



Artificial intelligence and journalism: guidelines for ensuring information quality

Intelligence artificielle et journalisme : des règles
pour garantir la qualité de l'information

Laurence Dierickx – Carl-Gustav Lindén
University of Bergen

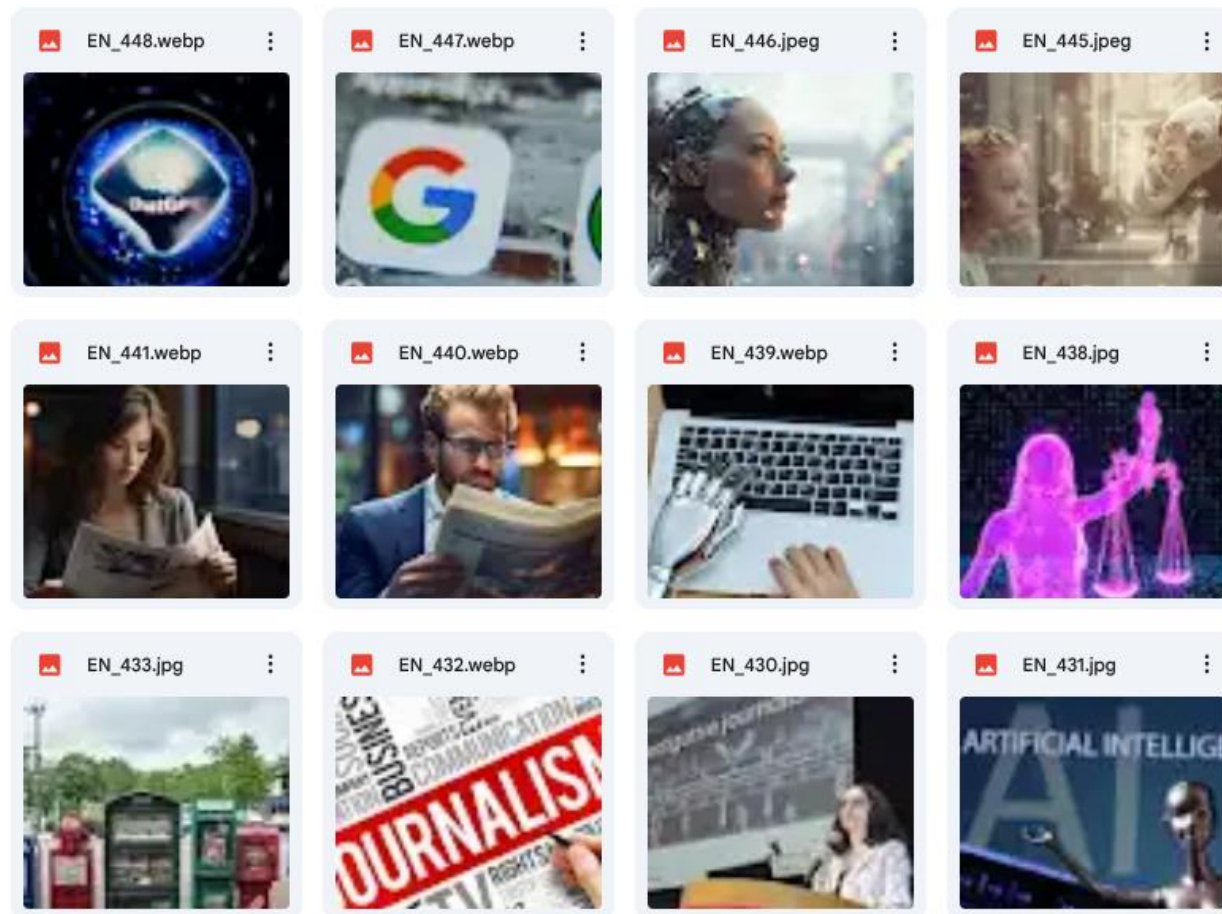
Nordic Observatory for Digital Media and Information Disorder (NORDIS)

Document numérique & Société
September 12, 2024



AI in journalism

- Breaking news detection
- News discoveries
- Networks analysis
- Audiovisual search engines
- News verification
- Automated fact-checking
- Data and text analysis
- Text summarisation
- Audio transcription
- Machine translation
- Text-to-speech applications
- Text generation
- Multimedia creation
- Personalisation and recommendation
- Etc.



LLMs: the game changers?

Operational complexity: LLMs learn patterns from large datasets, raising challenges in ensuring ethical use and reliable result (stochastic parrots).

- Data quality of training data (UGC, biased data)
- Risks of plagiarism (copyrighted data)
- Failures in logical deduction ($A = B$ is not necessarily $B = A$)
- Artificial hallucinations (training data and processes)
- Generation is not verification (designed for providing an answer)
- Risk of being fooled by a convincing tone (anthropomorphism)
- Socio-professional risks (replacement of human work, impact on critical thinking and creativity)

How do media and professional organisations develop and apply guidelines to ensure the quality of information in response to the challenges posed by large language models (LLMs), and what are the common features and key principles of these guidelines?

Focus on Northern and Western Europe

- Self-regulatory bodies
- Ethical codes agree on accuracy, fairness/balance, independence, privacy, protection of sources
- High level of professionalism
- Impact of news media Public sphere
- Government support (public service broadcaster)

Method

- Search engines (Google, Ecosia)
(AI OR "artificial intelligence") AND (journalism OR media OR press) AND (charter OR principles OR ethics) AND (country)
- (Human) Monitoring on Twitter/X
- Limits: publicly available texts (no internal documents)
- 36 text from 10 countries between February 2023 and April 2024
(Belgium, France, Germany, the Netherlands, Switzerland, United Kingdom, Denmark, Finland, Norway, Sweden)
- Non-English text translated with DeepL (human supervision)

Anatomy of the corpus

- News media (15)
- Public broadcasters (6)
- Press agencies (4)
- Press councils (4)
- Press groups (3)
- Professional organisations (4)
- “Principles”, stating fundamental principles and including texts presented as “Charter” in French (14)
- “Directives”, relating to recommendations and practical advice and including texts presented as “Guidelines” in English (14)
- “Positions”, focusing on commitment and perspective (6)
- Ethical codes published by press councils (2)

Indicator	Evaluation Criteria
Accuracy	Does the text emphasise truthfulness? Does it consider the ethical aspects of data quality?
AI Literacy & Training	Does it promote education or training of journalists in AI and data culture? What measures are in place to support such training?
Copyright	Does it address copyright issues related to data used in AI training? Does it address plagiarism and detection methods?
Fairness	Does it address fairness and impartiality in information production? Does it address balance and diversity in information?
Inclusivity	Does it address "new actors" in journalism? Does it encourage cross-sectoral cooperation?
Responsibility	Does it clearly define responsibilities, including editorial decisions and human involvement?
Risk Mitigation	Does it discuss risks such as bias and AI hallucinations? Does it suggest mitigation strategies?
Security & Privacy	Does it address system security? Does it respect privacy and consider legal obligations to protect it?
Sustainability	Does the text address system and data maintenance? Does it consider the environmental impact of AI technologies?
AI Ethics	Does it consider the principles of transparency and explainability?

Qualitative data analysis (Taguette)

General observations

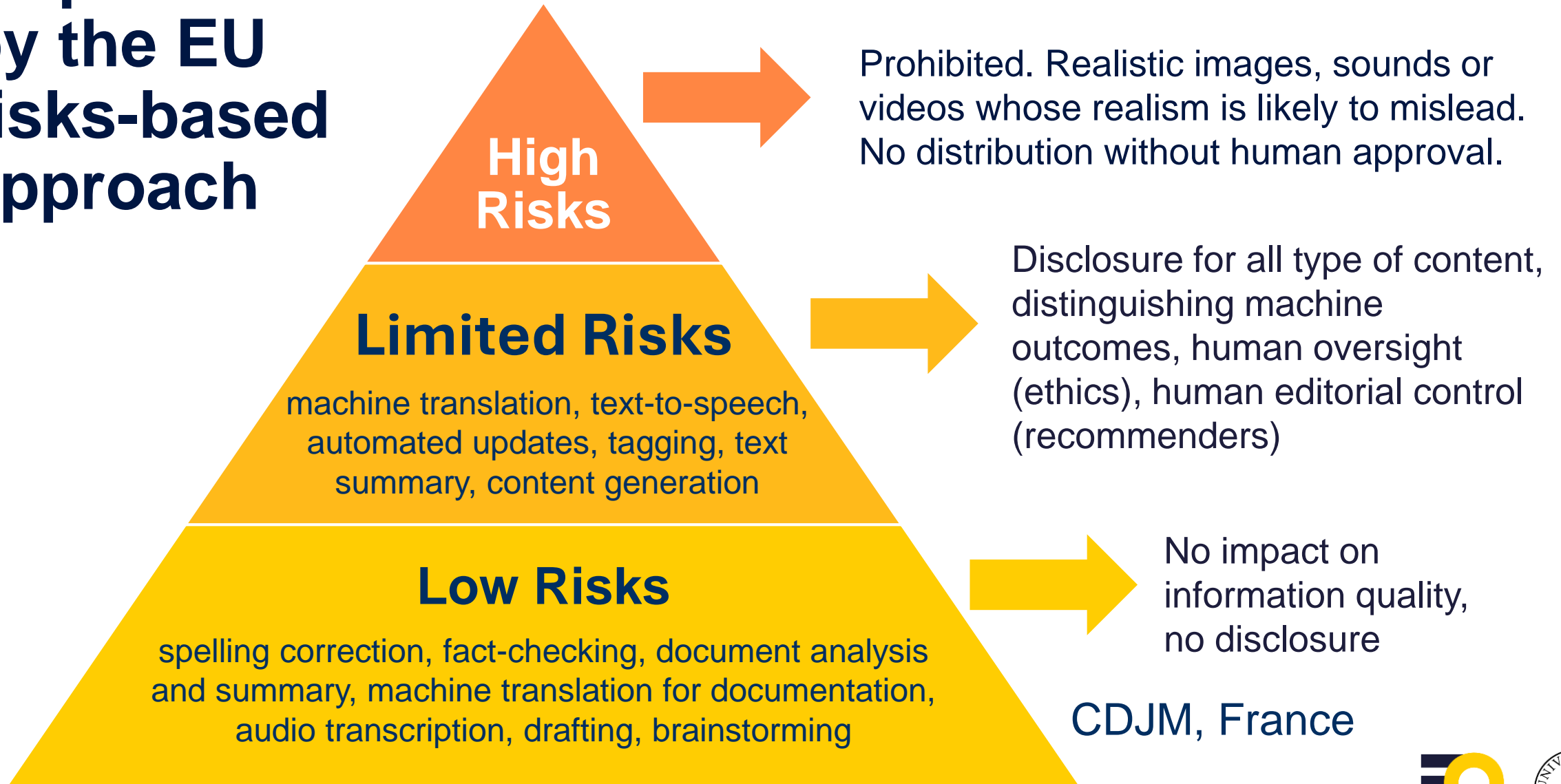
- News media organisations are using AI, both generative and non-generative, to assist with various tasks from information gathering to distribution
- These organisations recognise the risks of AI systems, including bias (AJP, ANP, *Financial Times*, Mediahuis, RTS, Yle), potential errors and untrue content (ANP, STT).
- AI systems may rely on copyrighted data (mentioned in half of the texts).
- There is an awareness of the paradox of AI technologies, which can be used to both inform and disinform (*Dagens Næringsliv*, *De Volksrant*, *Der Spiegel*, *Financial Times*, SVT).

Experimenting with caution

Recommendations or guidelines is part of a risk mitigation strategy that promotes responsible practices.

- Other strategies include testing and approval mechanisms to prevent AI "hallucinations" from being published (*Der Spiegel*, Germany).
- Due to potential errors, STT (Finland) avoids using AI for data exploration.
- Yle (Finland) emphasises ongoing risk assessment and vigilance to monitor and correct biases.

Inspired by the EU risks-based approach



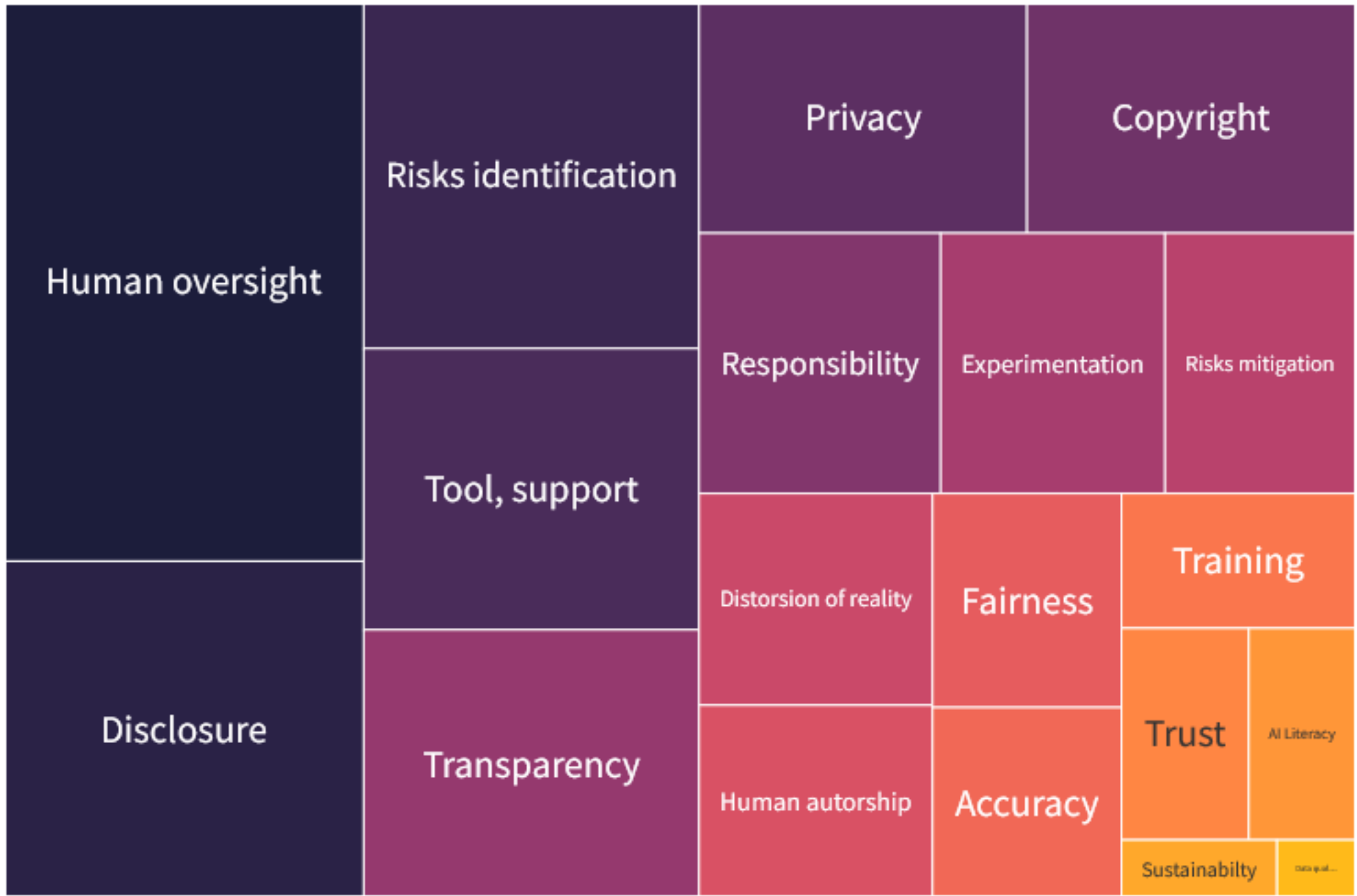
CDJM, France

1. Human oversight (100%)

- Editorial responsibility and control in the use of AI.
- Journalists and editors-in-chief are accountable for content, whether or not AI is used (BBC, *Der Spiegel*, *Financial Times*, Keystone-ATS, Mediahuis, Roularta, RTS).
- Trust and credibility are key issues, closely linked to media accountability.
- AI technologies are approached cautiously, with an emphasis on not replacing human judgement.
- For example, Aftonbladet will clearly label AI-generated content, Le Monde demands strict editorial control, and others like Dagens Næringsliv and *De Volksrant* use AI as a complementary tool, not a replacement for journalism.

2. Transparency (88,89%)

- Transparency can include disclosing when content is generated by AI (27 texts). This requirement is less common for AI used as an assistant.
- The French CDJM advises journalists to cite data sources and tools used (challenging due to time constraints and complex tools)
- Transparency is not included in most countries' codes of journalistic ethics, with the exception of Belgium and the UK, where press councils updated their codes in 2023. The Flemish Raad voor de Journalistiek requires transparency about the journalistic approach and human responsibility in editorial decisions. In the UK, Impress requires clear labelling of AI-generated content and information about data collection and use + emphasis on the outcomes' accuracy



Tags frequency

Dealing with biases and hallucinations

- Replication the existing societal perspectives, including historic biases
(*Financial Times*)
- It is complex to guarantee the reliability of facts presented as true
(ANP)
- Material created using generative AI raises significant issues around bias, ownership, plagiarism and intellectual property rights
(*The Guardian*)
- Watch out for biases in AI systems and work to address them
(Mediahuis)
- Any fully automated services must be considered very carefully
(Redaktørforeningens)

Reflecting journalism ethics

- Accuracy (critically assessing and verifying sources and facts)
- Trustworthiness (sources may be unreliable, harmful or inaccurate)
- Respect for facts (not manipulating or distorting them)
- Fairness (avoid bias and echo chambers, encourage diversity)
- Respect for (data) privacy
- Human responsibility and accountability

**AI algorithms are part of the editorial process.
Designed by humans for human purposes and uses.**

Current issues and gaps

- Professional organisations blurred messages about political and trade union positions (Belgium, Germany)
- Only Yle (Finland) considers the environmental impact of AI.
- Less focus on the possibility of private or internal data leaks (only SVT in Sweden and Ringier in Switzerland)
- New actors in journalism, such as data scientists and computer scientists, are not considered
- Ethics is a matter of practice (ethical dilemmas)
- Risks mitigation also involve data and AI literacy + training

Transparency is not enough

- Obscures decision-making process
- Risks of information overload
- Does not guarantee accuracy, fairness and reliability
- Transparency is not explainability
- Explainability has also several drawbacks, can lead to misinterpretation
- Transparent processes may lack trustworthy results
- Black-box of journalism (editorial, decision-making)

Thank you for your attention!

Contact: @ohmyshambles @Gusse

Data: <https://github.com/laurence001/AIguidelines/>

Dierickx, L., van Dalen, A., Opdahl A.L. and Lindén C.G. *Striking the Balance in Using LLMs for Fact-Checking: A Narrative Literature Review*, in Proceedings from 6th Symposium on Multidisciplinary International Symposium on Disinformation in Open Online Media (MISDOOM), Lecture Notes in Computer Science (Springer).

