Guardrails for the Algorithm: How Global Newsrooms Govern AI Ethics

Chang Sup Park

This study examines how news organizations across the world articulate AI ethics through formal guidelines, proposing a novel three-dimensional framework of AI governance:
- Epistemic Governance: Mechanisms ensuring accuracy, verification, and knowledge credibility
- Procedural Governance: Structures for transparency, accountability, human oversight, and newsroom workflows
- Protective Governance: Frameworks mitigating broader societal harms, promoting diversity, and preserving editorial independence.

Through comparative content analysis of 31 AI ethics guidelines across 19 countries and 6 world regions, I found a systematic governance asymmetry: news organizations heavily prioritize *epistemic* and *procedural* governance while leaving *protective* governance significantly underdeveloped.

Key Findings

RQ1: How are traditional journalism ethics reflected in AI policies?
- Core principles from the SPJ Code of Ethics—specifically Accuracy/Verification (A1), Transparency/Disclosure (D1), and Accountability (D3)—are nearly universally adopted across newsroom guidelines.

RQ2: How do news organizations prioritize the three governance dimensions?
- Epistemic Dominance: Accuracy and verification function as non-negotiable boundaries separating legitimate journalism from machine hallucination. However, governance stops at publication—none of the 31 guidelines cover post-publication updating/correcting of AI content (A3).

- Procedural Standardization: Human-in-the-loop oversight is enforced across all organizations. Larger public service media (e.g., BBC, Sveriges Radio) exhibit greater “governance depth” with formal advisory boards and ethics coordinators.

- Protective Deficit: Harm minimization (B1–B3) and editorial independence (C1) are rarely prioritized. Harm is frequently reframed as internal data security or corporate risk management rather than public/audience protection.

RQ3: What ethical concerns emerge beyond traditional frameworks?
- Intellectual Property & Asset Protection: Newsrooms actively defend content against unauthorized AI training and reuse, treating IP as a critical cross-cutting institutional concern.

Comparison: Three Dimensions of AI Governance

Governance Dimension	Core Focus & Definition	Primary SPJ Codes Mapped	Key Organizational Implementation / Examples	Identified Limitations & Gaps
Epistemic Governance	Production, validation, and reliability of journalistic knowledge.	A1 (Accuracy/Verification) A2 (Speed vs. Accuracy) A4 (Source Sourcing) A7 (Visual Distortion)	Independent fact-checking against human sources (Dawn, CNA, Rappler) Analogy to search engines/Wikipedia (WIRED) Strict bans on photorealistic visual generation (Agência Pública, WTWH Media)	Complete absence of post-publication correction/updating policies (A3). Focus is strictly ex-ante quality control.
Procedural Governance	Rules, accountability, disclosure, and human oversight in workflows.	D1 (Transparency) D2 (Corrections) D3 (Accountability)	Labeling and audience disclosure of AI deployment (Journo, Daily Maverick) Institutional roles (e.g., BBC's AI Risk Advisory Group, Journo's AI Ethics Coordinator) Non-delegable human responsibility (Agência Pública, NPR)	Commercial outlets often limit commitments to broad aspirational statements without formal oversight roles.
Protective Governance	Harm mitigation, fair trial, diversity, and editorial autonomy.	B1–B3 (Harm/Fair Trial) C1 (Independence) A5 (Diversity)	Data governance & local LLM hosting (Poynter) Indigenous rights frameworks (RNZ & Te Tiriti o Waitangi) Mandatory disclosure of vendor	Public protection is often reframed as corporate risk and data privacy. Diversity (A5) present in

			funding/partnerships (Dawn, Poynter)	only 7 of 31 policies.

Key References

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973–989.

Carlson, M. (2017). *Journalistic authority: Legitimating news in the digital era*. Columbia University Press.

Diakopoulos, N. (2019). *Automating the news: How algorithms are rewriting the media*. Harvard University Press.

Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707.

Helberger, N., & Diakopoulos, N. (2023). ChatGPT and the AI turn in journalism: A framework for governance and newsroom decision-making. *Digital Journalism*, 11(8), 1429–1435.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.

Napoli, P. M. (2019). *Social media and the public interest: Media regulation in the disinformation age*. Columbia University Press.

Society of Professional Journalists (SPJ). (2014). *SPJ Code of Ethics*. https://www.spj.org/ethicscode.asp