific system itself**. Failure to clearly communicate this distinction risks conflating legitimate dissent with invalidated knowledge, thereby weakening both scientific integrity and freedom of expression frameworks.

Finally, the problem highlights the need to rethink responsibility across the knowledge ecosystem. **Publishers, journalists, platform providers, AI developers, and policymakers all interact with scientific outputs, yet no single actor currently bears responsibility for ensuring that retracted research is appropriately flagged** and contextualised beyond academia.

In sum, the continued circulation of retracted articles is not a marginal issue of scholarly hygiene. It is a governance problem with direct implications for media reliability, AI accountability, and evidence-informed policymaking. Addressing it requires coordinated policy responses that extend the logic of scientific self-correction into the public and digital domains.

The continued circulation of retracted scientific articles poses a concrete challenge for both **EU AI governance** and **science journalism policy**, as it exposes a structural weakness in how scientific reliability is operationalised outside academic publishing.

IMPLICATIONS FOR EU AI GOVERNANCE

Large Language Models and AI-based search systems are increasingly used as intermediaries between scientific knowledge and society. In practice, they function as epistemic authorities: synthesising research, summarising evidence, and answering policy-relevant questions. Empirical evidence shows that current LLMs do not reliably detect retractions, expressions of concern, or serious methodological flaws, and often reproduce invalidated claims as credible knowledge.

This directly intersects with EU objectives under the AI Act and broader AI governance frameworks, which emphasise **trustworthy AI, risk mitigation**, and **systemic responsibility** for high-impact AI systems. The issue of retracted research illustrates that AI-related risks are not limited to hallucination or bias, but include the **uncritical amplification of formally invalidated scientific knowledge**.

From a governance perspective, this represents a failure of provenance and reliability safeguards. AI systems may technically comply with transparency requirements while still disseminating unreliable science, because retraction status is neither consistently machine-readable nor systematically integrated into AI pipelines. Without intervention, this undermines the EU's ambition to ensure that AI systems used in knowledge-intensive domains do not distort evidence-based decision-making.

IMPLICATIONS FOR SCIENCE JOURNALISM POLICY

Science journalism occupies a pivotal role in translating scientific findings for public and policy audiences. However, retractions reveal a structural asymmetry: initial findings receive disproportionate attention, while subsequent corrections or withdrawals rarely achieve comparable visibility.

Empirical studies show that **retracted research continues to be cited positively in media coverage and policy-related discourse**. In particular, studies show that post-retraction updates of news reports are rare. This is not due to journalistic malpractice, but to the absence of institutionalised safeguards that integrate retraction checks into editorial workflows. **Our research found that incidences of new post-retraction mainstream media reporting without acknowledgment of the retraction are not common**, whilst this phenomenon is prevalent in social media.

In the context of an increasingly fragmented media ecosystem and time-pressured news production, journalists rely heavily on secondary sources, summaries, and algorithmically surfaced content. **When retraction signals are weak or absent, invalidated science remains socially operational as evidence**. This poses a reputational risk for science journalism and contributes to public confusion about the reliability of scientific knowledge.

Crucially, this challenge does not undermine the role of journalism as a forum for scientific debate or dissent. Rather, it highlights the need for clearer differentiation between legitimate scientific disagreement and knowledge that has been formally invalidated within the scientific system itself.

UNDERSTANDING SCIENTIFIC KNOWLEDGE PRODUCTION AS A BASIC SKILL IN VET:

Beyond regulatory and technological measures, strengthening societal understanding of how scientific knowledge is produced and validated should be considered an important complementary policy response. Integrating basic knowledge about the scientific communication process—including peer review, publication practices, and the role of retractions as mechanisms of scientific self-correction—into education and training curricula could **improve the public's ability to interpret scientific claims critically**. This is particularly relevant for vocational education and training (VET) programmes that prepare future journalists and media professionals, but it is equally valuable for broader educational contexts.

Such competencies align closely with the European Union's policy framework on basic skills and lifelong learning. The European Commission's Action Plan on Basic Skills emphasises the importance of strengthening foundational competencies such as literacy, science literacy, digital skills, and civic understanding across all stages of education and adult learning . Embedding knowledge about the scientific publication and correction process within these skill sets could help future professionals and citizens better recognise the difference between emerging scientific debate, unreliable claims, and formally retracted research.

[3] **Auste Valinciute & Willem Halffman (24 Sep 2025):** Do Journalists Update Retracted Science News?, Journalism Practice, DOI: 10.1080/17512786.2025.2540460

POLICY RECOMMENDATIONS

- **Mandatory Integration of Retraction Metadata in High-Risk and GPAI Systems**
  Require AI systems used for information retrieval, summarisation, or decision support in science-related domains to integrate authoritative retraction databases (e.g. Crossref / Retraction Watch-linked metadata).
- **Machine-Readable Retraction Standards**
  Support EU-wide standards for structured, persistent, and interoperable retraction signals embedded in publication metadata and DOIs.
- **Reliability-by-Design Requirements**
  Treat failure to detect retracted research as a systemic reliability risk under AI governance frameworks, not merely as an accuracy issue.
- **Auditability of Scientific Knowledge Sources**
  Require providers of GPAI and AI-assisted search systems to document how scientific sources are selected, filtered, and updated when reliability status changes.
- **Clear User Signalling**
  Oblige AI systems to explicitly flag retracted or invalidated research outputs to users, including explanations of the nature of the retraction where available.
- **Institutionalised Retraction Checks in Editorial Workflows**
  Encourage or require media organisations to integrate automated retraction checks into science and health reporting processes.
- **Post-Retraction Responsibility Frameworks**
  Promote editorial guidelines that define responsibilities for updating, contextualising, or correcting coverage when cited research is retracted.
- **Visibility Parity for Corrections**
  Develop best-practice standards ensuring that retractions and corrections receive visibility proportional to the original coverage.
- **Training and Capacity Building**
  Support targeted training for science journalists on interpreting retractions, expressions of concern, and different forms of unreliable science.
- **Collaboration with AI and Platform Providers**
  Foster cooperation between media organisations, AI developers, and research infrastructure providers to ensure consistent handling of retracted research across systems.

SOURCES

Auste Valinciute & Willem Halffman (24 Sep 2025): Do Journalists Update Retracted Science News?, Journalism Practice, DOI: 10.1080/17512786.2025.2540460

Barbour V, Bloom T, Lin J, et al.: Amending Published Articles: Time To Rethink Retractions And Corrections? bioRxiv. 2017; 118356

Thelwall, M., & Lehtisaari, M., Katsirea, I., Holmberg, K., & Zheng, E.-T. (2025). Does ChatGPT ignore article retractions and other reliability concerns? Learned Publishing. 38(4), e2018. https://doi.org/10.1002/leap.2018