SYMPOSIUM

ON THE RESPONSIBLE USE OF TECHNOLOGY IN HUMANITARIAN ACTION

4 - 5 DECEMBER 2025 | BEIJING, CHINA

POST-SYMPOSIUM REPORT

REPORT

清华大学战略与安全研究中心
CENTER FOR INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

ICRC

TABLE OF CONTENTS

ORGANIZERS OF THE SYMPOSIUM

CENTER FOR INTERNATIONAL SECURITY AND STRATEGY (CISS) OF TSINGHUA UNIVERSITY

Established on 7 November 2018, the Center for International Security and Strategy (CISS) of Tsinghua University is a think tank in the field of international security and strategy. CISS has two objectives: one is to follow the changes in global dynamics, offering policy suggestions for decision-making through research on topics of international order, international relations, and security and strategic studies while conveying rational and professional messages to society; and the other is to elucidate and disseminate China's perspectives and policy propositions through various forms of international exchanges and cooperation, to enhance the international community's understanding of China, and to improve Tsinghua's global influence in the fields of international relations and strategic studies.

The CISS consists of research divisions including the Research Project on the US and Europe, the Research Project on Global Governance, the Research Project on Eurasia, and the Research Project on Artificial Intelligence Governance. It sets up "China Forum" – a sub-brand committed to international communication – as well as organizing the "CISS Youth" Research Exchange Program across top universities at home and abroad, and overseeing the Tsinghua University Student Association of International and Strategic Studies (SAISS) and the Tsinghua University Student Association for China-U.S. People-to-People Exchange.

INTERNATIONAL COMMITTEE OF THE RED CROSS

The International Committee of the Red Cross (ICRC) is an impartial, neutral and independent organization whose exclusively humanitarian mission is to protect the lives and dignity of victims of armed conflict and other situations of violence and to provide them with assistance.

The ICRC's Regional Delegation for East Asia was established in Beijing in 2005. As China has steadily taken on a prominent role in international relations over the past decades, the delegation's mission is to encourage and support China's contribution to the international humanitarian cause. The delegation fosters and sustains strategic dialogue with Chinese authorities, military and police, the Red Cross Society of China, think tanks and academic institutions, as well as media and the corporate sector, on humanitarian issues and the promotion of international humanitarian law. The delegation also serves as one of the ICRC's largest procurement hubs worldwide.

The ICRC's Global Cyber Hub in Luxembourg was established in response to the growing impact of digital transformation on armed conflict and humanitarian action. The hub enables the ICRC to integrate cyber and digital dimensions into its neutral and independent humanitarian operations, while prioritizing the needs of people affected by conflict. Its mission is to ensure that the ICRC remains a neutral, impartial and independent humanitarian actor in cyberspace and throughout the digital era. This mission is pursued through strategic research, legal and policy engagement, outreach initiatives and operational innovation.

EXECUTIVE SUMMARY

The Beijing Symposium on the Responsible Use of Technology in Humanitarian Action brought together more than 80 humanitarian practitioners, academic researchers and industry representatives from almost 20 countries to discuss how emerging technologies are reshaping humanitarian action and what it takes to use them responsibly in crisis settings.

Roundtable 1 on responsible artificial intelligence (AI) focused on governance, ethics and accountability. Participants agreed that human control and accountability remain the anchor of AI governance, since legal and ethical responsibility cannot be shifted to machines. Despite differing regulatory approaches across regions, the discussion showed convergence on shared principles such as safety, transparency, human oversight and inclusiveness. At the same time, participants stressed that principles alone are no longer sufficient: responsible AI requires concrete safeguards, testing and evaluation practices, clear responsibility chains, and enforceable accountability.

Roundtable 2 on tech for good discussed how technology can support effective and collaborative crisis response. Participants emphasized that robust digital infrastructure is the most important form of "tech for good", including connectivity, power and offline-capable systems. The discussion then turned to the contested neutrality of technology, with some participants arguing that code can be neutral and others stressing that technology always reflects power and incentives, creating dependencies that humanitarian actors must manage. Participants agreed that dialogue with the tech industry is essential to make "tech" a force for "good" – something that requires responsibility, safeguards and exit strategies, as well as the courage to say "no" to technologies that create harm or dependency, or that undermine humanitarian principles.

Roundtable 3 on AI safety and security examined how to responsibly deploy AI in crisis settings. Participants agreed that AI risks are highly context-dependent, with lower-risk applications in logistics or back-office functions and higher stakes for systems that influence triage or early warning. Because AI systems directly embed data into algorithms, safety-by-design require hard choices around data, transparency, and model training. The discussion highlighted that staff and communities must be equipped to understand biases and risks, and to question and distrust AI systems when needed, and that humanitarian organizations cannot manage these risks alone, making cooperation with academia, industry and the public sector essential.

Roundtable 4 on digital trust explored how to build and sustain confidence in a connected world. Participants described trust as relational and practice-driven, and agreed that trust depends on consistent, transparent and predictable behaviour. In conflict settings, digital trust is particularly fragile, as misuse or leakage of sensitive data could directly endanger people and undermine humanitarian access and acceptance. Looking ahead, participants stressed that building and sustaining trust requires preparedness for, and a forecast of, future risks, including contingency planning for system failure or misuse and practical initiatives such as the digital red cross, red crescent, and red crystal emblems.

INTRODUCTION

The humanitarian sector stands at a crossroads in today's age of digital and other new technologies. From connectivity and data management to artificial intelligence (AI) and satellite systems, technology shapes how humanitarian organizations assess needs, plan operations, communicate, deliver assistance and conduct protection work. These tools offer gains for humanitarian action in terms of speed, scale and precision, yet they also introduce new risks, dependencies and dilemmas. The rapid spread of harmful information online, for example, fuels division and incites violence, amplifying the human cost of conflict. Data breaches risk undermining the trust that civilians and parties to armed conflicts place in humanitarian organizations, thus limiting humanitarian access to people affected by armed conflict and other violence, and potentially endangering the safety of humanitarian personnel. The development of military technology and autonomous weapon systems poses a great number of humanitarian, legal and ethical challenges on the battlefield.

As humanitarian action becomes increasingly mediated by digital systems mainly developed outside the sector, questions of governance, accountability, neutrality and protection have moved from the margins to the centre of humanitarian practice. This means that the humanitarian sector cannot solve these technical challenges on its own. It requires networked, sustained engagement with stakeholders across industry, academia, civil society and the humanitarian sector.

To this end, in 2022, the International Committee of the Red Cross (ICRC) – through its Global Cyber Hub in Luxembourg – initiated a Symposium series to address the challenges for humanitarian action in cyberspace and the digital age. The series serves as an open, trusted space for cross-sectoral and interdisciplinary dialogue at the intersection of new technology, cybersecurity, data protection and humanitarian action, with sessions held under the Chatham House Rule. Earlier events in the series were hosted in Luxembourg in 2022 and 2024, in Nairobi, Kenya, in November 2024, in Geneva, Switzerland, in June 2025, and in Vienna, Austria, in November 2025. Further details about the Symposium series are available on the ICRC website.

The following report is a detailed summary of the key findings and outcomes of a Symposium held as part of the series in Beijing, China, in December 2025.

THE BEIJING SYMPOSIUM ON THE RESPONSIBLE USE OF TECHNOLOGY IN HUMANITARIAN ACTION

The Beijing Symposium, which focused on the responsible use of technology in humanitarian action, marked an important step in the series in terms of both content and collaboration. Co-organized by the ICRC's Global Cyber Hub in Luxembourg, the ICRC's Regional Delegation for East Asia in Beijing and the Center for International Security and Strategy (CISS) at Tsinghua University, it situated humanitarian debates on technology within an academic environment that works at the confluence of security, technology policy and global governance. The event drew together more than 80 humanitarian practitioners, policy experts, industry representatives and researchers from almost 20 countries in Asia, Europe, Africa and North America, helping to connect operational concerns with wider discussions on AI governance and safety, cyber risks, and the digital transformation.

As the largest ICRC technology-focused event held in China to date, it also fostered a conducive environment for engagement with China's research and technology community. The Symposium featured a strong academic presence and a clear focus on long-term research, governance models and strategic risk thinking. At the same time, discussions stayed grounded in practical humanitarian realities, linking questions of digital trust, neutrality and independence with the opportunities and risks digital technologies pose to civilians in conflict settings.

OPENING REMARKS AND KEYNOTES

The opening session framed emerging technologies as both enablers of humanitarian action and sources of significant ethical, operational and governance risks. The opening session of the Symposium, moderated by **Xiao Qian**, Deputy Director of the CISS, featured reflections by representatives of the organizers:

> *Throughout the long arc of human civilization, technology has always been a driving force for social progress. … But precisely because technology carries so much promise, we must ensure it is used responsibly. Without proper norms and ethical frameworks, the application of technology in humanitarian settings may fall short of expectations or even introduce new risks to vulnerable populations.*
>
> *— Professor Yang Bin*

> *The kind of [digital] dilemmas we grapple with in situations of war are like a magnifying glass that bring out certain issues in a very stark way.*
>
> *— Balthasar Staehelin*

Professor Yang Bin, Vice Chancellor of the Tsinghua University Council, and **Balthasar Staehelin**, Personal Envoy of the ICRC President to China and Head of the ICRC Regional Delegation for East Asia. Together, they situated the Symposium at the crossroads of rapid technological change and humanitarian responsibility, emphasizing that while AI, big data and cloud computing are transforming societies and humanitarian operations at unprecedented speed, their adoption in crisis contexts demands careful governance. Yang Bin highlighted China's growing engagement in ethical AI development and global governance initiatives, which had become one of the incentives to host the Symposium in Beijing. Balthasar Staehelin described humanitarian settings as a magnifying lens for broader digital dilemmas, where issues such as the sensitivity of data, algorithmic bias, online misinformation and digital inclusion could be directly connected with physical harm. Balthasar Staehelin described humanitarian settings as a magnifying lens for broader digital dilemmas, where issues such as the sensitivity of data, algorithmic bias, online misinformation and digital inclusion could be directly connected with physical harm. Both stressed the importance of cross-sector and cross-regional dialogue, including engagement with Chinese and Global South perspectives.

The keynote speakers expanded these themes from complementary angles. **Gong Ke**, Executive Director of the Chinese Institute of New Generation Artificial Intelligence Development Strategies, emphasized that AI is already being used in humanitarian contexts to respond to crises more quickly and effectively.

> *AI can help us respond to crises better, quicker and more accurately. But this potential will only be realized if we ground AI firmly in the core humanitarian principles of humanity, impartiality and independence. Vulnerable people should be the first to benefit from AI and not the last.*
>
> *— Gong Ke*

He called for further prioritization of AI applications tailored to humanitarian needs and stressed the importance of developing AI in a humane and accountable manner. Guardrails could include embedding ethical principles from the initial design and deployment stages throughout AI's entire lifecycle, rather than addressing ethics only after incidents occur. He also proposed establishing AI incident learning platforms, no-fault AI compensation funds and an accountability framework for responsible AI. **Dai Huaicheng**, Secretary-General of the China Arms Control and Disarmament Association, warned of the continued escalation of the AI arms race, stressing that military uses of AI must comply with international humanitarian law and calling for global governance, restraint by major powers and clear limits on unacceptable uses. **Professor Zeng Yi**, founding Dean of the Beijing Institute of AI Safety and Governance, cautioned that AI models would not become safer by default and called for global red lines on high-risk and military uses of AI. **Els Debuf**, Head of the ICRC's Global Cyber Hub in Luxembourg, highlighted how digital technologies have become a lifeline for humanitarian response, and called for early integration of humanitarian principles into technology design, as well as for sustained cross-sector cooperation and continued dialogue with Chinese partners.

OPENING PANEL: SETTING THE SCENE FOR THE SYMPOSIUM

The opening panel, moderated by Balthasar Staehelin, translated the keynote messages into a more practical discussion on technology governance and implementation. **Chen Qi**, Deputy Director of the CISS, argued that prohibiting lethal autonomous weapon systems was unrealistic, and instead stressed the need to embed international humanitarian law into these systems and retain human decision-making at all costs.

> *Instead of only thinking about optimization through AI, a more interesting challenge is to rethink institutions and organizations with the availability of these systems [in mind]. How can an organization that is 160 years old … be rethought in terms of having AI to serve the purpose it serves in a different way through the use of AI.*
>
> *— Professor Andrea Cavallaro*

> *The technological fragmentation … we are facing in the humanitarian sector today [is] a mirror of the state of the early environment of internet development. It is 'walled gardens' in which various tech companies have developed powerful but closed data systems. In crisis scenarios, this data non-interoperability constitutes one of the greatest humanitarian challenges.*
>
> *— Xiaodong Lee*

Professor Andrea Cavallaro at the Swiss Federal Institute of Technology in Lausanne, underscored the fragility of sophisticated AI systems despite huge technological advancements, explaining that a large language model with 7 billion parameters could collapse if only four of those parameters were manipulated. He observed how academia and humanitarian action operated on vastly different time horizons – 72 months versus 72 hours – yet argued for a middle ground between long-term "imagination" and short-term "optimization". **Xiaodong Lee**, Founder and CEO of the Fuxi Institution for Digital Economy and member of ICRC's Global Advisory Board on digital threats during conflict, highlighted the risks arising from the "walled gardens" of technological fragmentation and proposed two guiding principles from internet governance for the governance of AI: rough consensus, where progress is driven by broad, practice-based agreement rather than formal unanimity, and running code, which prioritizes testing, learning and incremental improvement over waiting for complete regulatory solutions. **Blaise Robert**, Global AI Adviser of the ICRC, wrapped up the points made by the previous

speakers by re-emphasizing the dual-use nature of digital technologies and the need to balance immediate operational needs with longer-term governance, using humanitarian principles as a practical guide.

Together, the opening remarks, keynotes and panel set the intellectual frame for the four thematic round tables at the centre of the Symposium – which were governed by the Chatham House Rule – as well as foregrounding the tensions between innovation and restraint, and establishing responsible design, governance and cross-sector dialogue as prerequisites for the responsible use of technology in humanitarian action.

ROUNDTABLE 1

RESPONSIBLE AI: GOVERNANCE FOR ETHICS, SAFETY AND ACCOUNTABILITY

BACKGROUND

Humanitarian organizations increasingly rely on digital tools to communicate and coordinate their activities, and to offer digital services directly to communities in need. In doing so, they depend on a range of off-the-shelf commercial software products, often developed by private technology companies, which give rise to specific challenges. The values and interests driving tech companies may not always be aligned with those of humanitarian organizations. For example, the limited transparency of digital tools can be at odds with the transparency expected of humanitarian organizations vis-à-vis affected populations. In addition, as recent events have highlighted, private companies may base their decisions on geopolitical factors and be subject to international sanctions. If humanitarian organizations rely solely on solutions developed by these companies, their neutrality and independence may be questioned, and their operational capacity reduced. It is also important to consider that commercially developed tools have limited interoperability which may result in a situation of "vendor lock-in", making humanitarian organizations even more exposed to unilateral decisions taken by private companies, such as pricing changes.

SUMMARY OF THE ROUNDTABLE DISCUSSION

The conversation unfolded along two intertwined threads: how to maintain human responsibility as systems grow more autonomous, and how to build governance frameworks that are both globally coherent and locally operational. In opening statements, participants highlighted the limitations of AI systems, the risks posed by such systems, and strong divergences in governing philosophies – precautionary in some regions, innovation-driven in others – as well a widening gap in AI capabilities, and global tensions that risk fragmentation across regions. At the same time, participants underscored broad agreement globally on a human-centred approach to AI, on accountability and transparency as shared principles, and on the idea that responsibility and control must remain with people, not machines.

The conversation then turned to the gap between principles and implementation. Participants noted that while many governance initiatives and responsible AI declarations already exist, the priority now lies in implementation – namely, establishing concrete safeguards, clear chains of responsibility, robust testing and evaluation practices, and mechanisms to recognize limitations rather than blindly relying on AI outputs. The complexity of distributed decision-making in AI systems was highlighted as a challenge, and participants debated whether this meant that accountability was impossible or if AI makes it easier to escape responsibility. Several speakers rejected the idea of an "accountability gap", arguing that the humans who were designing, authorizing and deploying these systems were doing so with intentionality, and they must always be held accountable.

A discussion emerged around the issue of dual-use technologies – a particularly relevant topic in the context of AI systems in humanitarian action and conflict settings. Participants described how AI is increasingly being integrated into weapon systems, raising concerns about increased unpredictability in their engagement, which may undermine the ability to ensure compliance with applicable legal obligations. Participants also noted the growing reliance on AI within critical digital infrastructure, raising concerns that attacks on data centres or power networks could blur the lines between civilian and military objects. Similar issues with dual-use technologies also exist in satellite systems, cybersecurity and biological research. This led one participant to propose explicitly considering "AI-driven crises" and adapting disaster-management tools to the AI context, including incident-sharing, monitoring, drills and post-crisis investigations.

The discussion returned to the topic of inclusion several times. A number of participants stressed the widening "AI divide", which means that a large share of the world's population has limited access to AI technologies, and that low-resource countries lack the capacity to develop AI systems or meaningfully shape AI governance. Others pointed out that actors from China, African countries and the broader Global South are not yet being adequately heard in global debates. This observation strengthened calls to treat AI as serving global public interests and to design inclusive AI governance.

KEY TAKEAWAYS

Human control and accountability remain the anchor of responsible AI governance. Regulations and norms bind humans, not machines, and AI systems cannot be allowed to hinder or replace human judgement or blur the chains of responsibility. Legal and ethical responsibility must always remain with people.

Although governance models differ across regions, the roundtable identified growing convergence around core values such as safety, transparency, human oversight and inclusiveness. Collaboration among governments, industry, academia and humanitarian actors remains essential to avoid fragmentation and to ensure the responsible development and use of AI.

Principles are no longer enough; governance requires concrete tools, procedures and safeguards. Several participants stressed that the global community must move from broad principles (e.g. safety, transparency and responsibility) to concrete operational measures such as testing and evaluation, traceability mechanisms, a clear allocation of responsibility, and the enforcement of accountability across the AI lifecycle.

ROUNDTABLE 2

TECH FOR GOOD: HOW CAN TECHNOLOGY SUPPORT EFFECTIVE AND COLLABORATIVE CRISIS RESPONSE?

BACKGROUND

Leading tech firms have increasingly embraced "tech for good" initiatives by rolling out programmes for social responsibility, digital inclusion and sustainability. For humanitarian action, this creates opportunities to adapt innovations such as medical imaging support tools, supply-chain traceability solutions, and satellite or mobile connectivity for crisis response. Against this backdrop, the roundtable set out to examine how technology can meaningfully support effective humanitarian response, which technological innovations were most relevant in fragile contexts, and how technology could be harnessed responsibly and sustainably through partnerships between humanitarian organizations, governments, academia and the private sector.

SUMMARY OF THE ROUNDTABLE DISCUSSION

The discussion repeatedly returned to a fundamental point: for humanitarian organizations operating in crisis settings, "tech for good" starts with basic infrastructure, not advanced tools. Several participants emphasized that connectivity and power systems, as well as those that operated without the internet or electricity, were foundational to humanitarian action. One speaker described the critical importance of connectivity for people in crisis – comparable to water and electricity – and argued that without it, more advanced tech solutions would lose their value.

From there, the roundtable turned to the question of neutrality in tech, sparking a lively debate. Some speakers argued that technology, at least at the level of code, was neutral. Others argued that technology was never neutral, pointing to design philosophies, company incentives, geopolitical context, export controls and sanction regimes as factors shaping how tools were developed and deployed. When connecting this to humanitarian action, participants raised concerns about how the tech neutrality in question could be managed to avoid conflicting with the neutrality of humanitarian organizations.

This led directly to a broader discussion of power, accountability and trust. Several speakers warned that digital tools could reshape relationships between and among tech companies, humanitarian organizations and affected communities. Improving dialogue with the tech industry was therefore cited as being essential to ensure that the technologies and products designed by tech companies could be a force for good rather than harm. Biometrics were cited as an example of technology that could invert accountability, making affected people prove themselves to organizations rather than strengthening mechanisms that held organizations accountable to the people they serve. Participants also referred to ethnographic research showing that digital tools could affect how communities perceive and trust humanitarian organizations, particularly when interactions shifted from face-to-face to screens and systems.

The discussion was grounded with concrete examples illustrating both potential and risk. One participant described how, following the Wang Fuk Court fire tragedy, a digital tool had been built to connect affected people with volunteers. In that case, developers had deliberately decided against real-name registration to avoid placing additional burdens on people in distress and raising further data-protection concerns. Participants also highlighted cautionary cases, including biometric databases that lacked proper ways to erase data when organizations withdrew, leading to the potential exposure of highly sensitive data.

In closing, and reflecting on the examples highlighted, the discussion converged on the view that good intentions alone did not make technology "good". Participants stressed the need for accountability, ethical design from the outset, and the capacity to refuse technologies that create dependencies, expose sensitive data or undermine trust. Several speakers called for greater collaboration between and among humanitarian organizations, governments, academia and the private sector. Participants concluded that technology should be judged not by its sophistication, but by whether it works under crisis conditions and genuinely serves people in need – which sometimes meant that the simplest solution is the most appropriate.

KEY TAKEAWAYS

Robust digital infrastructure is the most important "tech for good". In crisis settings, "tech for good" starts with resilient basics: connectivity, power and offline-capable systems. Without robust infrastructure, even advanced technologies fail to deliver value.

The neutrality of technology is contested. Some participants argued that technology could be neutral, while others argued that technology could never be neutral and always carries political, legal and power implications. For humanitarian organizations, protecting their own neutrality relies on managing dependencies, vendor lock-in and the risk of creating long-term reliance for affected populations.

Good intentions are not enough. Goodwill and innovation do not guarantee positive outcomes. Meaningful dialogue between the tech industry, the public sector and humanitarians is essential to make "tech" a force for "good" – which requires responsibility, safeguards, exit strategies and the courage to say "no" to technologies that create harm or dependency, or that undermine humanitarian principles.

ROUNDTABLE 3

AI SAFETY AND SECURITY: PRACTICAL SAFEGUARDS FOR RESPONSIBLE DEPLOYMENT IN CRISIS SETTINGS

BACKGROUND

As AI systems are deployed worldwide, they are also increasingly being used in crises. Models such as open-source medical assistants and satellite-imagery classifiers, or the use of algorithms to monitor conflicts, illustrate how AI can enhance humanitarian response. Yet high-stakes environments also expose serious risks: bias and discrimination, fragile data protection, opaque decision-making, and the potential for errors to directly affect the safety and dignity of vulnerable populations. The fact that most AI research and technologies originate in the commercial sector adds an additional layer of complexity when it comes to safely deploying these tools in humanitarian contexts. It is therefore essential to ensure that comprehensive practical safeguards – such as data governance, transparency and robustness – are put in place. The roundtable set out to examine how to assess risks linked to the use of AI in crisis contexts, what red lines and mitigation strategies were needed, and how multi-stakeholder cooperation could help ensure AI served affected communities.

SUMMARY OF THE ROUNDTABLE DISCUSSION

The roundtable opened with agreement that AI could meaningfully support humanitarian operations, but that its risks varied sharply by context. Tools used for logistics or back-office management introduce less severe risks than systems that influence triage, early warning or aid distribution, with tools in this latter category demanding strict safeguards because they compress decision time and could entrench biases and existing discrimination. While a context-specific approach to AI design and deployment was cited as being useful in principle, its impact might be diluted by the need to create perfect taxonomies that capture every possible context and its associated risk factors. Instead, participants argued that risk frameworks should align on clear principles – do no harm, neutrality and allowance for human override – and that these principles could then be applied flexibly to specific use cases.

Participants highlighted that many risks – such as cybersecurity, data protection and fragile infrastructure – are not unique to AI. But since AI systems directly embed data into algorithms, they introduce additional vulnerabilities: training data shapes outputs, could expose sensitive information and could be manipulated. Participants called for risk-based procurement that evaluates vendors' data practices, for restrictions preventing the reuse of humanitarian data, and for preferences for models that allow for inspection and constraint. They stressed data minimization (i.e. collecting only what is strictly necessary) as a core safeguard and acknowledged how hard this was to implement in real emergencies. Consent, for example, was mentioned as being essential in principle but as something impossible to obtain in acute crises, with several participants noting that deciding when to make exceptions must remain a human call, not an automated one.

The discussion highlighted transparency and AI literacy as equally essential. First, participants called for the introduction of AI user interfaces that reveal uncertainty, explain how outputs are produced and make overriding models easy. Second, staff training must move beyond using the tool to understanding biases, fragility and operational risks. People must be empowered to be "brave enough to distrust AI" when something feels wrong.

Finally, the group agreed that humanitarian organizations could not manage AI safety on their own. They rely on academia, the tech industry and the public sector for expertise, safe model development, evaluation, incident-sharing mechanisms and capacity-building. Humanitarian organizations that are rolling out AI, as well as the communities where it is being deployed, must be allowed to participate early in the design process to ensure tools remain safe, appropriate and aligned with humanitarian principles.

KEY TAKEAWAYS

AI safety and security risks are highly context-dependent. Participants agreed that AI has valuable use cases before, during and after crises, but that risk profiles vary drastically. Back-office and logistics tools pose manageable risks; front-line systems that influence triage or early warning demand far greater scrutiny.

Safety-by-design means making hard operational choices about data, transparency and model training. Because AI systems directly embed data into algorithms, model behaviour mirrors the data used to train it. Beyond traditional cybersecurity and data protection, AI introduces new vulnerabilities and risks: the entanglement of data and algorithms makes systems susceptible to data poisoning and bias arising from poor data quality.

Training must equip staff and communities not just to use AI tools, but also to understand inherent biases and risks, and to question and distrust AI systems. Beyond individual organizations, participants called for the building of broader societal literacy so that communities understood what AI could and would not do, and when to challenge its outputs.

ROUNDTABLE 4

DIGITAL TRUST: BUILDING AND SUSTAINING CONFIDENCE IN A CONNECTED WORLD

BACKGROUND

Digital trust refers to the expectation that digital technologies and services, and the organizations providing them, will operate securely, reliably, and in line with societal values and shared expectations such as accountability and oversight. As humanitarian operations increasingly rely on digital and digitally enabled services, trust in how these systems function, how data is handled and how organizations behave becomes essential to effective and principled humanitarian response. In crisis and conflict settings, where affected populations face heightened risks of exploitation, discrimination and harm, failures in digital trust can have serious consequences for safety, dignity and humanitarian access. This round table examined digital trust as a foundational condition for the responsible use of digital technologies in humanitarian action. It focused on how technical safeguards, governance frameworks and cross-sector cooperation could be combined to build, sustain and protect digital trust in a rapidly evolving and contested digital environment.

SUMMARY OF THE ROUNDTABLE DISCUSSION

The roundtable framed digital trust as a foundational condition for humanitarian action in an increasingly connected and contested digital environment. When prompted to define digital trust in their own terms, participants offered different but complementary definitions. Rather than being framed in terms of a fixed definition, digital trust emerged as a relational, contextual and practice-driven concept. One participant defined it as the trust required to maintain relationships with patients, authorities and armed actors in order to operate safely in conflict settings. Another proposed a taxonomy of trust across four dimensions: policy, technical, ethical and legal. Meanwhile, a third defined digital trust as the confidence to use a digital tool, platform or service with the expectation that it will function as intended and that the data will be handled responsibly. Throughout the discussion, digital trust was never framed as a purely technical concept; it meant confidence in how digital tools behaved and how data is handled, and in an organization act predictably and responsibly.

Building on this shared understanding, participants discussed why digital trust is increasingly difficult to establish and sustain. Speakers highlighted that trust becomes fragile when digital systems fail, are misused or are not sufficiently understood. They pointed to awareness and education gaps, cybersecurity threats and data breaches, misinformation, and rapid technological change as factors that eroded trust over time. Participants noted that, in conflict and crisis settings, these challenges became more acute, as humanitarian organizations often operated under conditions of insecurity, with cross-border data flows and weak regulatory protections, while handling highly sensitive information about vulnerable populations. They expressed particular concern about the misuse, repurposing or exploitation of humanitarian data during conflicts, noting that such incidents

could expose affected people to harm and quickly erode trust, leading communities to disengage from digital services and undermining humanitarian access and acceptance.

The discussion on building trust emphasized that technical measures – such as cybersecurity, encryption and privacy-by-design – mattered but were not sufficient on their own. Instead, building trust requires consistent, transparent and predictable behaviour. Participants highlighted governance and policy commitments, as well as visible cybersecurity practices such as multi-factor authentication, as important trust-building measures. One speaker argued that transparency in explaining organizational choices and risk trade-offs is itself trust-building. To address the challenge of dual-use data and humanitarian organizations' inability to match the resources of belligerents, one proposed best practice was to avoid holding dual-use data in the first place or to keep it in non-digital form.

Finally, the ICRC's "digital emblem" project was raised as a concrete trust-related initiative, framed as an emerging effort to build a recognizable system in cyberspace that could support humanitarian protection goals. While the digital red cross, red crescent and red crystal emblem is by itself, does not offer protection against cyberattacks, it was conceptualized to serve as a signal of the protected status of humanitarian and medical digital assets. Participants discussed the need for a clear and widely adopted technical standard, as well as the capabilities that the digital emblem should enable, including the traceability of digital certificates.

KEY TAKEAWAYS

Digital trust in humanitarian action results from consistent, transparent and predictable behaviour across safe and reliable technological systems, and from cross-border collaboration, especially when handling sensitive data in high-risk contexts.

In conflict settings, digital trust is particularly fragile because the misuse, repurposing or leakage of humanitarian data can directly endanger affected populations and undermine neutrality, access, and acceptance by communities.

Building and sustaining digital trust requires preparedness for, and a forecast of, future risks, including contingency planning for technological failure or misuse, and practical mechanisms such as the ICRC's "digital emblem" project, which is intended to signal the legal protection afforded to humanitarian and medical digital assets.

ACKNOWLEDGEMENTS

This report aims to capture the discussions held and insights gained during the Symposium on the Responsible Use of Technology in Humanitarian Action, which took place in Beijing, China, on 4 and 5 December 2025. The Symposium was planned and organized by the International Committee of the Red Cross (ICRC) Global Cyber Hub in Luxembourg, the ICRC's Regional Delegation for East Asia in Beijing, and the Center for International Security and Strategy (CISS) of Tsinghua University.

The organizers would like to extend special thanks to the more than 80 contributors and attendees who travelled to Beijing from almost 20 countries to participate in the intense two-day event, as well as to a large number of people in the ICRC's Global Cyber Hub and its Regional Delegation for East Asia, and at Tsinghua University, who made the event possible.

This report was prepared by the ICRC Symposium organizing team. The authors wish to thank all those who contributed to the preparation process. The report is based on the authors' interpretation of the notes collected under the Chatham House Rule. While every effort has been made to capture the essence of the conversations, certain omissions or inaccuracies are inevitable. The report does not necessarily reflect the official opinions of the event organizers, participants or facilitators. The organizers do not guarantee the accuracy or reliability of any recommendations, opinions or other information presented. Any errors or misrepresentations in the report are solely the responsibility of the authors.

The ICRC helps people around the world affected by armed conflict and other violence, doing everything it can to protect their lives and dignity and to relieve their suffering, often with its Red Cross and Red Crescent partners. The organization also seeks to prevent hardship by promoting and strengthening humanitarian law and championing universal humanitarian principles.

People know they can count on the ICRC to carry out a range of life-saving activities in conflict zones and to work closely with the communities there to understand and meet their needs. The organization's experience and expertise enables it to respond quickly and effectively, without taking sides.

@CISS Tsinghua

@CISS–China Forum

CISS Podcast

ICRC Website

www.icrc.org
facebook.com/icrc
x.com/icrc
instagram.com/icrc

International Committee of the Red Cross
19, avenue de la Paix
1202 Geneva, Switzerland
T +41 22 734 60 01
shop.icrc.org

ICRC

2529/002 2026