IAEA
International Atomic Energy Agency
Atoms for Peace and Development

Technical Meeting on Public Communication in Emergencies: Tackling Misinformation and Retaining Public Trust in Disruptive Information Environments

IAEA Headquarters, Vienna, Austria

23-27 June 2025

Ref. No.: EVT2304696

Information Sheet

Introduction

Misinformation [i] and disinformation [ii] are ever-more frequent disruptive factors in the public communication environment during disasters, emergencies, and crises[1].

During the response to the COVID-19 pandemic, human-made disinformation and misinformation distributed on digital channels significantly amplified public anxiety [iii], creating doubt [iv], fostering divisiveness, and encouraging distrust of authoritative protective guidance[v].

Generative artificial intelligence (GAI) is now widely available, creating novel challenges and opportunities for emergency responders. Artificial intelligence (AI) has the potential to both help diminish the disruption caused by dis- and misinformation[vi], as well as to worsen its harms[vii]. While GAI's potential to be misused to produce and distribute disinformation on a mass-scale remains under study, its capabilities continue to advance: GAI currently can be used to generate multilingual content easily, affordably, and swiftly in text, audio, images, or video formats[viii,ix]. AI generated text, audio, image, and video are

[1] The UN Secretary-General's report on countering disinformation notes that *dis*information is "information that is inaccurate, intended to deceive and shared in order to do serious harm". In contrast, UNESCO defines *mis*information as "misleading information created or disseminated without manipulative or malicious intent".

increasingly indistinguishable from human-made content. Experimental research demonstrates that GAI can utilize publicly available personal information to tailor persuasive arguments in direct dialogue with humans that succeed in persuading human interlocutors more effectively than their human counterparts[x]. In controlled studies that cannot fully replicate the authoritative, dynamic and contextual nature of real-world, communication by emergency response organisations, regulatory authorities and operators, among other official sources of information, GAI can produce social media disinformation that is more convincing than human-produced disinformation on the same subject[xi].

Preparedness enhancement is needed since persuasive and deceptive AI generated content is already maliciously used to manipulate the public, highlighting GAI's potential for emergency response disruption. Artificial content is now used to defraud and extort consumers and businesses[xii]; deceive investors[xiii, xiv]; spread harmful comments[xv], confuse the public about the authenticity of political leader's photographs, audio recordings and video statements[xvi]; and engender false, divisive narratives about the cause of emergencies and the equitability of the response[xvii, xviii]. More broadly, research indicates that human-made, and in future AI-enabled, disruptive disinformation can act as a hybrid threat, harming "collective decision-making" processes[xix], which in turn reduce emergency response resilience.

Radiological and nuclear emergencies are perceived by the public as "high risk" events, regardless the actual hazards[xx]. The public's misperception of radiological and nuclear emergencies risks is exacerbated by extensive publicly accessible misinformation on the internet and social media[xxi]. The public's overestimation of the radiological or nuclear risks is complicated by the health consequences the public fears, as well as the "societal risks" such as evacuation from contaminated areas[xxii]. Deployed during the response to a severe emergency, AI generated deceptive content could trigger significant socio-economic disturbance, potentially creating or exacerbating a transnational challenge.

To help mitigate these harms, multi-sectoral collaboration could be coordinated to monitor, analyze and counteract AI generated mis/disinformation given its disruptive consequences. For instance, research and other activities undertaken could improve detection of deceptive or harmful content dealing with radiological and nuclear emergencies, help develop capabilities for evidence-based attribution to models and/or perpetrators of disinformation and further deployment of standards and practices to enhance the ability to verify the provenance and legitimacy of emergency preparedness and response information.[xxiii, xxiv]. Cross-sectoral collaboration could support the exchange of knowledge on risks to emergency preparedness and response measures from AI generated mis/disinformation, while strengthening the public's resilience against manipulation via disinformation.

In an emergency, the public must be able to rely on credible, actionable, authoritative information that is provided by emergency response organisations, regulatory organisations, operators, local authorities, and the mass media. In a disruptive communication environment, it may be more difficult to find authoritative information that helps the public avoid or reduce safety risks. Mis/disinformation could encourage both inappropriate public responses and increased sharing of false narratives, which in turn increases public anxiety and confusion, undermining trust in authoritative information and public protective instructions, thereby potentially further increasing public safety hazards. There is also a risk that widely shared AI-generated content could deceive emergency responders into misallocating response resources that are more critically needed elsewhere.

Objectives

The purpose of the Technical Meeting is to share good practices and experience, gather expert advice, operational knowledge, and research results to develop more resilient and effective emergency public communication preparedness and response measures to mitigate the harms caused by human-made and AI-

produced disruptive disinformation during both routine operation and emergencies, including supporting emergency response organisations, regulatory authorities and operators in identifying dis/misinformation.

Target Audience

The target audience includes emergency response managers, planners, and public communicators dealing with nuclear and radiological emergencies, as well as emergencies and disasters triggered by other causes, among emergency response organisations, nuclear regulatory agencies, critical infrastructure operators, nuclear facility operators, social media platforms, digital network operators, generative artificial intelligence model providers, and mass media.

Working Language(s)

English.

Expected Outputs

The insights gathered during the Technical Meeting will support the development of recommendations on means to reduce the risk that deceptive disinformation erodes public trust and fosters greater public anxiety during an emergency response.

Structure

The Technical Meeting will be conducted as a five-day in-person meeting and may include side events, tabletop exercises, technology demonstrations, breakout sessions, and panel discussions.

Topics

The following topics will be discussed at the Technical Meeting. Participants are encouraged to submit proposals for presentations on these topics, please see the instructions for the submission of abstracts and presentations below. The presentations may include an overview of research, case studies, lessons learned and proposals on:

- Disruptive impact:
    - Human-made and/or AI-generated mis/disinformation generating misperceptions, polarizing and/or misleading narratives that disrupt emergency response;
    - Erosion of public trust in emergency response organisations caused by human-made and/or AI-generated mis/disinformation, including false visual information at scale.
- Public trust:

    - Building public trust in an emergency response organisation as a credible information source in emergencies, including among constituencies with comparatively low trust in institutions.
- Practical measures:
    - Counteracting the disruptive influence of mis/disinformation in emergencies;
    - Lessons learned in deploying multilingual AI-supported tools to counteract the disruptive impact of mis/disinformation during emergency response;
    - Lessons learned in proactive communication and building the public's media literacy.
- Current and future AI-capabilities:
    - Prognosis of the evolution of AI-produced disruptive messaging's capacity and methodology to influence the public;
    - Overview of current and/or projected AI-supported tools to identify both human-made and AI-produced mis/disinformation that disrupts public protective measures during emergencies;
    - AI-supported means to amplify fact-based messaging to counteract mis/disinformation;
    - AI support and tools for routine emergency preparedness communication with the public, including crisis call centres and hotlines and mass-scale fact checking.
- Enabling multisectoral and multidisciplinary collaboration:
    - Review of existing multisectoral collaborative models for monitoring, analysing, characterizing, and sharing information about disruptive and/or abusive activities in digital networks;
    - Proposals for collaborative arrangements to:
        - Exchange in real-time, authenticated, secure information about mis/disinformation that may disrupt the public's implementation of protective actions during emergency response;
        - Share technical guidance and feedback on measures to mitigate disruptive disinformation;
        - Support benchmarking and evaluation of Large Language Models' (LLM) vulnerability to be misused to generate content that is disruptive to emergency response, and enable rectification by the LLM's operator, either prior to roll-out, and/or during active operation;
        - Organize, implement, and assess exercises that test collaborative, mitigatory procedures;
        - Develop and offer incentives for detecting instances of misleading or disruptive messaging to undermine public protective actions.
- Proposals on policy, methods, and technical options to:
    - Direct the public to accurate, science-based, authoritative information on protecting their safety;
    - Assure compliance with international humanitarian law by respecting rights to freedom of expression and security of the person, including epistemic security.
    - Minimize legal liabilities in using AI tools in emergency public communication preparedness and response.

Participation and Registration

All persons wishing to participate in the event have to be designated by an IAEA Member State or should be members of organizations that have been invited to attend.

In order to be designated by an IAEA Member State or invited organization, participants are requested to submit their application via the InTouch+ platform (https://intouchplus.iaea.org) to the competent national authority (Ministry of Foreign Affairs, Permanent Mission to the IAEA or National Atomic Energy Authority) or organization for onward transmission to the IAEA by **1 March 2025**, following the registration procedure in InTouch+:

1. Access the InTouch+ platform (https://intouchplus.iaea.org):
   - Persons with an existing NUCLEUS account can sign in to the platform with their username and password;
   - Persons without an existing NUCLEUS account can register here.

2. Once signed in, prospective participants can use the InTouch+ platform to:
   - Complete or update their personal details under 'Complete Profile' and upload the relevant supporting documents;
   - Search for the relevant event under the 'My Eligible Events' tab;
   - Select the Member State or invited organization they want to represent from the drop-down menu entitled 'Designating Authority' (if an invited organization is not listed, please contact InTouchPlus.Contact-Point@iaea.org);
   - If applicable, indicate whether financial support is requested and complete the relevant information (this is not applicable to participants from invited organizations);
   - Based on the data input, the InTouch+ platform will automatically generate the Participation Form (Form A) and/or the Grant Application Form (Form C);
   - Submit their application.

Once submitted through the InTouch+ platform, the application, together with the auto-generated form(s), will be transmitted automatically to the required authority for approval. If approved, the application, together with the applicable form(s), will automatically be sent to the IAEA through the online platform.

NOTE: The application for financial support should be made, together with the submission of the application, by **1 March 2025**.

For additional information on how to apply for an event, please refer to the InTouch+ Help page. Any other issues or queries related to InTouch+ can be sent to InTouchPlus.Contact-Point@iaea.org.

Selected participants will be informed in due course on the procedures to be followed with regard to administrative and financial matters.

Participants are hereby informed that the personal data they submit will be processed in line with the Agency's Personal Data and Privacy Policy and is collected solely for the purpose(s) of reviewing and assessing the application and to complete logistical arrangements where required. The IAEA may also use the contact details of Applicants to inform them of the IAEA's scientific and technical publications, or the latest employment opportunities and current open vacancies at the IAEA. These secondary purposes are consistent with the IAEA's mandate. Further information can be found in the Data Processing Notice concerning IAEA InTouch+ platform.

Papers and Presentations

The IAEA encourages participants to give presentations on the work of their respective institutions that falls under the topics listed above.

Participants who wish to give presentations are requested to submit an abstract of their work. The abstract will be reviewed as part of the selection process for presentations. The abstract should be in A4 page format, should extend to no more than 2 pages (including figures and tables) and should not exceed 600 words. It should be sent electronically to Nayana Jayarajan, the Scientific Secretary of the event (see contact details below), not later than **1 March 2025**. Authors will be notified of the acceptance of their proposed presentations by **1 April 2025**.

In addition to the registration already submitted through the InTouch+ platform, participants have to submit the abstract, together with the Form for Submission of a Paper (Form B), to the competent national authority (e.g. Ministry of Foreign Affairs, Permanent Mission to the IAEA or National Atomic Energy Authority) or organization for onward transmission to the IAEA not later than **1 March 2025**.

Expenditures and Grants

No registration fee is charged to participants.

The IAEA is generally not in a position to bear the travel and other costs of participants in the event. The IAEA has, however, limited funds at its disposal to help meet the cost of attendance of certain participants. Upon specific request, such assistance may be offered to normally one participant per country, provided that, in the IAEA's view, the participant will make an important contribution to the event or, will make an important contribution to his or her State's emergency preparedness and response arrangements.

The application for financial support should be made, together with the submission of the application, by **1 March 2025**.

Venue

The event will be held at the Vienna International Centre (VIC), where the IAEA's Headquarters are located. Participants must make their own travel and accommodation arrangements.

General information on the VIC and other practical details, such as a list of hotels offering a reduced rate for IAEA participants, are listed on the following IAEA web page:

www.iaea.org/events.

Participants are advised to arrive at Checkpoint 1/Gate 1 of the VIC one hour before the start of the event on the first day in order to allow for timely registration. Participants will need to present an official photo identification document in order to be admitted to the VIC premises.

Visas

Participants who require a visa to enter Austria should submit the necessary application to the nearest diplomatic or consular representative of Austria at least four weeks before they travel to Austria. Since Austria is a Schengen State, persons requiring a visa will have to apply for a Schengen visa. In States where Austria has no diplomatic mission, visas can be obtained from the consular authority of a Schengen Partner State representing Austria in the country in question.

IAEA Contacts

Scientific Secretary:

Ms Nayana Jayarajan
Incident and Emergency Centre
Department of Nuclear Safety and Security
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 VIENNA
AUSTRIA

Tel.: +43 1 2600 26102
Email: N.Jayarajan@iaea.org

Administrative Secretary:

Ms Reka Kalrenbach
Incident and Emergency Centre
Department of Nuclear Safety and Security
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 VIENNA
AUSTRIA

Tel.: +43 1 2600 21482
Email: r.kaltenbach@iaea.org

Subsequent correspondence on scientific matters should be sent to the Scientific Secretary and correspondence on other matters related to the event to the Administrative Secretary.

Event Web Page

Please visit the following IAEA web page regularly for new information regarding this event:

www.iaea.org/events/2304696

Enclosure: Form for Submission of a Paper (Form B) (if 'Abstract/Paper submission requested)

---

[i] "Countering Disinformation for the Promotion and Protection of Human Rights and Fundamental Freedoms Report of the Secretary-General." A/77/287, United Nations, 29 Aug. 2022, https://digitallibrary.un.org/record/3987886?v=pdf

[ii] Cherilyn Ireton and Julie Posetti. Journalism, 'Fake News' & Disinformation, Handbook for Journalism Education and Training, UNESCO, Paris, France, 2018, https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0_0.pdf . Accessed 25 Mar. 2024.

[iii] Verma, G., Bhardwaj, A., Aledavood, T. et al. Examining the impact of sharing COVID-19 misinformation online on mental health. Sci Rep 12, 8045 (2022). https://doi.org/10.1038/s41598-022-11488-y

[iv] Jacob, Cécile, et al. "The Effect of Communication and Disinformation during the COVID-19 Pandemic: Think Tank: European Parliament." Think Tank | European Parliament, Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Jan. 2023, www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2023)740063 .

[v] Ibid.

[vi] Feuerriegel, S., DiResta, R., Goldstein, J.A. *et al.* Research can help to tackle AI-generated disinformation. *Nat Hum Behav* **7**, 1818–1821 (2023). https://doi.org/10.1038/s41562-023-01726-2

[vii] Bontridder N, Poullet Y. The role of artificial intelligence in disinformation. Data & Policy. 2021;3:e32. doi:10.1017/dap.2021.20

[viii] Bontcheva, Kalina. "New White Paper on Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities." EDMO, European Digital Media Observatory (EDMO), AI4TRUST, vera.ai, TITAN, AI4Media, 13 Feb. 2024, edmo.eu/edmo-news/new-white-paper-on-generative-ai-and-disinformation-recent-advances-challenges-and-opportunities/

[ix] L. Whittaker, T. C. Kietzmann, J. Kietzmann and A. Dabirian, ""All Around Me Are Synthetic Faces": The Mad World of AI-Generated Media," in IT Professional, vol. 22, no. 5, pp. 90-99, 1 Sept.-Oct. 2020, doi: 10.1109/MITP.2020.2985492.

[x] Salvi, Francesco, et al. "On the Conversational Persuasiveness of Large Language Models: A Randomized Controlled Trial." arXiv.Org, 21 Mar. 2024, arxiv.org/abs/2403.14380.

[xi] Giovanni Spitale et al., AI model GPT-3 (dis)informs us better than humans. *Sci. Adv*. 9, eadh1850 (2023). DOI: 10.1126/sciadv.adh1850

[xii] Ferrara, E., GenAI against humanity: nefarious applications of generative artificial intelligence and large language models. J Comput Soc Sc (2024). https://doi.org/10.1007/s42001-024-00250-1

[xiii] Sytsma, Tobias, James V. Marrone, Anton Shenk, Gabriel Leonard, Lydia Grek, and Joshua Steier, Technological and Economic Threats to the U.S. Financial System: An Initial Assessment of Growing Risks. ; Santa Monica, CA: RAND Corporation, 2024. https://www.rand.org/pubs/research_reports/RRA2533-1.html .

[xiv] Atherton, Daniel. (2023-05-22) Incident Number 543. in McGregor, S. (ed.) *Artificial Intelligence Incident Database.* Responsible AI Collaborative. Retrieved on March 28, 2024 from incidentdatabase.ai/cite/543.

[xv] Yang, Kai-Cheng, and Menczer, Filippo. "Anatomy of an AI-Powered Malicious Social Botnet." arXiv.Org, 30 July 2023, arxiv.org/abs/2307.16336.

[xvi] Sison, Alejo Jose G. and Daza, Marco Tulio and Gozalo-Brizuela, Roberto and Garrido Merchán, Eduardo César, ChatGPT: More than a "Weapon of Mass Deception" Ethical Challenges and Responses from the Human-Centered Artificial Intelligence (HCAI) Perspective (April 6, 2023). http://dx.doi.org/10.2139/ssrn.4423874

[xvii] A Fragile State of Preparedness: 2023 Report on the State of the World's Preparedness. Global Preparedness Monitoring Board, Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO

[xviii] Gupta, Ritwik, et al. "Trustworthy Disaster Response: Technology, Policy, and Society." Center for Security in Politics (CSP) , University of California Berkeley, 17 Oct. 2023, csp.berkeley.edu/2023/10/17/trustworthy-disaster-response-technology-policy-and-society/.

[xix] Arcos R, Gertrudix M, Arribas C and Cardarilli M. Responses to digital disinformation as part of hybrid threats: a systematic review on the effects of disinformation and the effectiveness of fact-checking/debunking [version 1; peer review: 2 approved]. Open Res Europe 2022, 2:8 (https://doi.org/10.12688/openreseurope.14088.1)

[xx] Slovic, P. (2012). The perception gap: Radiation and risk. Bulletin of the Atomic Scientists, 68(3), 67-75. https://doi.org/10.1177/0096340212444870

[xxi] James Mc Laughlin, Some observations on perceptions of radiation risks in the context of nuclear power plant accidents, Radiation Protection Dosimetry, Volume 199, Issue 18, November 2023, Pages 2169–2173, https://doi.org/10.1093/rpd/ncad184

[xxii] Mclaughlin, J., et.al.
[xxiii] Feuerriegel, S., et.al.
[xxiv] Micallef, N., Sandoval-Castañeda, M., Cohen, A., Ahamad, M., Kumar, S., & Memon, N. (2022). Cross-Platform Multimodal Misinformation: Taxonomy, Characteristics and Detection for Textual Posts and Videos. Proceedings of the International AAAI Conference on Web and Social Media, 16(1), 651-662. https://doi.org/10.1609/icwsm.v16i1.19323