

AI and Media Ethics Existing References Overview

This document offers a brief summary of current reference materials that outline ethical guidelines and standards in media and artificial intelligence, as well as where these two domains intersect.

1. The evolution of the right to information

The Universal Declaration of Human Rights makes freedom of opinion and expression a fundamental human right (art.19). This right includes *freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers.*

The [International Declaration on Information & Democracy](#) was published in November 2018 as an effort to specify fundamental principles for the global information and communication space, beyond journalism ethics and standards. It defines this space as a *common good of humankind that should be protected as such*. Such space should serve the exercise of freedom of expression and opinion and respect the principles of pluralism, freedom, dignity, tolerance, and the ideal of reason and understanding.

The Declaration enshrines the right to reliable information and states that access to knowledge is a fundamental right. It outlines several principles to ensure this right : **truthfulness, accuracy, plurality** of viewpoints, "**do-not-harm**" principle, respect of **privacy** (with exceptions), **responsibility** and **liability** of public debate actors, and **transparency** of entities with influence power.

The Declaration also specifies the social function of journalism : that of a "*trusted third party*" for societies and individuals. Journalism's task is also *to allow the public to distinguish the important from the trivial*. Pursuing these goals requires that journalists handle information in such a way that serves the public interest, does not treat information as a commodity, presents the facts **fairly** and acts in complete **independence** from all forms of power and undue influence.

2. Journalism ethics and professional standards

Journalism ethics and standards can be traced back to the beginning of modern journalism in Europe. Numerous organisations have published codes of ethics which aim at defining what journalists and news organisations should do, given their role in society.

As of today, most professional journalists and media organisations have pledged to respect and implement ethical principles by elaborating their own codes of ethics, or directly referring to reference declarations, including, among others, the [Declaration of duties and rights of journalists](#) (1971) (also known as Munich Charter) or the International Federation of Journalists (IFJ), and [Global Charter or Ethics for Journalists](#) (2019). The EU's parliamentary assembly also issued [a resolution](#) (1993) that affirms ethical principles for journalism. None of these declarations are legally binding.

While each code has its own scopes and specificities, it is possible to outline some widely-agreed ethical principles and professional standards :

- **Truthfulness** and **accuracy**: journalists should strive for truth and accuracy, give all the relevant facts in their possession and verify them.
- **Independence**: journalists should not act on behalf of any kind of special interests.

- **Impartiality** and **fairness**: journalists must try to be as objective as possible, by reporting different viewpoints fairly.
- "**Minimise harm**" principle: most codes underline the importance of weighing the pros and cons of publishing content. Journalists are sometimes asked to act in the **public interest**.
- **Privacy**: Journalists must respect the right to privacy, unless the disclosure of information is essential for the purpose of the published content.
- **Accountability** and **liability**: Any piece of content that is published should be signed by an individual or a moral entity. Corrections are published when errors are discovered.

3. General AI ethics

The exponential rise of AI systems has been met with an equivalent surge of declarations and guidelines from scholars, public organisations, NGOs and private companies. Several of these address the challenges brought about by the growth of AI in the news and information sector.

Academic knowledge

The ethics of AI is a rapidly growing field of research. Works can broadly be classified into two general categories: those that mainly focus on the present-day and short-term challenges posed by the ubiquity of algorithms in our information systems; and those which delve into medium/long-term challenges associated with the rapid advancement of AI capabilities. As AI systems grow in sophistication and power, the boundaries between these two fields of study begin to blur.

The first field considers the contemporary challenges raised by the development and adoption of AI systems into our societies. Authors in this field have outlined issues such as the subordination of global information flows to attention-seeking algorithms, the global surveillance system emerging from AI being trained on personal data to predict and influence human behaviour, or the amplification of biases within algorithms. This field provides numerous useful insights for designing regulations targeting the uses of AI in the global information space, including the media.

The second field emphasises the long-term consequences of AI development, particularly reflecting on the challenges associated with the emergence of Artificial General Intelligence (AGI), broadly defined as a type of AI that can understand, learn, and perform any intellectual task that a human being can. Authors in this field focus on questions relating with the alignment problem (how to align AI with human interests), itself subdivided into numerous epistemological, philosophical and ethical questions. Although extending well beyond the realm of contemporary challenges, this field of study offers perspectives valuable for anticipating the future risks posed by AI.

Intergovernmental and governmental initiatives

The [Montréal Declaration for a Responsible Development of Artificial Intelligence](#) (2018) is a set of ethical guidelines aimed at ensuring the responsible development and deployment of AI. The OECD has established an AI Policy Observatory, which in 2019 published some broad [AI principles](#), including human-centred values, transparency and explainability, robustness and security, and accountability. UNESCO published a document in 2021 titled "Recommendation on the Ethics of Artificial Intelligence". One chapter is dedicated to "Information and communication". It emphasises that AI should be leveraged to enhance access to information and knowledge for various stakeholders, including researchers, journalists, and the public.

In July 2023, the United Nations held its first Security Council meeting on AI, discussing its potential threats to global peace, and emphasising the need for global collaboration on AI safety and governance. The UN is launching a [multistakeholder advisory body on AI](#), whose main goal is

to "help build global capacity for the development and use of AI in a manner that is trustworthy, human rights-based, safe and sustainable, and promotes peace."

The past two years have seen a push to operationalize principles, with governments trying to convert them into concrete policy frameworks (Canada's directive on the use of AI in government, Singapore's Model AI Governance Framework, Japan's Social Principles of Human-Centric AI, or the U.K. guidance to using artificial intelligence in the public sector). The EU Artificial Intelligence Act (AI Act), approved by the EU Parliament in June 2023, is the world's first comprehensive legislation regulating AI use. The final version is expected by the end of 2023.

Private and non-governmental initiatives

Several NGOs and corporations working on AI have crafted their own charters and constitutions. These documents mainly offer general principles for AI design and training. The [Asilomar Principles](#) (2017) are a set of guidelines developed to ensure the safe and beneficial development and application of AI. Anthropic, an AI company backed by Google parent's company Alphabet, has developed a set of written moral values, referred to as [Claude's constitution](#), to train its AI systems (Claude being its main LLM project). This "constitutional AI" approach is designed to ensure AI systems like chatbots adhere to specific ethical guidelines, drawing inspiration from various sources such as the UN Declaration on Human Rights or Apple's data privacy rules.

AI Ethics: Diverse Approaches with Little Common Ground

A [research paper](#) published in Nature Machine Intelligence in 2019 investigated a corpus of principles and guidelines on ethical AI. While the authors observed a global convergence around five ethical principles (transparency, justice and fairness, non-maleficence, responsibility and privacy), they underlined a substantive divide in how "*these principles are interpreted, why they are deemed important, what issue, domain or actors they pertain to, and how they should be implemented*".

4. Media ethics regarding AI technology

Media outlets tend to remain discreet and not prescriptive about their uses of ICT, including AI driven technologies. Some have published specific guidelines that frame their use of social media (such as [The Guardian](#), [The Washington Post](#) or the [BBC](#)). Others have disclosed extensive editorial standards that regulate their use of digital content (such as [Al Jazeera](#) and [AP](#)).

Overall, [publishers and broadcasters](#) have focused more on intellectual property issues and the responsibilities that should be imposed on AI developers, operators or users, than on the ethical guidelines for the use of AI-based technologies in journalism.

Following the explosion of generative AI at the beginning of 2023, a [minority](#) of newsrooms have published ethics guidelines. Most of these [focus](#) on issues related to:

- human oversight of published content
- transparency regarding the synthetic or human origin of published content
- the limits set on the use of generative AI by journalists
- the need to not disclose private information to/on AI platforms
- the claim of human responsibility for any published content
- the risk of bias embedded in generative AI tools.

The scope and the level of demand of these guidelines vary greatly. Some have no overlapping areas. Regarding the use of other types of AI, such as recommender systems, moderation tools, or forecasting models, it is worth noting that media outlets typically have not published any general rules or recommendations.