

AI ETHICS RESEARCH THEMATIC AREA NARRATIVE IN ENGLISH ARABIC FRENCH PORTUGUESE AND SPANISH

Rachel Adams , Kelly Stone

Rachel Adams , Kelly Stone

©2024, RACHEL ADAMS , KELLY STONE



This work is licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted use, distribution, and reproduction, provided the original work is properly credited. Cette œuvre est mise à disposition selon les termes de la licence Creative Commons Attribution (<https://creativecommons.org/licenses/by/4.0/legalcode>), qui permet l'utilisation, la distribution et la reproduction sans restriction, pourvu que le mérite de la création originale soit adéquatement reconnu.

IDRC GRANT / SUBVENTION DU CRDI : - GLOBAL INDEX ON RESPONSIBLE ARTIFICIAL INTELLIGENCE

Global Index on Responsible AI

Dimension: National Responsible AI Capacities

Sub-dimension: Investments

Thematic area: [AI Ethics Research](#)

Definitions

[Ethics](#) can be defined as ‘a set of **moral principles** that govern a person’s behavior or the conducting of a particular activity’. In this regard, ethics can be thought of as a set of principles [based on notions of right and wrong](#) that help determine what constitutes acceptable and unacceptable conduct and/or behavior.

[AI Ethics](#) can therefore be defined as ‘a system of moral principles and techniques’ intended to guide the design, development, use and deployment of AI technologies. While standards for AI Ethics are always evolving, the UNESCO Recommendation on the Ethics of Artificial Intelligence (Recommendation) has articulated four core values underlying ethical AI: (1) human rights and human dignity; (2) living in peaceful, just and interconnected societies; (3) ensuring diversity and inclusiveness; and (4) environment and ecosystem flourishing. In addition, the Organisation for Economic Cooperation and Development (OECD) has identified five [value-based principles](#) for AI, which include: (1) inclusive growth, sustainable development and well-being; (2) human-centred values and fairness; (3) transparency and explainability; (4) robustness, security and safety; and (5) accountability.

[AI Ethics Research](#) is an emerging [field of study](#) within applied ethics and the philosophy of technology concerned with how ethical principles are currently applied to AI and whether they serve as [adequate safeguards](#) against some of the risks posed by the development and deployment of AI systems and tools.

Justification

Artificial Intelligence has long been considered a field of study which raises ethical concerns. The recent applications of AI have led to consequences like discrimination, widespread surveillance, and a lack of accountability. Moreover, it has raised concerns in disputed areas of AI use in [warfare](#). To address these negative consequences, [ethical principles](#) have been drafted by global experts aiming to design AI with more accountability, fairness and privacy. This includes the OECD AI Principles as well as the UNESCO Recommendation on Ethical AI which aim to create a ‘dynamic basis for the [normative evaluation and guidance](#) of AI technologies, referring to human dignity, well-being and the prevention of harm’.

[Normative frameworks](#) for AI Ethics aim to establish ‘[guardrails](#)’ for the use and deployment of AI technologies, in order to reduce various risks such as bias, discrimination, and threats to human rights and freedoms. However, it's crucial to acknowledge the [complexity](#) of applying these principles in practice and anticipate new ethical challenges that may arise with the advent of new AI-based technologies, like generative AI. Furthermore, AI could lead to varied ethical issues or emphasize specific AI principles differently across countries and regions. Therefore, nations should enhance their capabilities in AI Ethics Research by investing in this field. The objective is not only to adhere to a developing set of ethical principles that guide the responsible use of AI tools and systems but also to comprehend the local conditions necessary to implement these ethics. This will help determine their relevance and usefulness to the local population. Investments in AI Ethics Research should look to: (1) promote an [interdisciplinary understanding](#) of ethical issues; (2) identify and anticipate the social, economic and political implications of AI technologies; and (3) pose [critical questions](#) around the appropriateness of AI use in various facets of life.

This highlights the importance of AI Ethics Research in informing the debate with data about current guidelines, drafting new principles, criticizing approaches as well as highlighting important areas that AI Ethics need to tackle. This thematic area therefore aims to measure and explore how governments and other stakeholders pave the way for this research to take place by investing for example in AI ethics publications, research, and institutional set-ups.

Identification

This thematic area measures steps countries have taken to invest in AI Ethics Research. In particular, evidence must account for (1) frameworks relating to policies, programmes or guidelines that outline funding and practices of AI Ethics Research (2) government actions to support AI Ethics Research by investing programmes that focus on the integration of AI ethics across government departments, agencies and functions, building of research networks, funds or tenders that focus on AI Ethics Research, and (3) non-state actors working to advance AI ethics by the establishment of independent research networks or institutes for AI Ethics Research Units, publications in that area or funding streams dedicated to AI Ethics.

Frameworks may take the form of adopted policies, white papers, or guidelines. Government actions can encompass the creation of draft policies or guidelines, the formation of government entities like oversight bodies tasked with providing policy recommendations on the matter and/or enforcing regulations. Additionally, these actions can involve the execution of policies aimed at addressing the issue and/or promoting awareness or collecting more data about AI Ethics Research. Non-state actors (NSAs) may include non-governmental organisations (NGOs), but also multinational corporations, private military organisations, media outlets, organised ethnic groups, academic institutions, lobby groups, labour unions or social movements working to advance the relevance and importance of AI ethics research

Examples:

Frameworks

The German federal government outlined in an [AI-Strategy](#) that a joint European networking approach should be used to foster research ‘when it comes to establishing clear ethical guidelines, in basic and applied research in the medium and long term’. In the strategy the government also

advocates for an ‘ethics by design’ approach [focusing on technical solutions] and in general committed to provide [EUR 5 billion](#) by 2025 for the promotion of AI. But, one has to be aware of the distinction between general research into AI applications versus those specifically targeted to AI Ethics Research.

Government Actions

Government action visible is for example the financing of competence centers for AI that are funded by the German Ministry for Education and Research and participating federal states and the [DFKI](#) (one of those centers) which also has an [AI Ethics team](#). Moreover, through the German Research Foundation single projects are funded, for example, the [paper](#) from Thilo Hagendorff (2022) ‘Blind spots in AI ethics’ by the Cluster of Excellence Machine Learning—New Perspectives for Science.

Non-State Actors

There are civil society organizations like [Algorithm Watch](#) specifically committed ‘to watch, unpack and analyze automated decision-making (ADM) systems and their impact on society’ and produced for example a global database of [167 ethics guidelines](#). Moreover, the Bertelsmann Stiftung (foundation) runs a project on [Ethics of Algorithms](#) with what they want to contribute to a ‘design of algorithmic systems that leads to more participation for all’. In 2019, Meta provided an [initial funding grant](#) required for the founding of the Institute for Ethics in Artificial Intelligence at Munich University.

Research Guidance

With regard to relevant frameworks, start by identifying any law, policy or programme that calls for investment in AI Ethics Research. It may be helpful to consult National AI Policies or Strategies that call for the development of Ethical Guidelines for AI and ongoing investment in AI Ethics Research as a national priority or policy agenda. Check if there might be commitments to funding AI Ethics Research or programmes of study.

In relation to government actions, start by searching for the plans or programmes driven by government departments or ministries (e.g. Ministry of Economics, Ministry of Technology, Ministry of Science and Innovation, etc.) to allocate funding to AI Ethics and Research. Look for parliamentary or congressional hearings and/or events, monitoring mechanisms or information related to AI Ethics Research. To identify non-state actors actively involved in AI Ethics Research, start by looking at whether any academic institutions and universities have established AI research units and are looking specifically into the ethical implications of AI. Look for civil society organisations and private sector companies that have done work in AI Ethics Research at a national or regional level, specifically in the promotion of ethical guidelines and research on implementing AI ethics into practice. Several of these non-state actors may have also released papers or publications or blogposts on emerging issues in AI Ethics.

Some useful sources

- [The OECD Artificial Intelligence Policy Observatory - OECD.AI](#) provides information on AI policies across the globe - search in the respective guideline whether “ethical research” (or a similar term) is mentioned.

Search

- For details of how government plans to invest in Artificial Intelligence with regard to ethical research.
- Search for details of what funds are planned or already allocated to AI research specifically targeting ethical research.
- General searches for "AI Ethics Research," "researching AI Ethics".
- Search academic search engines (e.g., [Google Scholar](#), [arXiv](#), [ResearchGate](#), etc.) for recent papers on "AI Ethics Research" + [country].
- Non-governmental organizations search for NGOs working on “Ethical AI” or “Ethical Automated Decision Making” and search in their publications for research done by them.
- Search for Investments into AI Ethics Research from national and international companies in your country.
- Search for research units targeted to AI Ethics from big national companies.

Consult

- Government officials in ministries/bodies that are working on the allocation of research funds regarding the topic of AI ethics.