

PROMPT

Best Practice Guidelines for Responsible AI and
Disinformation Response in (Transnational) Journalism
Perspectives from the 2nd European Narrative Observatory

Deliverable 8.1



This project has received funding from the European Commission under grant agreement CNECT/LC-02629302.

Table of Contents

1. Introduction: A Transnational Approach to Responsible AI and Journalism.....	5
2. Literature Review: Recurring Principles & Practical Gaps in AI-in-Journalism Guidance: A Literature Review.....	6
2.1. Core Recurring Principles.....	6
2.2. Practical Gaps and Limitations.....	8
2.3. Implications for the Design of Our Best Practice Guidelines.....	9
3. AI in Modern Journalism.....	11
3.1. AI and Quality of Sources: Lessons from Wikipedia.....	11
3.2. AI in a transnational context.....	14
3.2.1. Insights from Interviews with Fact-checkers across Europe.....	16
3.2.2. Case studies: Romanian and Moldovan elections.....	18
3.3. Editorial Decision Checklist for AI Use in Newsrooms.....	19
4. Media Literacy for Media Professionals in the Age of AI and Disinformation.....	21
4.1. MOOC Module on “Narratives, News, and Noise: AI, Journalism and the Integrity of Information”.....	21
4.2. PromptED: Interactive Coaching for Trustworthy AI Use in Media.....	26
5. Further Resources.....	28
6. References.....	30

List of Abbreviations

AI	Artificial Intelligence
AUR	Alianța pentru Unirea Românilor
CJR	Columbia Journalism Review
CoE	Council of Europe
COPE	Covering Cohesion Policy in Europe
DSA	European Union's Digital Services Act
ECHR	European Convention on Human Rights
EU	European Union
MOOC	Massive Open Online Course
NATO	North Atlantic Treaty Organization
NGO	Non-Governmental Organization
PROMPT	Predictive Research On Misinformation & Propagation Trajectories
USR	Uniunea Salvați România
VLOP	Very Large Online Platform

1. Introduction: A Transnational Approach to Responsible AI and Journalism

Disinformation has become one of the defining challenges of our digital era. It spreads across borders with ease, exploiting the same infrastructures that connect us. What begins as a local rumour can quickly evolve into a transnational narrative, amplified by algorithms, and echoed across platforms, networks, and languages. In this environment, journalists and researchers face a moving target: not just falsehoods, but shifting ecosystems of influence that shape how people understand truth, trust, and sometimes even democracy.

The EU-funded [PROMPT](#) project (09/2024-02/2026) was created to respond to this challenge. It brings together journalists, researchers, and (fact-checking) organisations from across Europe to study how disinformation narratives are being formed, how they can be detected, and how information can be coordinated across borders. PROMPT's work bridges practice and research - combining interviews with fact-checkers, case studies of national elections (Romania and Moldova), and the development of AI-tools such as the **Corpus Analyser**, the **Disinfo Scanner**, the **Wikipedia Sensitivity Meter and Barometer**, all designed to make disinformation patterns and networks visible and traceable.

These Best Practice Guidelines grow directly out of that collaborative work. Its aim is not to reinvent the wheel: there are already numerous ethical frameworks and policy recommendations addressing (generative) artificial intelligence (AI) and information integrity (see e.g. [UNESCO 2024](#)). Instead, PROMPT seeks to bring existing principles together, test them in practice, and ground them in the real experiences of journalists, fact-checkers, and media professionals working in diverse European contexts.

Artificial intelligence plays a central role in this conversation. AI tools increasingly shape how information is created, verified, and distributed. They offer powerful opportunities for detection, translation, and analysis - but also introduce new risks of bias, opacity, and fabrication. Within PROMPT, AI is treated not as a replacement for human work and judgment, but as a tool that must be used responsibly, transparently, and under clear editorial oversight.

The chapters that follow reflect this perspective: The following chapter reviews existing ethical and professional guidelines around AI, journalism, and fact-checking - identifying shared principles and areas that need further adaptation, especially for cross-border work. Chapter 3 turns to PROMPT's empirical research, including sections from interviews and case studies that illustrate how these principles apply in practice, from collaborative knowledge spaces like Wikipedia to the information dynamics of national elections in Romania and Moldova.

Taken together, these chapters provide a foundation for journalists and civic actors to engage with disinformation and AI in a principled, practical, and transnational way. They recognise that

the fight against disinformation is not just technical - it is ethical, collaborative, and evolving. PROMPT's contribution lies in connecting these dimensions across borders and languages, ensuring that Europe's diverse media and research communities can learn from one another and act together in safeguarding the integrity of public debates and knowledge.

2. Literature Review: Recurring Principles & Practical Gaps in AI-in-Journalism Guidance: A Literature Review

As the deployment of artificial intelligence (AI) in journalistic workflows accelerates, a significant body of guidance has emerged. Several key themes recur across these documents, even as they often stop short of offering fully operational frameworks for newsrooms. This section first outlines the core recurring principles found in existing guidance, and then identifies the practical gaps that persist.

2.1. Core Recurring Principles

1. Human oversight & editorial decision-making

Many guidelines underscore that the decision to implement or deploy AI systems is fundamentally an editorial decision, not merely a technological or commercial one. For example, the Council of Europe (CoE) Guidelines emphasise that news organisations should designate a “problem owner” responsible for oversight of AI implementation, and that editorial values must drive AI adoption rather than efficiency alone.

<https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>

Similarly, practitioner accounts note that journalists regard AI tools as support, not substitute, to human judgement - for instance, in an interview, Jacqueline Dalton, Head of Editorial at Fondation Hironnelle, asserts that “Journalists are fully responsible for editorial decisions and content validation.”

<https://www.hironnelle.org/en/how-to-reconcile-artificial-intelligence-and-journalistic-work-fondation-hironnelles-directive>

The documentary “*Journalism in the Age of AI*” by Euractiv provides a nuanced examination of how artificial intelligence is reshaping newsroom practices. Through expert interviews and concrete use cases - including automated reporting, real-time translation, content moderation, and audience segmentation - it highlights both the innovative potential of AI and the risks associated with increased automation. The film consistently emphasizes the importance of clear editorial accountability, transparency in AI-supported workflows, and sustained human oversight to ensure journalistic

integrity, trustworthiness, and ethical standards as algorithmic systems become more deeply integrated into news production.

<https://aicode-project.eu/2025/07/22/journalism-in-the-age-of-ai/>

2. Transparency & disclosure of AI use

There is broad agreement that media organisations must disclose to audiences when AI is used, how it is used, and what its limitations are. The CoE guidance calls for “standardised forms of labelling that AI systems were used in the workflow (in natural language and machine-readable code)”.

<https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>

For example, the initiative Content Authenticity with members including Adobe, the BBC, AFP is working on restoring trust and transparency in the age of AI.

<https://contentauthenticity.org/>

Empirical analysis confirms that disclosure remains rare in practice, highlighting the importance of transparency as a principle.

<https://www.cjr.org/analysis/can-ai-tools-meet-journalistic-standards.php>

3. Accuracy, verification & provenance

The application of AI to journalistic tasks carries risks of error, “hallucination”¹, or bias². Guidance emphasises that AI-generated content must be verified against independent sources, and provenance of data/model must be documented. For example, newsrooms using AI in investigative workflows call for multiple sources and human validation of AI output.

<https://www.cjr.org/feature/how-were-using-ai-tech-gina-chua-nicholas-thompson-emilia-david-zach-seward-millie-tran.php>

4. Bias, fairness & data governance

The risks of algorithmic bias and discriminatory outcomes are repeatedly flagged. The CoE Guidelines instruct that media organisations evaluate data quality, fairness, and the risk of stereotypes when acquiring or developing AI systems.

<https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>

Moreover, academic work on AI guidelines in media emphasises fairness, accountability, and the preservation of journalistic values as recurring themes in guideline documents.

<https://doi.org/10.48550/arXiv.2405.04706>

¹ A hallucination occurs when an AI system produces information that is factually incorrect, fabricated, or not supported by its training data, yet presents it with confidence.

² Bias refers to systematic distortions in AI output that stem from imbalances or problematic patterns in the underlying data, model design, or usage context.

5. **Accountability, rights & legal compliance**

The democratic function of journalism, including freedom of expression and pluralism, is central in the governance discourse. The CoE frames AI use in journalism explicitly within the context of human rights (e.g., Articles 8, 10 and 14 of the ECHR) and editorial independence.

<https://dig.watch/resource/guidelines-on-the-responsible-implementation-of-artificial-intelligence-systems-in-journalism>

6. **Training, capacity building & newsroom adaptation**

Guidance emphasises that organisational capacity (skills, awareness, infrastructure) must keep pace with technology. The CoE states that news organisations should provide training programmes bringing together technologists and journalists.

<https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>

Practitioner organisations reflect this: Fondation Hirondelle reports they “must implement training sessions with concrete examples that journalists can relate to... Our goal is to train all journalists in our newsrooms by end of year.”

<https://www.hirondelle.org/en/how-to-reconcile-artificial-intelligence-and-journalistic-work-fondation-hirondelles-directive>

2.2. Practical Gaps and Limitations

While the principles above are widely shared, the literature and practitioner accounts reveal some recurring gaps in how they are operationalised:

1. **Lack of operational checklists and everyday workflows**

Many guidelines remain high-level and abstract. A recent study by the Columbia Journalism Review found that most AI policies in newsrooms “do not address a journalist’s daily workflow”. Yet, conversations with journalists repeatedly highlight that this gap is especially difficult to close: rigid checklists often fail to reflect the dynamic, fast-paced, and situational nature of everyday newsroom routines. As a result, even when guidelines exist, they tend to remain unused in practice.

<https://www.cjr.org/analysis/i-tested-how-well-ai-tools-work-for-journalism.php>

2. **Insufficient public disclosure formats and metadata standards**

Although disclosure is emphasised, there is little standardisation of *what* must be disclosed - e.g., model version, data provenance, prompt engineering, human role. The CoE mentions labelling and machine-readable code but offers limited practical templates.

<https://dig.watch/resource/guidelines-on-the-responsible-implementation-of-artificial-intelligence-systems-in-journalism>

3. **Limited audit frameworks and evaluation metrics specific to journalism**

While technical audit frameworks for AI exist in adjacent fields, few guidelines tailor metrics to journalistic values (accuracy, fairness, audience trust). The literature notes that evaluating AI tools in newsrooms remains under-studied.

<https://www.cjr.org/feature/how-were-using-ai-tech-gina-chua-nicholas-thompson-emilia-david-zach-seward-millie-tran.php>

4. **Accessibility for smaller newsrooms / less-resourced languages**

Most guidance and case studies focus on large, well-resourced news organisations. The CoE acknowledges that “some news organisations, such as well-resourced public service media ... are better positioned” to implement AI.

<https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>

This creates a risk of widening inequalities in the media sector.

https://www.cjr.org/tow_center_reports/artificial-intelligence-in-the-news.php

5. **Vendor/third-party risk & procurement issues**

Reliance on external AI providers can reduce editorial autonomy and increase dependencies (also on the specific training data). Guidance mentions procurement risk but offers limited practical templates for contracts or vendor management.

https://www.cjr.org/tow_center_reports/artificial-intelligence-in-the-news.php

6. **Cross-border harmonisation and policy-maker tools**

At a European level, harmonised standards are still emerging. While the EU’s regulatory framework for AI (such as the Artificial Intelligence Act) is evolving, specific tools for policy-makers addressing journalism would be helpful.

<https://www.consilium.europa.eu/en/policies/artificial-intelligence/#AI%20act>

Together, these gaps mean that despite shared ethical commitments, the *practice* of AI in journalism risks undermining trust, transparency and editorial independence unless supported by robust operational frameworks. For example, if disclosure is inconsistent, audiences cannot assess whether a story was AI-assisted; if audit metrics are missing, it becomes difficult to evaluate whether AI is enhancing or impairing quality; if small newsrooms are excluded, media pluralism may weaken.

In short, the gap between principle and practice limits the potential of AI to contribute positively to journalism’s democratic role - and may even intensify the very risks (bias, automation of journalism, reduced transparency, platform dependence) that guidance seeks to mitigate.

2.3. Implications for the Design of Our Best Practice Guidelines

In developing the Best Practice Guidelines, it is therefore crucial to bridge the gap between high-level principles and newsroom reality. That means including operational tools (checklists, metadata schemas, audit templates), being sensitive to newsroom size and resource constraints, addressing procurement and vendor risk explicitly, and promoting transparency and comparability across outlets and EU member states. The Guidelines should also embed a mechanism for continuous learning - monitoring, evaluation, and revision - to ensure that AI-journalism praxis remains aligned with evolving ethics, technology and democratic demands.

The following table outlines key principles and thematic areas, complemented by examples drawn from existing best-practice guidelines (highlighted in green). The final items (highlighted in orange) are also referenced in current guidelines but are either less consistently defined, unevenly applied, or represent emerging issues that require further clarification, standardization, and critical attention.

Table 1. Recurring Principles in AI-in-Journalism Guidance

Principle / Theme	Source(s)	Example(s) in Practice / Notes
Human oversight & editorial decision-making	CoE Guidelines (2023); Fondation Hirondelle (2025)	Journalists remain fully responsible for editorial decisions; AI tools are assistants, not decision-makers. Designated “problem owner” oversees AI deployment.
Transparency & disclosure of AI use	CoE Guidelines (2023); CJR (2025)	Label AI-assisted content; provide audiences with information on model, dataset, human role. Limited standardization in practice.
Accuracy, verification & provenance	CoE Guidelines (2023); CJR (2025); DisinfoCode (2023)	AI outputs must be verified with independent sources; include model use etc to track the process.
Bias, fairness & data governance	CoE Guidelines (2023); arXiv review (2024)	In a best case scenario, newsrooms would be able to evaluate AI models for algorithmic bias; audit datasets; mitigate stereotyping risks; consider fairness in news coverage.
Accountability & legal/ethical compliance	CoE Guidelines (2023); EU AI Act (2023)	Compliance with human rights; there are currently court processes awaiting decision on how to handle copyright; in any case, editorial independence needs to be maintained.

Training & capacity building	CoE Guidelines (2023); Fondation Hironnelle (2025); Pulitzer Center (2025)	Staff training in AI tools, ethics, verification; workshops for journalists and editors.
Operational workflow & checklists (gap)	CJR (2025); DisinfoCode (2023)	Lack of concrete daily workflow checklists in most newsrooms; operational guidance is minimal.
Disclosure formats & metadata standards (gap)	CoE Guidelines (2023)	Standardized schema for AI-assisted content largely missing; practical templates needed.
Audit frameworks & evaluation metrics (gap)	CJR (2025); DisinfoCode (2023)	Metrics for accuracy, bias, audience impact are underdeveloped. No standardized newsroom audit protocols.
Accessibility for smaller newsrooms / languages (gap)	CoE Guidelines (2023)	Tools and guidelines mainly focus on large, well-resourced newsrooms; risk of excluding smaller outlets.
Vendor / procurement & risk management (gap)	CJR (2025); CoE Guidelines (2023)	Limited guidance on contracts, IP, vendor risk, external AI dependencies.
Cross-border harmonisation & policy tools (gap)	EU AI Act (2023)	Lack of unified EU-wide operational standards for journalism; guidance mostly national or abstract.

3. AI in Modern Journalism

The growing presence of artificial intelligence in journalism is transforming how news is researched, verified, and distributed. Recent studies note that AI tools are increasingly used in areas such as content recommendation, data analysis, translation, and automated transcription, yet their adoption remains uneven across media systems (Diakopoulos & Koliska, 2017; Cools & Koliska, 2024). While large, well-resourced news organisations experiment with

algorithmic support for editorial tasks, smaller outlets often lack the technical capacity or funding to do so (Council of Europe, 2023).

This chapter examines how AI interacts with journalistic quality, verification practices, and cross-border collaboration. It begins by analysing Wikipedia as a case study for understanding source integrity in the digital knowledge commons (McDowell & Vetter, 2020; Saez-Trumper, 2019), then explores findings from interviews with European journalists and fact-checkers conducted within the PROMPT project, and finally presents aspects to ensure responsible and transparent AI use in newsrooms.³ Together, these sections connect normative principles to the everyday realities of journalism in Europe's diverse media landscape.

3.1. AI and Quality of Sources: Lessons from Wikipedia

As artificial intelligence becomes increasingly integrated into journalistic workflows, the quality, reliability, and provenance of information sources emerge as critical concerns. AI-driven systems depend heavily on large, publicly available datasets, which makes the governance of such sources a matter of direct relevance for journalism. Wikipedia plays a central role in this ecosystem - not as a primary source for journalistic verification, but as a widely reused reference point within search engines, data pipelines, and AI training corpora.

This section therefore does **not** propose Wikipedia as a model source for fact-checking. Rather, it examines Wikipedia as a case study for understanding how open, collaborative information infrastructures shape - and sometimes distort - the information environments in which journalistic and AI systems operate.

Wikipedia as a Digital Knowledge Commons

Wikipedia represents one of the most influential and widely accessed information repositories worldwide (McDowell & Vetter, 2020; Saez-Trumper, 2019). Its radical openness, community-driven governance, and commitment to verifiability have enabled it to function as a global digital knowledge commons. At the same time, these characteristics also introduce structural vulnerabilities that are highly relevant for journalism in the age of AI.

While Wikipedia itself is rarely considered a legitimate primary source for professional fact-checking, its prominence in search results and its extensive reuse in downstream systems mean that errors, biases, or manipulations can have far-reaching consequences. When misinformation enters Wikipedia, even incrementally, it can be amplified across the broader information ecosystem - including journalistic research practices and algorithmic content generation.

Disinformation Dynamics and Structural Vulnerabilities

³ More details on the interviews in the PROMPT's Second Narrative Report: <https://disinfo-prompt.eu/posts/3pioVYvnxak9OkJT1V8byu>

Research and investigative reporting show that disinformation on Wikipedia often emerges through gradual, long-term editing patterns. Individual edits may appear neutral or procedurally valid, while their cumulative effect subtly reshapes narratives. Articles in smaller language editions or on politically sensitive topics are particularly exposed due to limited editorial capacity and reduced community oversight.

A well-documented example is the Croatian Wikipedia case. Over more than a decade, ideologically motivated administrators and editors systematically altered historical and political content to reflect radical-right perspectives, producing articles that combined encyclopaedic tone with historical revisionism (“Croatian Wikipedia Disinformation Assessment,” 2021). An external assessment commissioned by the Wikimedia Foundation confirmed sustained bias and abuse of administrative power, demonstrating how decentralised governance structures can be exploited by persistent actors (Wikimedia, n.y.). While the Croatian Wikipedia is somewhat dated, it remains a valuable and informative resource. At the same time, findings from the [PROMPT Narrative Report 2](#) chapter on the Moldovan elections in 2025 highlight relevant risks that also apply in this context. As the report points out, certain Wikipedia pages have become targets of historical revisionism, with content based on missing, weak, or problematic sources. This underlines the importance of critically assessing Wikipedia entries, particularly on politically or historically sensitive topics.

These dynamics are especially relevant for journalism because Wikipedia content frequently informs background research, contextual framing, and - increasingly - AI-assisted summarisation and retrieval systems. The case highlights the risk of treating widely available digital sources as neutral or stable simply because they are familiar or institutionally embedded.

Regulatory Context: Wikipedia as a VLOP under the DSA

Wikipedia’s designation as a Very Large Online Platform (VLOP) under the European Union’s Digital Services Act (DSA) further underscores its systemic relevance. While the DSA establishes a framework for platform accountability and risk mitigation, its application to a volunteer-driven, consensus-based project raises distinct challenges.

As the Wikimedia Foundation notes:

“The EU DSA imposes a wide range of obligations on platforms, search engines, and app stores of all sizes. All Wikimedia projects are covered by these. The handful of services that get designated as VLOPs - in our case, only Wikipedia - face supplementary obligations and shorter implementation deadlines.” (Bradley-Schmiege, 2023)

For journalism, this development illustrates a broader shift from informal, community-based accountability toward hybrid models of governance that combine legal regulation, institutional oversight, and civic participation - a trajectory increasingly relevant for AI governance in newsrooms.

Looking at the [DSA SRAM for 2025](#), elections (“Disinformation regarding civic and electoral processes, and conflicts”) and historical revisionism (“Disinformation regarding historical/geographical narratives”) are highlighted by Wikimedia as high-salience systemic risks, particularly because of their impact on information integrity, the reliability of sources, and public trust during politically sensitive periods.

Empirical Tools from the PROMPT Project

In addition to [existing tools developed within or with the Wikimedia community](#), The PROMPT project, in collaboration with Wikimedia France, directly addresses these challenges by operationalising transparency and accountability through analytical tools designed for journalists, researchers, and fact-checkers.

1. **PageChecker Tool** (<http://37.59.112.214:8300/>)

This tool enables users to analyse Wikipedia pages for signs of manipulation, bias, or inconsistency by examining edit histories, contributor patterns, and source networks. It supports journalists in detecting early indicators of coordinated editing campaigns and serves as a practical model of human-AI collaboration in verification.

2. **WikiSensiBarometer** (<https://wikisensibarometer.disinfo-prompt.eu/>)

The WikiSensiBarometer uses existing Wikimedia portals or large customized datasets of thematic pages. Through 16 metrics measuring heat, quality & user behaviour on Wikipedia, it provides shortcuts to the Wikimedian community to identify suspicious content/anomalies/accidents.

Together, these tools illustrate how AI-assisted transparency mechanisms can strengthen journalistic integrity and mitigate vulnerabilities within open knowledge environments. They represent a key step toward the co-regulation of digital commons, where community norms, technical infrastructure, and professional journalism intersect.

Implications for Journalism and AI Governance

From a journalistic perspective, Wikipedia should be approached with caution as a source - but with analytical interest as an infrastructure. Its hybrid governance model offers transferable lessons for AI-supported journalism, including:

- **Transparency and traceability:** All edits are logged, timestamped, and publicly accessible, providing audit trails that mirror best practices for AI-assisted editorial workflows.
- **Participatory oversight:** Distributed review mechanisms demonstrate how plural forms of oversight can complement professional editorial control.
- **Resilience through redundancy:** The combination of human review, technical signals, and community debate highlights the value of layered safeguards against disinformation.

Positioned at the end of this chapter, Wikipedia serves not as a recommended evidentiary source, but as a reflective case study. It functions as a living laboratory for understanding how information quality, governance structures, and AI systems interact - offering insights that can inform more robust, transparent, and accountable journalistic practices in automated information environments

Ultimately, the PROMPT project's integration of Wikipedia-based monitoring tools demonstrates how journalistic innovation can enhance both platform accountability and media resilience.

3.2. AI in a transnational context

The diffusion of disinformation across national boundaries challenges both traditional journalism and emerging AI-supported detection strategies, particularly in multilingual environments. Differences in political, linguistic, and media-systemic contexts shape how disinformation emerges, circulates, and is addressed. But **they also influence the availability and effectiveness of technical responses.**

As Unger et al. (2025) demonstrate, disinformation dynamics are highly contextual, shaped by political, linguistic, and media-systemic conditions within individual countries. Their expert interviews with German political, governmental, media, and business elites reveal that while definitions of disinformation converge around intentionality and harm, perceptions of its key features and actors differ widely across sectors and institutional contexts. These divergences underscore how fragmented national approaches can hinder coherent, transnational counter-strategies - a challenge particularly acute in multilingual Europe, where digital information flows disregard borders (Unger et al., 2025).

Conversely, current automatic detection of disinformation for what AI scientists call "low-resource languages" (e.g. Estonian, Swahili, etc.) face limitations, such as the lack of available and high-quality training datasets.

These conditions point to the need for tools and approaches that can operate across linguistic and national boundaries without assuming uniform resource availability.

National fragmentation has practical implications for the responsible implementation of AI in journalism. Automated content analysis, network detection, and linguistic models often reflect the biases of their training data, national context and language coverage. For example, smaller linguistic communities remain underrepresented in major AI training corpora, which reduces the sensitivity of automated systems to local narratives and disinformation cues. As a result, fact-checkers and journalists working in languages beyond English face disproportionate limitations when using AI to detect or contextualize manipulative content and to uncover (cross-border) networks spreading disinformation.

A study published written by Soorya Balendra in 2025 examines how AI-driven content moderation on Meta's social media platforms can reinforce global inequalities, especially outside Western contexts. The platforms increasingly rely on automated systems trained on Western languages, norms, and data. As a result, AI often misinterprets cultural expressions from the Global South, leading either to excessive removal of lawful content or delayed action on harmful material. These errors disproportionately affect users in under-resourced regions, where languages are poorly represented in training data and governments have limited influence over platform policies. The study shows that these problems stem from broader power imbalances in how AI systems are designed and governed. Using a human rights lens, the research argues for more inclusive, multistakeholder approaches to AI governance so that content moderation becomes fairer and better adapted to diverse global contexts.(Balendra 2025)

To mitigate these asymmetries - at least within the European context -, initiatives such as the multilingual AI-tool as part of the PROMPT-project (<https://prompt.socialdata.studio/>) demonstrate how shared infrastructures can enable cross-border collaboration. PROMPT allows fact-checkers, journalists, and researchers to examine disinformation networks in different linguistic and geopolitical contexts, revealing patterns of coordinated behaviour that would remain hidden within isolated national datasets. By aggregating and visualizing sources across eight European languages, the platform supports a comparative understanding of disinformation ecologies that extend national boundaries.

Such tools exemplify the potential of collaborative AI infrastructures to strengthen journalistic resilience at the European level. At the same time, they also underscore that significant global inequalities persist, as similar infrastructures and resources are far less accessible in many parts of the world, reinforcing uneven capacities to analyse and counter disinformation.

In parallel, Unger et al. (2025) note that elites across sectors emphasize the need for contextual awareness and sector-specific expertise when identifying disinformation. This observation supports a core ethical principle for AI use in journalism: algorithms must augment, not replace, domain-specific knowledge. Cross-border AI initiatives should therefore embed editorial and cultural contextualization into their design, ensuring that outputs remain interpretable and verifiable by human experts within each national setting.

3.2.1. Insights from Interviews with Fact-checkers across Europe

Empirical insights from interviews conducted within the framework of the PROMPT project offer a nuanced understanding of how journalists and fact-checkers across Europe perceive and apply AI-supported verification practices. The study employed semi-structured interviews with fact-checkers and journalists from 6 different European countries (Italy, France, Lithuania, Estonia, Romania, Latvia), focusing (amongst other aspects) on the practical integration of AI

and digital tools in their professional routines.⁴ Each interview lasted approximately one hour and was conducted under conditions of anonymity to encourage open reflection on newsroom practices, institutional constraints, and ethical challenges.

The interviews explored how fact-checkers collaborate with journalists (oftentimes within the same organisation), the role of AI-assisted technologies in verification workflows, and the structural barriers that shape these interactions. While participants represented various media systems, the cases presented here - Lithuania and Estonia - were selected because respondents from these contexts provided particularly detailed and illustrative accounts of their national realities. Together, these insights illuminate both the opportunities and obstacles in building a transnational, AI-enhanced fact-checking ecosystem in Europe.

Estonia: Collaboration Between Fact-Checkers and Journalists

In Estonian newsrooms, fact-checkers work closely with journalists, though the intensity of collaboration depends on the type of reporting. Regular journalists often alert fact-checkers to suspicious claims they encounter on social media, enabling immediate verification. Fact-checkers also summarize the most prominent disinformation narratives for television segments, reaching a wider audience and raising public awareness. For investigative stories, collaboration becomes more direct: reporters with unique contacts or field access provide critical information for verification. Fact-checkers support journalists with practical guidance, including verifying images, cross-checking claims, or confirming sources. This approach has fostered a positive cultural shift, with reporters increasingly consulting fact-checkers before publishing, especially during crises or fast-moving events. This reduces the risk of spreading misleading or false information while promoting rigorous journalism.

Lithuania: Fact-Checking as an Integrated Editorial Function

At Lithuanian news organizations, journalists and fact-checkers maintain a structured yet responsive collaboration. Reporters flag suspicious social media content for fact-checkers, who verify these claims and occasionally present the most significant narratives on television, ensuring broad public outreach. While daily collaboration is limited, fact-checkers are frequently consulted on verification questions - how to independently confirm a source, verify an image, or assess a claim's credibility. This collaborative model is increasingly proactive. Reporters now often seek guidance from fact-checkers before publication, particularly in urgent or high-stakes situations. Such interactions strengthen the newsroom's verification culture, helping to prevent misleading information from reaching the public and fostering trust in journalistic standards.

⁴ A full report with further details of the state of disinformation in Europe, focusing on narratives concerning the Moldovan parliamentary elections, the war in Ukraine, and threats on LGBTQIA+ rights and communities can be accessed via the PROMPT website: <https://disinfo-prompt.eu/posts/3pioVYvnxak9OkJT1V8byu>

Cross-Case Analysis

Across all cases, several recurring lessons emerge:

1. **Early Engagement:** Journalists benefit from involving fact-checkers at the beginning of the reporting process, particularly for politically sensitive or fast-moving topics.
2. **Reciprocal Expertise:** Effective collaboration depends on mutual recognition - journalists bring field knowledge, while fact-checkers contribute methodological rigour and digital investigation skills.
3. **Institutionalisation of Verification:** Fact-checking should be integrated into newsroom structures as a routine editorial function rather than an external correction mechanism.
4. **Use of Verification Technologies:** AI-assisted tools such as reverse image search, metadata extraction, and network analysis can accelerate verification, provided they remain under human oversight.
5. **Transnational Awareness:** Disinformation narratives increasingly cross linguistic and geographic boundaries, necessitating shared data infrastructures and cooperative monitoring frameworks.

Together, these insights highlight the need for European-level coordination mechanisms that bridge national differences in capacity, resources, and media cultures. They confirm that responsible AI use in journalism must be grounded in collaborative human practices - where technology amplifies rather than replaces professional judgement.

3.2.2. Case studies: Romanian and Moldovan elections

This section presents a case study of how journalists mobilise AI-based tools in their investigative and analytical work to examine coordinated disinformation and hybrid influence operations in democratic contexts. Rather than focusing on AI as an abstract technology, the analysis foregrounds its practical use within journalistic workflows, drawing on empirical research conducted as part of the PROMPT project.⁵

In Romania, analysis of two large-scale social media datasets - one derived from election-related keywords and another from problematic Facebook accounts identified by a local NGO - revealed 97 clusters of synchronised activity. These coordinated networks repeatedly disseminated emotionally charged, ideologically framed content. Prominent narratives included sarcastic, nationalist, and anti-Western rhetoric targeting liberal institutions, the EU, and NATO; conspiratorial and revisionist discourse with anti-globalist and antisemitic undertones; and populist victimhood narratives portraying parties such as AUR as persecuted defenders of sovereignty. At the same time, pro-European reformist actors, including Elena Lasconi and USR, were both endorsed and undermined through fluctuating moralised and gendered framings. The result was a fragmented and adversarial information

⁵ First Narrative Report: <https://disinfo-prompt.eu/posts/4utNbmaC1keX9Z60IYkxG9>; Second Narrative Report: <https://disinfo-prompt.eu/posts/3pioVYvnxak9OkJT1V8byu>

environment in which legitimacy was contested through competing truth claims, irony, and conspiratorial framing rather than substantive debate.

The Moldovan case extended these findings in a smaller, multilingual media system. Here, the investigation mapped hybridised disinformation networks spanning social media, messaging apps, and collaborative knowledge spaces such as Wikipedia. Manipulation patterns combined platform-specific tactics - notably short-form video amplification on TikTok - with cross-domain coordination via Telegram and Facebook. Identical narratives were tailored to reach both domestic and diaspora audiences across Romanian-, Russian-, and English-language spheres. Instances of narrative laundering were identified, where misinformation seeded on fringe or alternative-language platforms was repackaged through more credible intermediaries, including semi-legitimate news aggregators and user-edited encyclopaedic entries. Wikipedia editing activity, in particular, became a subtle vector of epistemic disruption, echoing prior research on vulnerabilities in collaborative knowledge environments.

Tools developed within the PROMPT framework, such as the WikiSensiBarometer (see section 3.1.), proved instrumental for visualising narrative clusters and detecting early coordination signals. However, despite these advances, both cases demonstrate a persistent asymmetry between transnational manipulation and nationally bounded responses. While hostile actors move fluidly across languages and jurisdictions, journalistic and civic defences remain fragmented.

Taken together, these case studies highlight three priorities for the forthcoming **Best Practice Guidelines**:

1. the integration of cross-border monitoring and early-warning systems into journalistic workflows;
2. the establishment of collaborative verification networks linking journalists, fact-checkers, and civic technologists; and
3. the inclusion of training modules that address not only content-level verification but also the infrastructural mechanisms - algorithmic amplification, virality, and emotional framing - that shape the circulation of disinformation.

Detailed methodologies and full datasets are available in the PROMPT [Narrative Reports 1 and 2 \(2025\)](#). The empirical lessons drawn from these case studies feed directly into the operational dimension of the Guidelines. While Romania and Moldova illustrate how disinformation circulates and adapts across borders, the next section translates these insights into practical newsroom instruments - outlining an Editorial Decision Checklist that supports responsible, transparent, and human-centred AI use in journalistic work.

3.3. Editorial Decision Checklist for AI Use in Newsrooms

As AI tools become embedded in journalistic workflows - from transcription and translation to content generation and verification - the need for clear, operational safeguards becomes essential. The following checklist translates the principles outlined in the Best Practice Guidelines into a step-by-step decision framework for editors and reporters.

Its purpose is to ensure that every instance of AI use in the newsroom upholds the core values of journalism: human oversight, factual accuracy, fairness, and transparency. It also helps newsrooms comply with emerging European regulatory standards (e.g., the EU Artificial Intelligence Act and Council of Europe Guidelines, 2023), while maintaining public trust. The checklist is designed for practical application - to be used as part of editorial review meetings, training sessions, or individual decision-making whenever AI systems are deployed in journalistic work.

Table 2. An attempt to formulate an editorial decision checklist for AI use in newsrooms.

Step	Question / Action	Decision / Notes
1. Purpose Identification	What is the goal of using AI for this task? (e.g., research, draft, fact-checking, language)	If not clearly defined or the public-interest relevance is low → reconsider AI use.
2. Human Oversight	Who will be responsible for reviewing the (AI-generated) output?	Must be an assigned journalist/editor with expertise in the specific field. PROMPT's tool can be used, particularly where AI tools assist transnational monitoring or fact-checking (with different languages) or content analysis.
3. Verification	Can AI output be independently verified? Are sources traceable?	If verification is impossible → AI output cannot be published.
4. Bias & Fairness Check	Does the mode or tool risk bias, stereotyping, or discrimination?	If high risk → apply mitigation measures (e.g., secondary review, diverse sources plus also check sources). The PROMPT tool can help with narrative clustering and language sensitivity modules to identify patterns of rhetorical techniques, narrative imbalance, or specific

		framing across different national media ecosystems.
5. Transparency & Disclosure	Will the AI use be disclosed to the audience?	Include model name/version, role of AI, human review.
6. Legal & Ethical Compliance	Does the AI output comply with laws (copyright, privacy) and editorial codes?	Non-compliance → stop or revise.
7. Quality & Accuracy Assessment	Does the content meet newsroom standards for accuracy, clarity, fairness?	Any factual gaps → revise, do not publish.
8. Post-Publication Monitoring	Who monitors AI-assisted content for errors or audience feedback?	Assign a monitoring point person; maintain logs of corrections. In the PROMPT project, tools such as the WikiSensiBarometer can be used to track text evolution - specifically on Wikipedia -, and identify miscontextualisation. The tool displays verified sources and edit histories to enhance transparency in (frequently edited) Wikipedia articles.

This checklist operationalises ethical AI principles within everyday editorial practice. It reinforces that AI should augment, not replace, human judgement, and that transparency, verification, and accountability remain non-negotiable.

In developing this framework, we do not seek to reinvent the wheel. Numerous guidelines - such as those by the Council of Europe (2023), and various other organisations - already provide valuable foundations. At the same time, both technology and newsroom practices continue to evolve. Any guideline therefore remains a living instrument, subject to regular revision as AI systems, journalistic standards, and regulatory frameworks develop.

Finally, the proposed best practice framework does not exist in isolation. It is grounded in a transnational research effort that acknowledges the inherently cross-border nature of both information manipulation and the journalistic responses required to address it. For this reason, these guidelines - developed within the PROMPT project - underscore the importance of cross-country cooperation. By operating in a multilingual, multiplatform context, PROMPT demonstrates how shared infrastructures can support journalists and fact-checkers across Europe, enabling them to analyse disinformation networks that span languages, platforms, and

national borders. All participating actors benefit from the shared knowledge, including smaller countries and less widely spoken languages, whose local contexts are equally important for understanding the broader disinformation landscape. This collaboration is crucial, as disinformation dynamics, platform logics, and AI-driven tools function well beyond domestic jurisdictions, while ethical and legal safeguards remain largely national. Strengthening such coordinated approaches is therefore essential to equip journalistic actors to navigate increasingly interconnected information environments.

4. Media Literacy for Media Professionals in the Age of AI and Disinformation

As disinformation spreads faster and more widely than ever, media literacy has become an essential skill for journalists and journalism students. The constant circulation of misinformation - amplified by bots, algorithmic recommendation systems, deepfakes, and other AI-driven tools - makes it increasingly challenging to separate factual reporting from manipulative narratives. Real-world events, from elections in Romania and Germany to global campaigns targeting marginalized communities, illustrate how disinformation operates across linguistic, cultural, and national boundaries, demonstrating the need for journalists to critically evaluate content and verify information before dissemination.

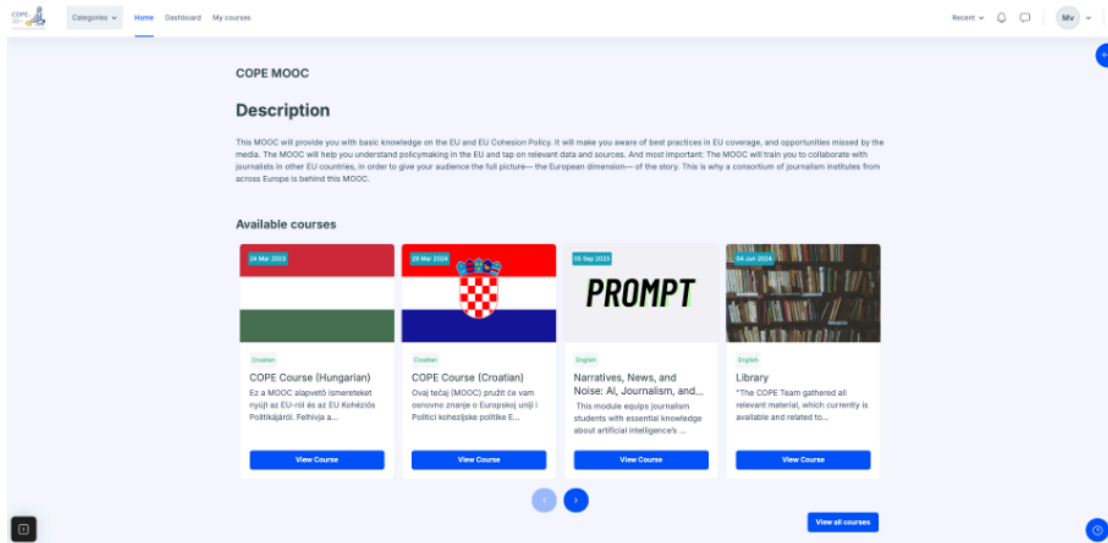
4.1. MOOC Module on “Narratives, News, and Noise: AI, Journalism and the Integrity of Information”

As malicious actors' tactics grow more elaborate, the challenge for journalists and journalism students becomes increasingly complex. That is why media literacy - the ability to critically evaluate and fact-check digital content - is an important tool to defend democracy in Europe today.

To address these challenges, the PROMPT project developed an online learning course integrating a media literacy component on AI, journalism, and disinformation. This new module, titled **“Narratives, News, and Noise: AI, Journalism, and the Integrity of Information,”** is available on the [COPE](#)⁶ (Covering Cohesion Policy in Europe) Training MOOC and equips journalism students and educators across Europe to understand, analyze, and respond to disinformation - with AI being one of the central instruments. The free, self-paced course invites participants to explore how false narratives are created, amplified, and countered through real-world examples and critical reflection. It combines interactive exercises, case studies, multimedia resources, and reflective questions to foster critical thinking and hands-on learning. Participants explore AI's dual role in spreading and detecting disinformation, apply the

⁶ More information on the EU-funded COPE project led by the Erich Brost Institute for International Journalism can be found here: <https://brost-en.org/cope/>

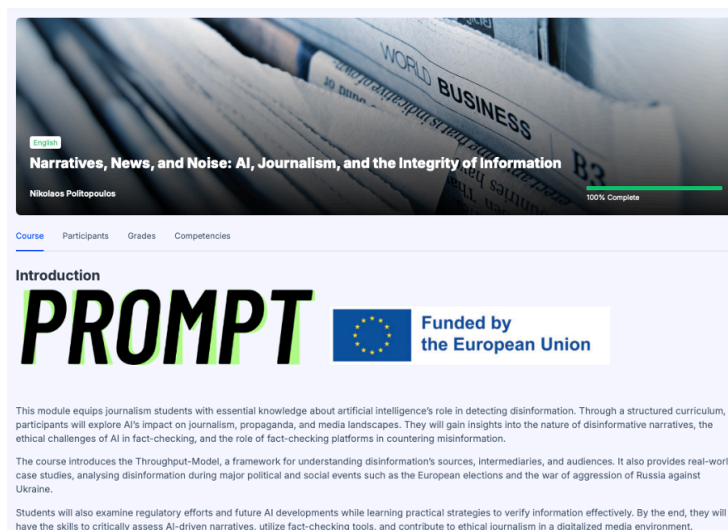
Throughput Model to map disinformation flows, practice fact-checking strategies, and engage with ethical dilemmas and evolving regulatory frameworks.



Screenshot of the COPE platform homepage showing multiple modules, including the PROMPT module.

Designed to be both flexible and pedagogically grounded, the module can be integrated into university curricula or used as a stand-alone resource, enabling learners to progress from understanding and analyzing disinformation to evaluating and applying ethical strategies in AI-driven media environments. By engaging with contemporary disinformation dynamics and journalistic challenges, the course strengthens media literacy across linguistic, cultural, and national contexts, supporting the development of journalists capable of critically navigating a complex, interconnected information landscape. Using this established MOOC infrastructure allowed PROMPT to efficiently integrate its module on AI, disinformation, and journalistic integrity, fostering interactive, cross-border journalism education across Europe

What You'll Learn



Course homepage of "Narratives, News, and Noise": Overview of the module

The module provides both conceptual foundations and practical insights, equipping participants to understand, analyze, and respond to disinformation in a rapidly evolving media landscape. Key areas of focus include:

- **AI's dual role in spreading and detecting disinformation:** Participants explore how artificial intelligence can both amplify misleading narratives - through algorithmic recommendations, bots, and deepfakes - and help journalists identify and verify false content, highlighting the complex ethical and technical challenges involved.
- **Case studies from real-world events:** Detailed examples from the Russian invasion of Ukraine, the Romanian elections, and anti-LGBTQ+ campaigns illustrate how disinformation operates in diverse political, cultural, and social contexts. These cases allow learners to connect theory to practice and understand the local and global impacts of misleading narratives.
- **The Throughput Model:** This framework offers a structured approach to analyzing the flow of disinformation, helping participants map how false information spreads, identify key actors and networks, and evaluate intervention strategies.
- **Fact-checking tools and journalistic strategies:** The module introduces practical resources for detecting, verifying, and countering misinformation, enabling learners to apply evidence-based techniques in real-world reporting and editorial decision-making.
- **Ethical dilemmas and evolving regulatory responses:** Participants examine the moral questions raised by AI-driven media, from algorithmic bias to content moderation, as well as current and emerging policies that seek to regulate disinformation while balancing freedom of expression.

Whether you are a journalism student, educator, or simply someone committed to defending democratic information ecosystems, this module provides the knowledge, frameworks, and practical tools to think critically, make informed decisions, and act responsibly in the face of disinformation.

Who It's For

This module is designed primarily for:

- **Journalism and media students**
- **Educators and journalism schools**
- **Media professionals** interested in AI and information integrity
- **Civil society actors** and everyone engaged with digital media literacy

It can be completed independently or integrated into university courses and training programs as part of the broader COPE MOOC curriculum, which includes 14 modules - plus the new PROMPT module - on journalism, public communication, and European policy.



Image: You're now a detective, putting the clues together./ Source: Pexels/Cottonbro Studio

Exercise: Match Examples to the Information Manipulation Techniques

Drag the words into the correct box Match this real-world examples to different manipulation techniques. The better you get at spotting these tactics, the stronger and more ethical your coverage will become.

A news outlet repeatedly describes female politicians as 'emotional' and 'irrational' while portraying male politicians as "decisive" and 'strong'.

A social media post claims: "Experts warn that if we allow this new technology, it will destroy millions of jobs overnight and make humans obsolete!"

A government report highlights crime statistics showing an increase in violent crimes but omits the fact that overall crime rates have actually declined over the past decade.

A viral post targeting a specific political group claims: "This study proves that renewable energy is completely unreliable!" - while ignoring other studies that demonstrate the stability of renewables.

"The housing crisis exists because of too many regulations. If we remove them, everyone will be able to afford a home!"

Cherry-Picking

Confirmation Bias Exploitation

Fearmongering

Stereotyping

Oversimplification

Exemplary screenshot of one of the interactive exercises in the MOOC module.

Pedagogical Approach and Integration

The module was designed to be both flexible and pedagogically grounded, allowing universities and journalism schools to integrate it seamlessly into existing curricula or to use it as a stand-alone learning resource. Its interactive structure carefully balances theoretical foundations with hands-on learning opportunities, enabling students to engage actively with the material rather than passively absorb content.

Key components of the module include:

- **Analytical exercises** based on real-world disinformation cases, giving students practical experience in identifying, tracing, and assessing misleading narratives across different platforms and contexts.
- **Reflective questions** that encourage critical thinking, helping students to interrogate sources, assumptions, and the broader societal implications of disinformation.
- **Multimedia resources** such as videos, infographics, and datasets that support independent study and allow learners to explore complex topics in multiple formats.

Grounded in **Bloom's Taxonomy**, the course is structured to progressively guide learners from foundational understanding and analysis of disinformation, to evaluating strategies and applying critical approaches for ethical journalism in an AI-driven media environment. By engaging with contemporary disinformation dynamics, AI tools, and ethical challenges, the module equips journalism educators with the means to prepare students for the complex, rapidly evolving realities of reporting in a media landscape increasingly shaped by artificial intelligence.

4.2. PromptED: Interactive Coaching for Trustworthy AI Use in Media

In the following, we present the **PromptED** tool as an external training and coaching resource. While PromptED is **not part of the PROMPT project**, it is relevant in a similar context, as it is specifically designed for media professionals and supports the responsible and informed use of generative AI in editorial workflows. The tool complements existing guidance by offering a practical, experiential approach to understanding how large language models work and where their limitations lie.

<https://prompted.eipcm.org>

PromptED is an innovative training tool designed specifically for media professionals to develop practical skills and foundational understanding of Large Language Models (LLMs). It is intended to be used as a shared learning resource in training workshops led by a skilled facilitator, supporting guided discussion, collective reflection, and hands-on experimentation. Unlike traditional courses or reference guides, PromptED uses an experiential learning approach that enables users to directly observe and understand how generative AI works - and where it can go wrong.

The Learning Experience

The platform features two complementary components:

- **Guided Simulator:** An interactive, step-by-step learning environment where users explore core LLM principles through carefully constructed examples from real journalistic tasks. Users can observe next-word prediction in action, experiment with prompt construction parameters, and see how subtle changes affect output quality and reliability. The simulator reveals how cognitive biases can be triggered by chat-bot-like interfaces and how prompts can inadvertently reinforce confirmation bias in AI outputs.
- **Prompt Playground:** A practical workspace where users apply their learning by creating, testing, and optimizing their own prompts. The system provides real-time suggestions for improvement based on criteria learned in the simulator, helping users develop an intuitive understanding of what makes prompts more reliable.

The Pedagogical Approach

PromptED moves beyond functional skills training by addressing a critical gap: helping users develop appropriate mental models of how LLMs actually work. By enabling direct observation of system behaviour through interactive examples, users build genuine understanding rather than just following rules. The platform includes an evaluation checklist aligned with journalistic standards, guiding systematic analysis of LLM outputs for accuracy, bias, plagiarism, logical fallacies and other quality factors.

This hands-on approach equips journalists and media professionals with both the knowledge and practical capabilities needed to use generative AI responsibly in content production while maintaining professional ethical standards.

5. Further Resources

Council of Europe - Guidance Note on Countering the Spread of Online Mis- and Disinformation Through Fact-Checking and Platform Design Solutions in a Human Rights-Compliant Manner

A comprehensive Council of Europe guidance document outlining strategies to counter online misinformation and disinformation through fact-checking methodologies and platform design, with an emphasis on human rights compliance.

<https://edoc.coe.int/en/internet/11885-guidance-note-on-countering-the-spread-of-online-mis-and-disinformation-through-fact-checking-and-platform-design-solutions-in-a-human-rights-compliant-manner.html>

OECD - *Facts Not Fakes: Tackling Disinformation and Strengthening Information Integrity* (2024)

A policy report from the Organisation for Economic Co-operation and Development that examines disinformation ecosystems, explores policy approaches to strengthen information integrity, and provides recommendations for governments and stakeholders.

https://www.oecd.org/en/publications/2024/03/facts-not-fakes-tackling-disinformation-strengthening-information-integrity_ff96d19f.html

AI Charter - *Charter for Artificial Intelligence (Les Surligneurs, 2025, in french)*

A principles-based AI charter developed by Les Surligneurs, outlining ethical commitments and democratic safeguards for the use and governance of artificial intelligence. It offers values and recommendations to guide responsible AI design, deployment, and societal impact assessment.

https://lessurligneurs.eu/new_website/wp-content/uploads/charte-ia-les-surligneurs-2025.pdf

UNESCO - AI and Media Ethics: Regional Declaration on Ethical and Transparent Use of AI in the Media

This article highlights a landmark *Regional Declaration* adopted by press councils in South-East Europe and Türkiye, outlining principles for ethical and transparent use of AI in journalism,

emphasizing human judgment and editorial responsibility.

<https://www.unesco.org/en/articles/ai-and-media-ethics-press-councils-south-east-europe-and-turkiye-adopt-landmark-declaration>

Trusting News - AI Trust Toolkit for Journalists

A practical resource from *Trusting News* designed to help newsrooms build audience trust around AI use. It offers strategies for transparency, ethical AI disclosure, and audience engagement when integrating AI in news production.

<https://trustingnews.org/trustkits/ai/>

Three Steps to an AI-Ready Newsroom - Responsible Policies Guide

A guide by the Thomas Reuters Foundation for newsrooms on how to craft responsible internal policies for AI use, including managing risks, defining ethical norms, and aligning AI workflows with journalistic values like fairness, accuracy, and accountability.

<https://www.trust.org/resource/ai-policies-newsroom-guide/>

6. References

AI Code Project (AICODE). *Journalism in the Age of AI*.

<https://aicode-project.eu/2025/07/22/journalism-in-the-age-of-ai/>

Balendra, Soorya (2025). "Meta's AI moderation and free speech: Ongoing challenges in the Global South". *Cambridge Forum on AI: Law and Governance*.

<https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/metas-ai-moderation-and-free-speech-ongoing-challenges-in-the-global-south/2DB952F896DB5744A43CD3E6C1A6DCB4>

Bradley-Schmieg, P. (2023). *Wikipedia is now a Very Large Online Platform (VLOP) under new European Union rules: Here's what that means for Wikimedians and readers*.

<https://diff.wikimedia.org/2023/05/04/wikipedia-is-now-a-very-large-online-platform-vlop-under-new-european-union-rules-heres-what-that-means-for-wikimedians-and-readers/>

Columbia Journalism Review (CJR). *Artificial Intelligence in the News*.

https://www.cjr.org/tow_center_reports/artificial-intelligence-in-the-news.php

Columbia Journalism Review (CJR). *Can AI tools meet journalistic standards?*

<https://www.cjr.org/analysis/can-ai-tools-meet-journalistic-standards.php>

Columbia Journalism Review (CJR). *How we're using AI in newsrooms.*

<https://www.cjr.org/feature/how-were-using-ai-tech-gina-chua-nicholas-thompson-emilia-dav-id-zach-seward-millie-tran.php>

Columbia Journalism Review (CJR). *I tested how well AI tools work for journalism.*

<https://www.cjr.org/analysis/i-tested-how-well-ai-tools-work-for-journalism.php>

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.

Council of Europe (2023). *Guidelines on the Responsible Implementation of Artificial Intelligence Systems in Journalism.*

<https://rm.coe.int/cdmsi-2023-014-guidelines-on-the-responsible-implementation-of-artific/1680adb4c6>

de-Lima-Santos, M.F., Yeung, W.N., & Dodds, T. (2024). "Guiding the Way: A Comprehensive Examination of AI Guidelines in Global Media". *Computers and Society*.

Diakopoulos, N. & Koliska, M. (2017). "Algorithmic Transparency in the News Media". *Digital Journalism*, 5(7): 809-828.

DigWatch (2023). *Guidelines on the responsible implementation of artificial intelligence systems in journalism.*

https://dig.watch/resource/guidelines-on-the-responsible-implementation-of-artificial-intelligence-systems-in-journalism?utm_source=chatgpt.com

European Council (2025). *Artificial intelligence act.*

<https://www.consilium.europa.eu/en/policies/artificial-intelligence/#AI%20act>

European Union (2023). *EU AI Act.*

<https://www.europarl.europa.eu/topics/en/article/20230601ST093804/eu-ai-act-first-regulation-on-artificial-intelligence>

Fondation Hironnelle (2025). *How to reconcile artificial intelligence and journalistic work.*

<https://www.hironnelle.org/en/how-to-reconcile-artificial-intelligence-and-journalistic-work-fondation-hirondelles-directive>

INA (2026). *Whether enthusiastic or cautious, journalism schools are adopting AI (translation, article in french).*

<https://larevuedesmedias.ina.fr/ecoles-de-journalisme-intelligence-artificielle-chatbots-etudiants-journalistes>

McDowell, Z. J., & Vetter, M. A. (2020). *Wikipedia and the Representation of Knowledge.*

Prompt (2025). Narrative Report #1.

https://assets.ctfassets.net/26u2dy66hbfc/2xc603sw106eULMfHefKmg/495535d6bbe0e950dcfc6a1b0b784d80/PROMPT_Case_Studies_Report_1_27.06..docx_1_.pdf

Prompt (2025). Narrative Report #2.

https://assets.ctfassets.net/26u2dy66hbfc/159V5Wv6rUpFieqkFfM472/062c20c525582fe6a05ad8a2b55bee8b/PROMPT_Narrative_Report_2.pdf

Pulitzer Center. *AI Accountability and Journalism resources.*

<https://pulitzercenter.org/>

Saez-Trumper, D. (2019). *Fighting Disinformation on Wikipedia: Past, Present, and Future.* [10.48550/arXiv.1910.12596](https://arxiv.org/abs/1910.12596)

Unger, S. et al. (2025). "Features of disinformation: an expert interview study on the perception of disinformation among political, governmental, media and business elites in Germany". *Journal of Elections, Public Opinion and Parties*, 35:3, 472-494, doi: 10.1080/17457289.2025.2514199.

UNESCO (2024). *AI and the future of journalism: an issue brief for stakeholders.*

<https://unesdoc.unesco.org/ark:/48223/pf0000391214>

Vera-Quiroz, V. (2024). *Stories from the anti-disinformation repository: Tools created and used by Wikimedians to counter disinformation.*

<https://diff.wikimedia.org/2024/06/27/stories-from-the-anti-disinformation-repository-tools-created-and-used-by-wikimedians-to-counter-disinformation/>

Wikimedia (2021). Croatian Wikipedia Disinformation Assessment. *Wikimedia Meta-Wiki.*

https://meta.wikimedia.org/wiki/Croatian_Wikipedia_Disinformation_Assessment-2021

Wikimedia (2024). *Wikipedia EU DSA Systemic Risk Assessment.*

https://foundation.wikimedia.org/wiki/File:WMF_24-25_Wikipedia_EU_DSA_Systemic_Risk_Assessment.zip