

BUILDING A RESPONSIBLE HUMANITARIAN APPROACH: THE ICRC'S POLICY ON ARTIFICIAL INTELLIGENCE



The rapid development of artificial intelligence (AI) is creating significant opportunities for human and societal progress, including in the humanitarian sector.

The International Committee of the Red Cross (ICRC) is committed to exploring if, how, when and where advances in AI can help us to achieve our mission to protect the lives and dignity of people living amid conflict or other situations of violence.

However, technologies using AI carry significant and multilayered risks that are difficult to identify and mitigate.

The hype, dystopia, politics and competition around AI also make it difficult to apprehend AI in a neutral and objective manner. For all these reasons, we have developed an AI policy to guide our approach when exploring the potential use of AI in support of our work and activities.

The ICRC's AI policy is anchored in a purely humanitarian approach driven by our [mandate](#) and [Fundamental Principles](#). It is meant to help ICRC staff learn about AI and safely explore its humanitarian potential.

This policy is the result of a collaborative and multidisciplinary approach that leveraged the ICRC's humanitarian and operational expertise, existing international AI standards, and the guidance and feedback of external experts.

Given the constantly evolving nature of AI, this document cannot possibly address all the questions and challenges that will arise in the future, but we hope that it provides a solid basis and framework to ensure we take a responsible and human-centred approach when using AI in support of our mission, in line with our 2024–2027 [Institutional Strategy](#).

By making this policy public, the ICRC wants to abide by our continuous commitment to be transparent and accountable for our actions.

ARTIFICIAL INTELLIGENCE POLICY

1. PURPOSE

This policy provides guidance on the ICRC's practice and decision-making in relation to the exploration, deployment, use and management of AI and technologies, solutions and tools related to machine-learning (ML) that support the ICRC's work and activities. It sets out a value-based approach and provides a set of principles to guide our work in relation to AI. It serves as an organizational framework to promote and support a human-centred, responsible and ethical use of AI in humanitarian action. The guidance provided in this policy is aspirational and meant to support the ICRC's continuous efforts to use digital technologies responsibly and in line with our humanitarian mission, principles, values and working procedures. Its implementation will require investments in specific operational and governance capabilities, within the limits of the ICRC's operational and financial capacities.

2. DEFINITION

There is no globally agreed upon definition of AI. Many definitions exist in different contexts and for different purposes and the term continues to evolve in line with technological developments. For the ICRC and for the purpose of this policy, AI is defined as "any system, tool or technology that involves the use of computer systems to carry out tasks that would ordinarily require human cognition, planning or reasoning".¹ ML is defined as "systems that use large amounts of data to develop their functioning and 'learn' from experience". For the purposes of this policy, the term "AI" refers to both AI and ML systems.

3. SCOPE

This policy applies to the use of AI by all staff members and for all activities, internal services and operational responses, including partnerships. It applies to all AI-related technologies, solutions and tools (i.e. generative AI, ML-enabled physical systems and robotics, digital tools featuring AI systems) and domains of application (i.e. data science, computer vision, natural language processing, predictive modelling and forecasting, information and data management, decision-making, programmatic responses, planning and optimization, internal and external communication).

Limitations – The guidance provided in this document is aspirational and general in nature and cannot address all possible strategic, technical or operational questions that the use of AI may trigger. More detailed policy and operational guidance for specific tools and use cases will be developed by departments as the need arises and ICRC practice evolves, in line with the approach and principles provided in this policy. Given the fast pace of AI-related innovations, and the continuous emergence and evolutions of applicable regulatory frameworks and norms, this policy is necessarily a living instrument that will be regularly updated as needed, as assessed by the relevant organizational entities, to reflect developments in the AI landscape and the ICRC's practice.

Sources – The guidance provided in this policy builds on standards, principles and recommendations adopted by leading international organizations and academic and scientific experts on AI (see Annex 1, "Sources and related documents"). It synthesizes international minimum standards and adapts them in line with the ICRC's needs, objectives, mandate, culture of integrity, values, principles and working procedures. This policy is informed by and builds on other internal reference texts related to the ICRC's strategic ambitions with regard to working procedures and information and communication technologies.

As the development and use of AI involve in most cases the use of personal data, the present policy complements, and is without prejudice to, the application of the ICRC's [Rules on Personal Data Protection](#) (the Rules), which are binding on the organization. Specific guidance on the application of the Rules in the context of AI is provided in the [Handbook on Data Protection in Humanitarian Action](#), and complaints for non-compliance with the Rules can be brought to the ICRC Data Protection Commission, through the Data Protection Office.

¹ Where a tool or platform that the ICRC is already using adds AI functionality after adoption, this additional functionality may be merely cosmetic – for example, a support chatbot or design recommendations. At times, the addition of AI may be so fundamental that it alters the nature of the tool or platform – for example, search recommendations being determined by AI rather than by text matching. In these cases, use of the platform may need to be reassessed under this policy.

This policy does not address the ICRC's [position](#) on the use of AI by parties to armed conflict in the context of international humanitarian law.

4. RATIONALE

AI is not a new technology. The ICRC already uses AI tools and technologies to support its activities (e.g. to improve [logistics](#) or provide [health care](#)) and several internal documents and guidelines already inform its practice. Yet, the speed and scale of AI-related innovations in the past few years, the emergence of new technological tools, their simplified accessibility, push-and-pull factors inside and outside the ICRC, and ongoing debates about the opportunities and risks they create – including for humanitarian action – have triggered a need to ensure that the organization is adequately equipped to explore and manage these developments in a timely, coherent and responsible manner.

Internally, multiple AI-related workstreams and initiatives are ongoing and emerging across different departments, projects and *métiers*. Yet, the ICRC currently lacks a common policy framework to ensure that these various activities are organized and managed by a clear, consistent, formal and professional methodology. This policy is intended to help bridge this gap and inform the ICRC's efforts to leverage technological innovation efficiently through effective governance mechanisms and a responsible approach that focuses on generating relevant and sustainable humanitarian impact and upholding the organization's principles and values.

Externally, AI-related debates – on opportunities, risks and regulatory needs and options – have become politicized in the context of global strategic competition between states that disincentivizes controls and limits, market forces that incentivize risk-taking by private sector actors and vastly differing assessments of the relevance and impact AI will have for societies and people. Positions are often influenced by a binary and overly simplistic paradigm that tends to pit AI enthusiasts against pessimists, and utopian against dystopian views, often fuelled by caricatural AI hype or fear-mongering. This includes debates on the alleged risk of “existential harm” for humanity, and the need to consider them in balance with the more predictable, quotidian harms that AI systems are already triggering or amplifying.

This policy sets out a humanitarian, ethical and learning approach that promotes a responsible, safe, coherent and human-centred use of AI. This approach will strengthen the ICRC's governance and its credibility and ability to engage in AI-related debates with humanitarian implications and influence them in the interests of people affected by conflict and principled humanitarian action.

5. VALUE-BASED APPROACH

5.1. A humanitarian and principled approach

Humanity – The ICRC's approach to AI is driven by purely humanitarian objectives and the Fundamental Principle of humanity. It places respect for humanitarian principles, human rights, safety, security, dignity, well-being and agency at the centre of our policies and practices. Through this approach, we are committed to using and exploring the potential of AI solutions to support the quality and efficiency of our work, within the limits of our humanitarian mandate, operational and financial capacities, and internal rules.

- The ICRC focuses on AI solutions that can help alleviate the suffering of people affected by armed conflict and other situations of violence and can improve the efficiency and effectiveness of the operational and organizational systems and processes that support the protection and assistance we provide.
- In line with the “do no harm” principle, the ICRC systematically assesses the potential negative short- and long-term secondary impacts that the use of AI solutions may have on the safety, dignity, agency and respect for the fundamental rights of affected people and takes all possible measures to avoid and mitigate them.
- The ICRC pays specific attention to ensure that the use and development of AI solutions do not jeopardize our ability to demonstrate humanity and empathy through direct and in-person human engagement in the delivery of our external and internal activities and programmes. We focus on AI solutions that can enhance human autonomy and agency and complement face-to-face human interactions, as well as our physical presence and proximity with affected people.

- In light of the above, the ICRC will refrain from using, deploying or providing AI systems that: deploy subliminal, manipulative, or deceptive techniques to distort behaviour and impair informed decision-making; exploit vulnerabilities related to age, disability, or socio-economic circumstances to distort behaviour; use biometric categorization systems that infer sensitive personal attributes, except for labelling or filtering lawfully acquired biometric datasets; rely on or infer social scoring, (i.e. evaluating or classifying individuals or groups based on social behaviour or personal traits); assess the risk of an individual committing a criminal offence based solely on profiling or personality traits, except when used to augment human assessments; compile facial recognition databases by untargeted scraping of facial images from the internet or CCTV footage; or infer emotions, except for medical or safety reasons.

Impartiality – We use AI solutions that strengthen our ability to identify, analyse, prioritize and respond to affected people's needs based on their scale and level of urgency, without making any adverse distinctions on the basis of race, nationality, gender, religion, class, political opinion or other potential sources of discrimination. At all times, the ICRC pays specific attention to affected populations and users that have specific needs and vulnerabilities, and to the specific requirements of the contexts in which AI solutions are considered and used.

- The ICRC acknowledges the significant risks of AI systems that could generate false or inaccurate data and replicate or amplify systemic and societal discriminations, biases and stereotypes – in particular around gender and race. We make proactive and sustained efforts to understand, avoid and mitigate these risks through our choice and use of AI solutions, and by avoiding their use where these biases cannot be effectively avoided or mitigated.
- The ICRC makes all possible efforts to ensure that the AI solutions and tools we use are inclusive by design and in their impact, and that they enable fair, non-discriminatory, equal and equitable delivery of assistance and services to all affected people and users, taking into account the varying degrees of need and capacity.

Neutrality and independence – AI, like other technologies, is not in and of itself good, bad or neutral. It can represent or promote the personal or political values of the individuals and/or companies who create, design or promote them. AI is an object of increasing politicization and competition between states at the global level, as well as private-sector and non-state actors. Additionally, many AI solutions are provided by tech companies that are increasingly active in the political domain or taking sides in ongoing armed conflicts. Therefore, as an organization we pay specific attention and take all feasible measures to ensure that our choices of AI tools and providers cannot be instrumentalized to challenge our neutrality and independence.

- The ICRC uses our ethical frameworks ([Framework on Engaging with the Private Sector to Mobilize Support](#), [Ethical principles guiding the ICRC's partnerships with the private sector](#)) and procurement processes ([Supplier Code of Conduct](#)) to ensure as much as possible that we do not use AI providers that (i) are directly engaged in activities contributing to armed conflict or closely associated with military activities of parties to conflict, or (ii) promote particular political interests or values that are contradictory to the ICRC's humanitarian mandate and ethical values.

When relevant, we strive to explore available, safe and adequate free and/or open-source AI solutions to mitigate risks of increased dependency on politicized or proprietary tools and systems that could (i) jeopardize the perception of our neutrality and independence from entities involved in political or conflict-related controversies; (ii) endanger our operational independence and autonomy; or (iii) restrict our ability to opt for alternatives without incurring disproportionate costs.

5.2. An enabling and learning approach

Fostering ethical and impactful innovation – The ICRC is committed to creating an organizational environment that facilitates and supports the safe and responsible exploration and use of impactful, sustainable and innovative AI initiatives and practices.

- The ICRC systematically adopts a context-based and ecosystem approach, relying on the analysis of the interconnectedness of AI-supporting environments, data and systems, and human and social factors. We leverage a responsible use of data; strong governance, classification, regulation and control mechanisms; and continuous evaluation and adjustments through a coordinated, collaborative, multidisciplinary and multistakeholder approach.
- The ICRC maintains a long-term perspective to ensure AI research and development investments are financially responsible and sustainable, focusing on relevance, impact and effectiveness, and also considers added-value, interoperability, scalability and cost-efficiency (including maintenance, training, verification and compliance-related costs).
- The ICRC pays specific attention to the alignment between the impact of considered AI solutions and the organization's short-and long-term objectives, operational and institutional constraints and requirements, and related programmatic and organizational risks at the security and reputational levels.

Competence and capacity building – The technical complexity and potential risks attached to AI systems and their functioning requires a significant level of knowledge and expertise to select, operate and manage them effectively and responsibly. The ICRC is committed to ensuring that staff members across the different levels of the organization are properly informed, equipped and supported to explore, develop and use AI responsibly and in line with our humanitarian mandate, Fundamental Principles, Code of Conduct and culture of integrity.

- At the institutional level, the ICRC incentivizes and enables adequate levels of AI and digital literacy, and training and upskilling for staff members at all levels. We leverage peer-review, support and collaboration with trusted and qualified external experts and partners to inform and sustain a continuously learning approach to the use of AI in humanitarian action.
- The ICRC invests in adapted learning and development programmes to support our staff members' understanding of what AI systems are, how they function and what opportunities and risks they bring in relation to staff members' professional activities. This includes awareness and understanding of the ethical considerations and different internal rules and regulations that apply to AI internal and external tools and technologies. When necessary, the ICRC restricts access to AI tools to specifically trained and authorized personnel.

6. GUIDING PRINCIPLES

6.1. Proportionality, precaution, and “do no harm”

The ICRC considers AI as a potential tool to help better respond to humanitarian and organizational needs, based on a problem identification and analysis that systematically considers the specific level of risk, effectiveness of expected results and alternative solutions in context.

- Potential AI solutions are only considered and used when necessary and adequate to achieve a positive impact and respond to clear, pre-defined objectives. Considerations of novelty and competitiveness are not in themselves justification for the use of AI solutions.

The ICRC's choice and use of AI systems is guided by a precautionary, pragmatic and science-based approach that carefully considers the multidimensional risks that they could create for affected populations, users and the organization, including from a legal perspective. Potential AI solutions are reviewed and assessed using a differential approach that distinguishes their type, purpose, scope, place of deployment and impact and associated risk level (high, medium or low).

- Specific additional risk analysis and mitigating measures are put in place when the AI solutions under consideration present higher levels of risks and/or have a potential impact on the safety, dignity and rights of affected populations and users; when they involve the use of personal, confidential, strictly confidential or sensitive internal data; or when they involve legal or reputational risks for the organization.

- When the potential benefits of AI outweigh identified risks, solutions under consideration are thoroughly tested before being rolled out. The necessary evidence is gathered through scientific experimentation and “sandbox” testing to ensure, as far as possible, the effectiveness and safety of the solutions in context.
- Systematic review and auditing systems supported by regular internal and external feedback mechanisms are put in place to ensure a continuous and multidimensional risks analysis based on their likelihood and the severity of their potential impact for affected populations, users or the organization, including from a legal perspective. When adequate and effective mitigating measures to identified risks are not available or sufficient, the ICRC rejects or terminates the use of relevant AI systems and solutions.

6.2. Safety and security

Considering the multiple and significant risks that AI solutions present for the confidentiality and protection of the data that the ICRC collects, processes and manages for the purposes of our activities – and the potential negative or harmful consequences arising from them – the ICRC explores and uses AI solutions through a human-centred approach anchored in the highest possible levels of safety, security and privacy in their design and ensuring sufficient human oversight and verification of AI outputs throughout their lifecycle.

- Risks of misuse, abuse, infiltration and attacks against AI systems and potential related harms and damage are systematically considered, and the highest possible standards of data protection and cybersecurity measures are put in place to prevent and mitigate them, in line with assessed risk levels.
- Additional systematic data quality control and meaningful and effective human oversight mechanisms are put in place to control and review relevant training models, algorithmic systems, inputs and outputs throughout the lifecycle of AI solutions used to support operational activities and decision-making processes that have an impact and present a risk of harm for affected people and users.

The ICRC ensures that the use of internal and external AI solutions are aligned with the rules and the principles of lawfulness, fairness, transparency, purpose limitation, data minimization, accuracy, storage limitation, security and accountability.

- At all times, the ICRC ensures that confidential or sensitive internal ICRC information or personal data are handled with the utmost care and are not used in or transferred to non-approved, external online AI systems, in line with the duty of discretion stated in the ICRC’s Code of Conduct.
- Mechanisms are put in place to ensure that the data required for AI systems is collected, used, shared, archived and deleted in ways that are consistent with the right to privacy of affected people and users.
- Systematic data protection impact assessments are carried out prior to the acquisition, deployment or use of considered AI solutions, and when relevant and feasible, solutions are put in place to restrict processing and enable rectification and/or erasure of the data used.

6.3. Transparency and “explainability”

Transparency and “explainability” are essential elements of a safe, ethical and responsible approach to AI, and to ensure respect for the ICRC’s principles, rules and values. The ICRC therefore ensures at all times the highest possible levels of transparency in our use and management of AI solutions, in accordance with the context, risks and impact of the systems and tools considered, and with due consideration to the principles of safety, security and privacy.

- As much as feasible and in accordance with the applicable risk level, context and use case, meaningful traceability, disclosure and notification measures are put in place in relation to the use of AI systems that support the ICRC’s activities and decision-making processes, including, as appropriate, releasing AI systems as open-source.

- Relevant and intelligible information about the use of AI systems is systematically provided to affected people and users when these systems could have an impact on them or their rights, or when decisions with significant consequences for the organization are concerned.
- This includes adequate and meaningful information about why, when and how AI solutions are used, and explanations on the causal link between AI inputs and outputs and the functioning of algorithmic processes.

6.4. Responsibility and accountability

The ICRC ensures the highest possible levels of responsibility and accountability in the use and management of AI solutions through continuous monitoring, assessment and reporting on the procurement, use, management and review of AI solutions – including through publicly accessible reporting, when relevant.

- When relevant and feasible, systematic impact, verifiability and replicability assessments are put in place and supported by effective evaluation and auditing processes to ensure the ICRC's AI systems and tools comply with this policy.
- When relevant and feasible, meaningful and accessible engagement mechanisms are put in place to consult affected people and users on the design, use and impact of considered AI solutions. Provided that it does not prevent affected people from benefiting from equivalent opportunities, the possibility of opting out of AI systems and having access to alternative solutions and complaint and remedy mechanisms are made available to them.
- Responsibility and accountability for decision-making supported by AI always lies with the ICRC and/or the staff using the tools or system concerned, in line with the ICRC's Code of Conduct and other applicable internal rules.

The ICRC puts in place effective governance mechanisms to ensure that our use of AI solutions is aligned with our broader organizational obligations and commitments and complies with applicable rules and regulations.

- Competent and qualified staff and organizational entities are appointed and equipped to ensure good governance and effective enforcement and compliance with this policy and other related internal rules and regulations applicable to AI systems.
- Decisions on the use or provision to others of AI systems, tools or technologies should be informed by legal considerations, and legal advice should be obtained during any AI-related decision-making process.
- Awareness-raising, training and compliance efforts are made to ensure that (i) the ICRC's use of AI solutions does not breach applicable international, regional and national legal restrictions, and that (ii) ICRC staff and users do not input any critical humanitarian information, other confidential and/or strictly confidential information, or any proprietary information protected by intellectual property rights when using external or commercial online AI services.
- All possible efforts are made to integrate and mitigate the environmental impact and carbon footprint of considered AI solutions through a life-cycle approach. This includes favouring solutions and tools that have a lighter footprint, in line with the ICRC's Plan of Action under the [Climate and Environment Charter for Humanitarian Organizations](#) and the ICRC's [Framework for Sustainable Development](#).

SELECTED SOURCES

Below are some of the many sources on AI ethics that the ICRC has drawn on to draft this policy.

Recommendations on the Ethics of Artificial Intelligence, UNESCO, available at: <https://unesdoc.unesco.org/ark:/48223/pf0000381137>

“Principles for digital development”, available at: <https://digitalprinciples.org/principles/>

Understanding Artificial Intelligence Ethics and Safety: A Guide for the Responsible Design and Implementation of AI systems in the Public Sector, The Alan Turing Institute, available at: <https://www.turing.ac.uk/news/publications/understanding-artificial-intelligence-ethics-and-safety>

A Framework for the Ethical Use of Advanced Data Science Methods in the Humanitarian Sector, Data Science and Ethics Group, available at: <https://www.hum-dseg.org/sites/g/files/tmzbdl1476/files/2020-10/Framework%20for%20the%20ethical%20use.pdf>

“Ethics guidelines for trustworthy AI”, European Union Commission, available at: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

European Union Artificial Intelligence Act, available at: <https://artificialintelligenceact.eu/the-act/>

White paper on trustworthy artificial intelligence, Chinese Academy of Information and Communication Technology, available at: <https://cset.georgetown.edu/publication/white-paper-on-trustworthy-artificial-intelligence/>

“Declaración de ética IA-LATAM para el diseño, desarrollo y uso de la inteligencia artificial”, IA-LATAM, available at: <https://aichallenge.utec.edu.uy/community/foro-bai-biases-in-ai-consultas-generales/declaracion-de-etica-ia-latam-para-el-diseno-desarrollo-y-uso-de-la-inteligencia-artificial/>

Continental Artificial Intelligence Strategy: Harnessing AI for Africa’s Development and Prosperity, African Union, available at: https://au.int/sites/default/files/documents/44004-doc-EN-_Continental_AI_Strategy_July_2024.pdf

ASEAN Guide on AI Governance and Ethics, ASEAN, available at: https://asean.org/wp-content/uploads/2024/02/ASEAN-Guide-on-AI-Governance-and-Ethics_beautified_201223_v2.pdf

Digital Ethics: A Guide for Professionals of the Digital Age, CIGREF, available at: <https://www.cigref.fr/wp/wp-content/uploads/2019/02/Cigref-Syntec-Digital-Ethics-Guide-for-Professionals-of-Digital-Age-2018-October-EN.pdf>

Artificial Intelligence and Data Protection, Council of Europe, available at: <https://rm.coe.int/2018-lignes-directrices-sur-l-intelligence-artificielle-et-la-protecti/168098e1b7>

Governing Artificial Intelligence: Upholding Human Rights and Dignity, Data and Society, available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

“The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems”, Amnesty International and Access Now, available at: <https://www.torontodeclaration.org/declaration-text/english/>

“How to design AI for social good: Seven essential factors”, Science and Engineering Ethics, available at: <https://link.springer.com/article/10.1007/s11948-020-00213-5>

“Humanitarian AI revisited: seizing the potential and sidestepping the pitfalls”, HPN-ODI, available at: <https://odihpn.org/Humanitarian-AI-revisited-Seizing-the-potential-and-sidestepping-the-pitfalls-Humanitarian-Practice-Network>

We help people around the world affected by armed conflict and other violence, doing everything we can to protect their lives and dignity and to relieve their suffering, often with our Red Cross and Red Crescent partners. We also seek to prevent hardship by promoting and strengthening humanitarian law and championing universal humanitarian principles.

