

**Forum:** Special Conference on Conflict Realities

**Issue:** Addressing Artificial Intelligence (AI) powered disinformation in conflict settings

**Student Officer:** Virginia Genitsaropoulou

**Position:** President

---

## INTRODUCTION

It is true that over the past twenty years, disinformation has been spread widely on social media. With many believing that it has been intensifying political unrest around the globe, the phenomenon has taken on a completely new form with the availability of generative AI, and more specifically with the public release of OpenAI's ChatGPT. Even though AI offers benefits, there are drawbacks as well, particularly in regard to data integrity.

One of the most worrying issues is AI-powered disinformation, because of its speed, convincing nature, aimed at the individual beliefs, and hence hard to trace and counter. It thereby erodes trust in media and institutions, manipulates public opinion, which creates and disseminates false information using algorithms and machine learning techniques. This phenomenon is especially noticeable during times of war, as false information has the power to incite violence, heighten tensions, and destabilize entire regions. AI-driven disinformation takes advantage of the speed at which modern communication channels can spread, making it difficult for individuals and institutions to distinguish between true and false information. In conflict situations where information can be used as a weapon, the effects of such disinformation are worse. Public sentiment that is stoked by false information and edited content can escalate conflicts, sabotage peace efforts, and erode public confidence in the government and media.

## DEFINITION OF KEY-TERMS

### *Artificial Intelligence*

The use or study of computer systems or machines that have some of the qualities that the human brain has, such as the ability to interpret and produce language in a way that seems human.<sup>1</sup>

---

<sup>1</sup> "Liar's War: Protecting Civilians from Disinformation During Armed Conflict." *International Review of the Red Cross*, <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914>.

## **Disinformation**

Is deliberately created to mislead or manipulate a person, a social group, organization, or a country.<sup>2</sup>

## **Misinformation**

Is false, but not created or shared with the intention of causing harm.<sup>3</sup> If one is not aware, why someone is spreading bad information, it's best to use misinformation.<sup>4</sup>

## **Malinformation**

Is based on fact, but used out of context to mislead, harm, or manipulate. An example of malinformation is editing a video to remove important context to harm or mislead.<sup>5</sup>

## **Machine learning**

The process of computers improving their own ability to carry out tasks by analyzing new data, without a human needing to give instructions in the form of a program, or the study of creating and using computer systems that can do this.<sup>6</sup>

## **Supervised learning**

Supervised learning is a subcategory of machine learning and artificial intelligence. It is defined by its use of labeled data sets to train algorithms that to classify data or predict outcomes accurately.<sup>7</sup>

## **Robot**

A computer program that works automatically, especially one that searches for and finds information on the internet.<sup>8</sup>

---

<sup>2</sup>"Foreign Influence Operations and Disinformation." *CISA*, [www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation](https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation).

<sup>3</sup>"Foreign Influence Operations and Disinformation." *CISA*, [www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation](https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation).

<sup>4</sup>"Misinformation" Vs. "Disinformation": Get Informed On The Difference." *Dictionary.com*, 29 Aug. 2022, [www.dictionary.com/e/misinformation-vs-disinformation-get-informed-on-the-difference/](https://www.dictionary.com/e/misinformation-vs-disinformation-get-informed-on-the-difference/).

<sup>5</sup>"Foreign Influence Operations and Disinformation." *CISA*, [www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation](https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation).

<sup>6</sup>"Machine Learning." *Cambridge Dictionary | English Dictionary, Translations & Thesaurus*, <https://dictionary.cambridge.org/dictionary/english/machine-learning>.

<sup>7</sup>"What is Supervised Learning?" *IBM - United States*, [www.ibm.com/topics/supervised-learning](https://www.ibm.com/topics/supervised-learning).

<sup>8</sup>"Bot." *Cambridge Dictionary | English Dictionary, Translations & Thesaurus*, <https://dictionary.cambridge.org/dictionary/english/bot?q=bots>.

## **Detection algorithm**

A detection algorithm is a technique used by an operating system to identify deadlocks in the system. This algorithm checks the status of processes and resources to determine whether any deadlock has occurred and takes appropriate actions to recover from the deadlock.<sup>9</sup>

## **Graphic violence**

Graphic violence means the visual depiction or representation of realistic serious injury to a human or human-like being where such serious injury includes amputation, decapitation, dismemberment, bloodshed, mutilation, maiming or disfiguration.<sup>10</sup>

## **User profiling**

User profiling is the process of grouping customers or website and application users into specific groups based on various metrics.<sup>11</sup>

## **Microtargeting**

A marketing strategy that uses consumer data and demographics to identify the interests and preferences of specific individuals or small groups to send targeted advertisements that align with their interests.<sup>12</sup>

## **Multinational corporations**

A company that operates in its home country, as well as in other countries around the world. It maintains a central office located in one country, which coordinates the management of all of its other offices, such as administrative branches or factories.<sup>13</sup>

## **Proxy war**

A proxy war, or proxy warfare, is a war in which opposite sides use third parties as substitutes for fighting each other directly. States sometimes use less powerful states as proxies but more often use violent non-state actors, mercenaries, and other third parties, such as during a civil war.<sup>14</sup>

---

<sup>9</sup>"Deadlock Detection Algorithm in Operating System." *GeeksforGeeks*, 12 Feb. 2024, [www.geeksforgeeks.org/deadlock-detection-algorithm-in-operating-system/](https://www.geeksforgeeks.org/deadlock-detection-algorithm-in-operating-system/).

<sup>10</sup> "Graphic Violence Definition." Law Insider, [www.lawinsider.com/dictionary/graphic-violence](https://www.lawinsider.com/dictionary/graphic-violence).

<sup>11</sup>"Just a Moment..." Just a Moment., [www.indeed.com/career-advice/career-development/what-is-user-profiling](https://www.indeed.com/career-advice/career-development/what-is-user-profiling).

<sup>12</sup>"What is Microtargeting? | Definition from TechTarget." CIO, 19 Sept. 2023, [www.techtarget.com/searchcio/definition/microtargeting](https://www.techtarget.com/searchcio/definition/microtargeting).

<sup>13</sup> CFI Team. "Multinational Corporation (MNC)." Corporate Finance Institute, 15 Oct. 2023, <https://corporatefinanceinstitute.com/resources/management/multinational-corporation/>.

<sup>14</sup> "Proxy War - Simple English Wikipedia, the Free Encyclopedia." *Wikipedia*, Wikimedia Foundation, Inc, 4 Jan. 2014, [www.simple.wikipedia.org/wiki/Proxy\\_war](https://www.simple.wikipedia.org/wiki/Proxy_war). Accessed 27 July 2024.

## ***Data integrity***

Data integrity is a concept and process that ensures the accuracy, completeness, consistency, and validity of an organization's data.<sup>15</sup>

## ***Affordable AI mechanisms***

AI systems and technologies that are cost-effective and accessible for a wide range of users, including businesses, governments, and individuals.<sup>16</sup>

# **BACKGROUND INFORMATION**

## ***AI and disinformation***

### AI as a tool of detecting disinformation

AI technologies have shown to be very successful in the information operations domain when it comes to identifying and eliminating unwanted and unlawful content from the internet. Advanced filtering capabilities let AI algorithms analyze text, images, and videos that may contain material harmful to users, such as hate speech and misinformation. Real-time monitoring provides for instant acknowledgment and reaction, which is an important element of the solution because harmful information spreads very fast across platforms. Its scalability makes it practical and effective to deal with a heavy volume of online content, making this quite practical compared to human moderation. This also helps AI recognize the patterns indicative of coordinated disinformation efforts and therefore be better placed in dismantling such networks. With continuous learning, AI updates these algorithms and trains on new data to adapt to new threats. Well-implemented AI also reduces human bias, so it is more consistent in its approach to content moderation. The aforementioned makes AI particularly good at ensuring the safety and legality of online platforms.<sup>17</sup> The detection and identification of fake robot accounts—a process referred to as "bot-spotting" and "bot-labeling"—has also been successfully accomplished by AI approaches. Social media companies are helping users make greater sense of the content they are interacting with and determine its authenticity by designating accounts that have been recognized as bots. Machine-learning algorithms are used by Google, Facebook and other Internet service providers to detect and eliminate fake bot accounts, eliminate trolls, and proactively flag information that may be offensive.

### User profiling and micro-targeting

Adversaries will be more adept at identifying people's distinct traits, needs, vulnerabilities, and views as machine learning progresses. They will then be able to target the people who are most susceptible to influence with greatest effectiveness by delivering highly personalized material. The creation of content automatically is a related development. Content can be automatically generated for individual users by using specific targeting combined with Natural Language Generation techniques, based on psychological, personal, or other attributes. Even

---

<sup>15</sup> "What Is Data Integrity?" *Fortinet*, [www.fortinet.com/resources/cyberglossary/data-integrity](http://www.fortinet.com/resources/cyberglossary/data-integrity).

<sup>16</sup> "Affordable AI?" *LinkedIn*, 19 May 2024, [www.linkedin.com/pulse/affordable-ai-subrata-roy-quesec](http://www.linkedin.com/pulse/affordable-ai-subrata-roy-quesec).

<sup>17</sup> *ArXiv.org E-Print Archive*, <https://arxiv.org/pdf/1802.07228>.

though consumers may think that the material they come across is unbiased, created by citizens, and shared by all users, user profiling and micro-targeting, which rely on the gathering and manipulation of user data to predict and shape voters' political opinions and election outcomes, may be dangerous for democracy, free speech, and voter choice.

### ***AI powered disinformation in conflict settings***

Inflaming pre-existing tensions by inventing fictional violence, wrongly attributing genuine violence, or accusing actors of violent intent is one of the most popular methods of creating conflict in areas affected by conflict. Since nearly 25%<sup>18</sup> of people in the area use social media, misinformation can spread swiftly due in part to the internet's ability to convey incorrect information to analog media like radio. This makes it possible for people without access to the Internet to become misinformed. While deception on the internet has always been a tactic in conflicts, tensions already present may make it worse. Recaptioning photos taken in many locations and countries and falsely attributing them to a fictitious conflict has been a special trend. For example, social media users have been re-captioning violent images and videos from other countries, such as a church massacre in Nigeria, and using this as false proof that Rwandans were killing Congolese, and vice versa.<sup>19</sup> This has led to increased tensions between the Democratic Republic of the Congo (DRC) and Rwanda.

Governments play a key role in this matter as well. There have been numerous instances worldwide where it seems that governments disseminated misleading information for political reasons. Many have reported that national governments are undermining traditional media, sometimes weakening the credibility of journalists in favor of social media influencers. This has led people to believe news from less reliable websites or even AI-powered bots more readily. Protests against multinational corporations exacerbate the issue. 75 % of UN Peacekeepers surveyed said that misinformation or disinformation has affected their safety and security. This information is noted in the UN Secretary-General's June 2023 policy brief, "Information Integrity on Digital Platforms."<sup>20</sup>

The subject of disinformation before and after AI is depicted in the following diagram. The proliferation of misinformation of this kind on internet platforms makes it harder and harder to distinguish alternative viewpoints. When misinformation about a particular group is spread, it can create a unity, leading participants to start believing hate speech.

---

<sup>18</sup> Congo Check, <https://congocheck.net/>.

<sup>19</sup> Congo Check, <https://congocheck.net/>.

<sup>20</sup> Welcome to the United Nations, [www.un.org/sexualviolenceinconflict/wp-content/uploads/2023/06/our-common-agenda-policy-brief-information-integrity-en.pdf](http://www.un.org/sexualviolenceinconflict/wp-content/uploads/2023/06/our-common-agenda-policy-brief-information-integrity-en.pdf).

Figure 1: Disinformation Before and After Generative AI<sup>18</sup>

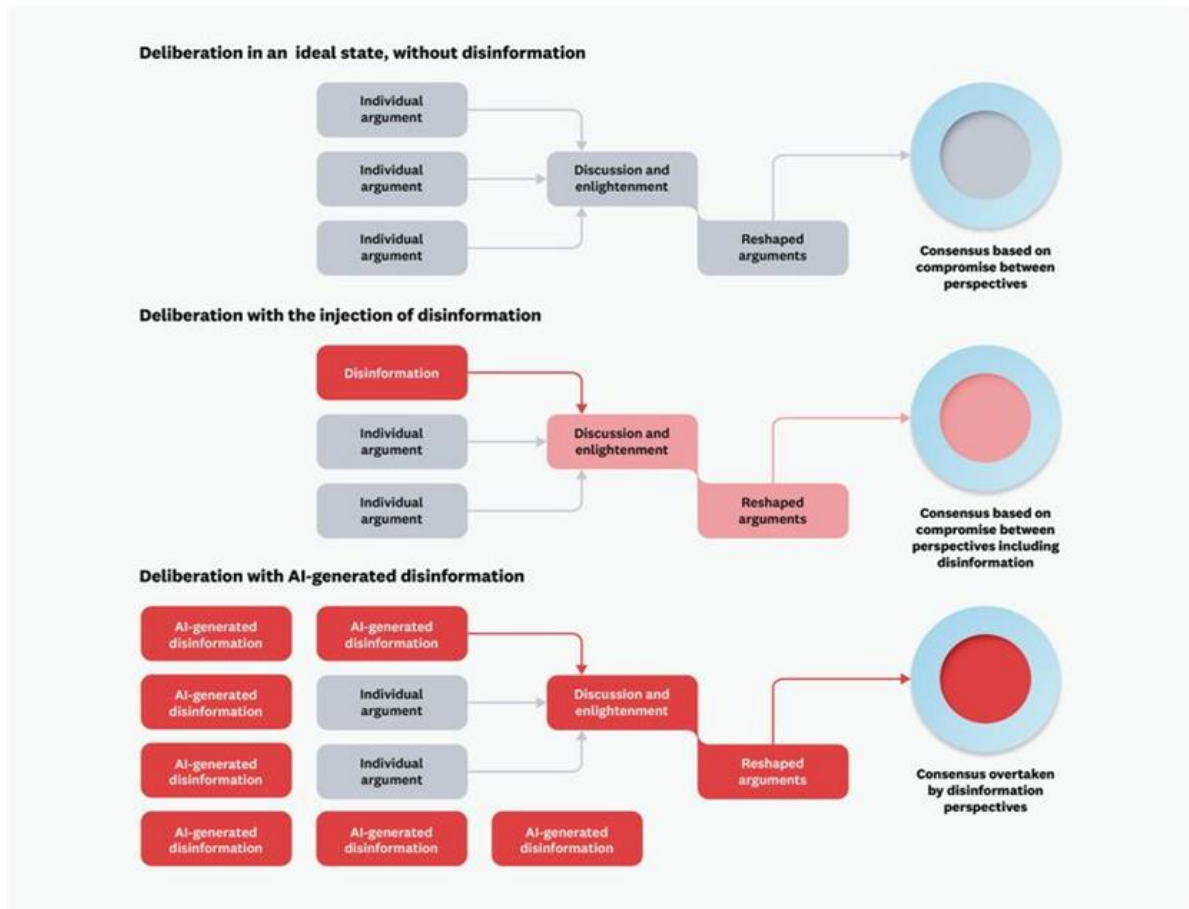


Figure 1: Disinformation before and after Generative AI<sup>21</sup>

The effect affordable AI mechanisms have on the spread of disinformation in conflict settings

Hate speech, disinformation, and misinformation are dividing societies, undermining trust, and ultimately endangering the advancement of humanity, as they create social divisions and erode public trust in institutions and information sources, thereby destabilizing societies and hindering collective progress and cooperation essential for addressing global challenges. It's standard procedure to use AI to assist in the dissemination of this content. For a considerable amount of time, disinformation operators have been using AI-powered bots on social media and training AI-powered algorithms to spread hateful and false information. However, up until now, the scope of disinformation operations has always been constrained by high maintenance and investment costs. The financial and labor barriers associated with producing and disseminating false information have decreased due to affordable, off-the-shelf generative AI tools. These days, hateful and deceptive content can be produced quickly, cheaply, and in large quantities, and it's

<sup>21</sup>"United Nations University Centre for Policy Research (UNU-CPR) on LinkedIn: #ai." LinkedIn, 11 Sept. 2023, [www.linkedin.com/posts/unucpr\\_ai-activity-7106999586749050881-6iZ0](https://www.linkedin.com/posts/unucpr_ai-activity-7106999586749050881-6iZ0).

even more difficult to identify. Artificial intelligence generated content leaves minimal trace, making it more difficult for fact-checkers, law enforcement, journalists, and regular users to distinguish it from authentic content. Effects can be seen in a variety of domains, including human rights and peace and security. It costs less to create and disseminate targeted misinformation, which is already a powerful weapon in any conflict.<sup>22</sup>

## Consequences

Disinformation can cause extreme fear, grief, or other painful emotions or unsound mental states that can directly affect the mental health of civilians. Disinformation aimed at civilian audiences may cause them to become suspicious of others or become paranoid; it may also cause them to doubt their own capacity to meet basic human needs; it may cause them to believe that friends or family have been or will be harmed; or it may cause them to reasonably fear death or physical harm. These injuries may be just as traumatic, harmful, and long-lasting even though they are harder to record and prove.<sup>23</sup>

## **MAJOR COUNTRIES AND ORGANIZATIONS INVOLVED**

### ***United States of America (USA)***

The United States has taken the initiative to enact laws and policies meant to combat misinformation. An example is the Cybersecurity and Infrastructure Security Agency (CISA) working on disinformation campaigns. CISA strengthens defenses against foreign influence operations and misinformation, lowering the risk to vital infrastructure in the United States. . Collaboration on this project takes place with international stakeholders, academia, the private sector, and the interagency. American giants like Microsoft, Facebook, Twitter, and Google are leading the way in creating artificial intelligence tools and technologies that are intended to identify and counteract misinformation. In addition, the National Security Agency (NSA), the Department of Defense (DoD), and the Central Intelligence Agency (CIA) make investments in AI technologies to identify and prevent adversaries' disinformation campaigns in conflict areas.

### ***European Union (EU)***

In reaction to misinformation and manipulation of foreign information that targets EU policies, the European Commission is stepping up its strategic communication. Raising awareness is essential to the EU's strategy for exposing and disproving false information. EUvsDisinfo, a flagship project to identify, analyze, and raise awareness of disinformation in its various forms, was started in 2015<sup>24</sup> by the European External Action Service (EEAS), the EU's diplomatic branch. It focuses on common narratives and strategies used by the Kremlin. Building

---

<sup>22</sup> "Investigation Finds AI Image Generation Models Trained on Child Abuse." *Cyber Policy Center | FSI*, <https://cyber.fsi.stanford.edu/news/investigation-finds-ai-image-generation-models-trained-child-abuse>.

<sup>23</sup> "Liar's War: Protecting Civilians from Disinformation During Armed Conflict." *International Review of the Red Cross*, <https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914>.

<sup>24</sup> "Strategic Communication and Tackling Disinformation." *European Commission*, [https://commission.europa.eu/topics/strategic-communication-and-tackling-disinformation\\_en](https://commission.europa.eu/topics/strategic-communication-and-tackling-disinformation_en).

resilience also includes raising awareness of cybersecurity. EU nations work together to get ready for any cyberattacks through the EU Agency for Cybersecurity and the Computer Emergency Response Team.

### **Russia**

Disinformation has been cited as common in Russia, as the country uses artificial intelligence to fabricate and disseminate stories in conflict areas in order to further its geopolitical objectives. According to a recent report from the Atlantic Council's Digital Forensic Research Lab, Moscow may have virtually endless access to digital ammunition in the form of chatbots like ChatGPT, AI-generated narration for videos, AI translation software, and the emergence of generative AI in general.<sup>25</sup>

### **India**

Through a variety of strategies including international cooperation, research, technology development, and regulation, India plays a major role in combating AI-powered disinformation in conflict situations. India's Ministry of Electronics and Information Technology (MeitY) is leading the country's efforts to control and regularly check on online disinformation. Social media companies are required by the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, to take down false content and fake news. The government has put out plans to stop the spread of false information. These include social media platform guidelines and a draft Data Protection Bill.

### **Global Disinformation Index (GDI)**

In 2018, GDI was founded as a non-profit organization based on the principles of neutrality, independence, and transparency. By June 2023, GDI had evaluated over 700 million<sup>26</sup> web pages in more than 40 languages, as well as websites viewed by people from more than 150 different nations. In order to best serve and inform advertisers, the ad tech industry, search and social media companies, and researchers, GDI combines artificial intelligence with extensive analyses of journalistic practice to create a neutral, independent, and transparent index of a website's risk of misinforming readers.

### **AI for Good organization**

The premier action-oriented, international, and inclusive UN platform on AI is called AI for Good. Its objective is to find useful AI applications to further the Sustainable Development Goals of the UN and scale those solutions for a global impact. The initiative is launched by the International Telecommunication Union (ITU) in partnership with other United Nations (UN) agencies and a broad coalition of stakeholders. Information and Communication Technologies (ICT) is the domain of the United Nations Information Technology (ITU) agency. Ultimately, it is a

---

<sup>25</sup> Sydney J. Freedberg Jr. "Brute Force: Russia 'doubled Down' on Often-crude Disinformation in 2023, Says Report." *Breaking Defense*, 4 Mar. 2024, [www.breakingdefense.com/2024/02/brute-force-russia-doubled-down-on-often-crude-disinformation-in-2023-says-report/](http://www.breakingdefense.com/2024/02/brute-force-russia-doubled-down-on-often-crude-disinformation-in-2023-says-report/).

<sup>26</sup> "The Global Disinformation Index." *The Global Disinformation Index*, [www.disinformationindex.org/product](http://www.disinformationindex.org/product).

network of stakeholders dedicated to applying AI for social good has been established by AI for Good.

## TIMELINE OF EVENTS

DATE	DESCRIPTION OF EVENT
23 June 2010	Google launched its Caffeine web indexing system.
6 January-8 November 2016	The use of AI-driven disinformation campaigns in the U.S. Presidential Elections attributed to the Russian Internet Research Agency (IRA).
23 November 2016	The European Union launches the “EU vs Disinformation campaign” to track and counter disinformation from foreign actors.
23 April and 7 May 2017	During the French presidential election, coordinated efforts were made to combat AI-driven misinformation.
10 April 2018	Social media executives testify before Congress on the role of AI in spreading disinformation.
28 May 2019	The AI for Good Global Summit was held at the International Telecommunication Union (ITU) Headquarters in Geneva, Switzerland.
30 March 2022	NATO enhances its StratCom Centre of Excellence efforts to counter disinformation in conflict zones, particularly focusing on AI-driven threats.
16 November 2022	The Partnership on AI (PAI) released its guidelines for the ethical use of AI to combat disinformation, which aim to provide a framework for the responsible creation and dissemination of synthetic media, with a particular focus on addressing the potential harms of AI-driven disinformation.

7 July 2023	AI for Good Summit featured panels specifically addressing AI-driven disinformation in conflict settings.
-------------	---

## RELEVANT UN RESOLUTIONS, TREATIES AND EVENTS

### ***Resolution on Information and Communications Technologies for Sustainable Development (A/RES/75/202)***<sup>27</sup>

The United Nations General Assembly Resolution A/RES/75/202 on Information and Communications Technologies for Sustainable Development acknowledges that it has been identified globally that ICTs could play a very important role in attaining the SDGs. Much emphasis by the resolution is put on digital inclusion, capacity building, infrastructure building, and development of comprehensive policy and regulatory frameworks.

### ***World Economic Forum (WEF)***

The WEF has been actively combating misinformation fueled by AI. The WEF convenes experts, stakeholders, and world leaders at its yearly meetings, like the ones held in Davos, to talk about the implications of artificial intelligence (AI), including how it both creates and counters misinformation. The WEF seeks to develop international collaboration and set moral standards for AI application. It highlights AI's dual nature, which allows it to be both a potent tool for mitigating and detecting misinformation as well as a vehicle for its dissemination. The AI Governance Alliance and the Center for the Fourth Industrial Revolution (C4IR) are two of the forum's initiatives that support responsible AI governance.

### ***Brookings Institution Event***

An event entitled "The Dangers Posed by AI and Disinformation During Elections" was held on March 13, 2024, at the Brookings Institution. The topic of discussion was the growing role that AI is playing in disseminating false information, particularly during election seasons. The event was held at the Falk Auditorium in Washington, D.C. from 2:00 PM to 3:00 PM EDT.<sup>28</sup> The Center for Technology Innovation at Brookings arranged it, and Senior Fellow Darrell West moderated a panel of experts. The main concerns raised by the recent proliferation of generative AI tools that can produce artificial text, images, videos, and audio were the main topic of discussion at the event. Election integrity is seriously threatened by these technologies' ability to quickly and cheaply create and spread misinformation.

<sup>27</sup> Official Document System - UN,

<https://documents.un.org/doc/undoc/gen/n20/379/41/pdf/n2037941.pdf?token=ZQ2nLpvWcSd98gqI2&f e=true>.

<sup>28</sup> "The Dangers Posed by AI and Disinformation During Elections." Brookings, 13 Mar. 2024, [www.brookings.edu/events/the-dangers-posed-by-ai-and-disinformation-during-elections/](http://www.brookings.edu/events/the-dangers-posed-by-ai-and-disinformation-during-elections/).

## PREVIOUS ATTEMPTS TO SOLVE THE ISSUE

### *UNESCO Media and Information Literacy*

The purpose of UNESCO's Media and Information Literacy (MIL) initiative is to improve people's capacity to access, comprehend, assess, and produce media in a variety of formats, thereby empowering both individuals and communities. In conflict-affected areas, where the dissemination of false information can heighten tensions and endanger peace efforts, this initiative is critical to fostering resilience against misinformation and disinformation.

### *European Digital Media Observatory (EDMO)*

Together with media outlets, academics, and decision-makers, EDMO creates and carries out plans to combat disinformation. Across Europe, independent fact-checking organizations are supported financially and with resources by EDMO. This support enables these organizations to disseminate accurate information to the public, verify information, and refute false claims. EDMO monitors internet platforms and identifies disinformation produced by AI by using AI and machine learning technologies.

## POSSIBLE SOLUTIONS

### *AI as means of detecting disinformation*

AI, through advanced machine learning algorithms, is crucial for detecting AI-generated disinformation by analyzing vast amounts of data to identify patterns and inconsistencies that human analysts might miss. This is especially important in conflict situations where timely, accurate information is critical. Automated systems can process large volumes of online content quickly, ensuring real-time detection and response to disinformation, a task beyond the capacity of human fact-checkers alone. Continuous training and updating of AI models help adapt to new disinformation tactics, such as deep fakes. Effective defense against digital disinformation requires significant R&D investment and collaboration between academic institutions, tech companies, and government agencies, with an emphasis on developing open-source AI models for continuous improvement and transparency.

## **Digital Watermarking**

Digital watermarking, which embeds metadata into digital content, helps mark the origin and authenticity of information, ensuring that images, videos, or documents have not been altered. This virtual fingerprint allows platforms and users to trace and verify content, which is crucial in conflict situations to prevent disinformation. By providing transparency and accountability, digital watermarking makes it easier to track malicious actors and verify authenticity, fostering a trustworthy digital environment. This approach is vital for maintaining information integrity, reducing tensions, and supporting peacebuilding and conflict resolution efforts.

## **Media Literacy Programs**

One of the most effective solutions to prevent AI-infused disinformation in times of conflict is implementing media literacy programs. These programs teach critical thinking and disinformation identification from an early age, making future generations more resilient. Public awareness campaigns are also crucial, using various media to educate society on the dangers of disinformation and the importance of verifying information before sharing. This comprehensive approach helps citizens critically analyze information, reducing the spread of false narratives and maintaining social stability in conflict situations.

## **International Standards and Agreements**

Developing international standards and agreements for regulating AI in information dissemination and fostering cross-border cooperation is crucial to counter disinformation. These standards should include ethical guidelines for AI developers and users, focusing on transparency, accountability, and human rights protection. For instance, global frameworks could mandate labeling AI-generated content and disclosing AI use in content moderation. Additionally, international agreements should establish protocols for regular audits and compliance checks to ensure AI systems are not used for spreading false information. Such measures would reduce AI-driven disinformation risks by promoting ethical and transparent AI practices.

## **BIBLIOGRAPHY**

"Artificial Intelligence-Powered Disinformation and Conflict." *United Nations University*, 8 Sept. 2023, [unu.edu/cpr/brief/artificial-intelligence-powered-disinformation-and-conflict](https://unu.edu/cpr/brief/artificial-intelligence-powered-disinformation-and-conflict).

"Formative Battles: Cold War Disinformation Campaigns and Mitigation Strategies." *Wilson Center*, [www.wilsoncenter.org/publication/formative-battles-cold-war-disinformation-campaigns-and-mitigation-strategies](https://www.wilsoncenter.org/publication/formative-battles-cold-war-disinformation-campaigns-and-mitigation-strategies).

UNU

*Collections*, [collections.unu.edu/eserv/UNU:9217/artificial\\_Intelligence\\_powered\\_disinformation.pdf](https://collections.unu.edu/eserv/UNU:9217/artificial_Intelligence_powered_disinformation.pdf)

"Artificial Intelligence-Powered Disinformation and Conflict." *United Nations University*, 8 Sept. 2023, [unu.edu/cpr/brief/artificial-intelligence-powered-disinformation-and-conflict](https://unu.edu/cpr/brief/artificial-intelligence-powered-disinformation-and-conflict).

"Global AI Governance: The Risks of Artificial Intelligence to International Peace and Security." United Nations University, [unu.edu/conversation-series/global-ai-governance-risks-artificial-intelligence-international-peace-and](https://unu.edu/conversation-series/global-ai-governance-risks-artificial-intelligence-international-peace-and).

"How AI is Boosting Disinformation." *LinkedIn*, 8 Feb. 2024, [www.linkedin.com/pulse/how-ai-boosting-disinformation-melissa-fleming-vbdpe](https://www.linkedin.com/pulse/how-ai-boosting-disinformation-melissa-fleming-vbdpe).

"Information in Conflict Settings in the Age of AI – a Humanitarian Perspective." *LinkedIn*, 27 Nov. 2023, [www.linkedin.com/pulse/information-conflict-settings-age-ai-humanitarian-robert-mardini-iup2e](https://www.linkedin.com/pulse/information-conflict-settings-age-ai-humanitarian-robert-mardini-iup2e).

"United Nations University Centre for Policy Research (UNU-CPR) on LinkedIn: #ai." *LinkedIn*, 11 Sept. 2023, [www.linkedin.com/posts/unucpr\\_ai-activity-7106999586749050881-6iZ0](https://www.linkedin.com/posts/unucpr_ai-activity-7106999586749050881-6iZ0).

"Foreign Influence Operations and Disinformation." *CISA*, [www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation](https://www.cisa.gov/topics/election-security/foreign-influence-operations-and-disinformation).

"Liar's War: Protecting Civilians from Disinformation During Armed Conflict." *International Review of the Red Cross*, [international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914](https://international-review.icrc.org/articles/protecting-civilians-from-disinformation-during-armed-conflict-914).

"The Global Disinformation Index." *The Global Disinformation Index*, [www.disinformationindex.org/product](https://www.disinformationindex.org/product)

AI for Good, 8 Mar. 2024, [aiforgood.itu.int/](https://aiforgood.itu.int/).

SITNFlash. "The History of Artificial Intelligence." *Science in the News*, 23 Apr. 2020, [sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/](https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/).

"The History of Artificial Intelligence, Machine Learning and Deep Learning." *Algotive*, [www.algotive.ai/blog/the-history-of-artificial-intelligence-machine-learning-and-deep-learning](https://www.algotive.ai/blog/the-history-of-artificial-intelligence-machine-learning-and-deep-learning).

"United States Presidential Election of 2016." *Encyclopedia Britannica*, 2 Feb. 2017, [www.britannica.com/topic/United-States-presidential-election-of-2016](https://www.britannica.com/topic/United-States-presidential-election-of-2016).

Partnership on AI - Synthetic Media, 1 May 2024, [syntheticmedia.partnershiponai.org/](https://syntheticmedia.partnershiponai.org/).

"The Dangers Posed by AI and Disinformation During Elections." *Brookings*, 13 Mar. 2024, [www.brookings.edu/events/the-dangers-posed-by-ai-and-disinformation-during-elections/](https://www.brookings.edu/events/the-dangers-posed-by-ai-and-disinformation-during-elections/).

"Register to Attend 'The Dangers Posed by AI and Disinformation During Elections'." *Open Main Menu*, [connect.brookings.edu/register-to-attend-ai-disinformation-elections](https://connect.brookings.edu/register-to-attend-ai-disinformation-elections).

"The Dangers Posed by AI and Disinformation During Elections – Michael V. Hayden Center for Intelligence, Policy, and International Security." *Michael V. Hayden Center for Intelligence, Policy, and International Security – Michael V. Hayden Center for Intelligence, Policy, and International Security*, [haydencenter.gmu.edu/event/the-dangers-posed-by-ai-and-disinformation-during-elections/](https://haydencenter.gmu.edu/event/the-dangers-posed-by-ai-and-disinformation-during-elections/).

"The Dangers Posed by AI and Disinformation During Elections." FyiGOV – You Choose the Events We Track It For You, [fyigov.com/event/the-dangers-posed-by-ai-and-disinformation-during-elections/](https://fyigov.com/event/the-dangers-posed-by-ai-and-disinformation-during-elections/).

"Governance of AI." United Nations University, [unu.edu/cpr/governance-ai](https://unu.edu/cpr/governance-ai).

EDMO – *United Against Disinformation*, 8 July 2024, <https://edmo.eu/>

"European Digital Media Observatory (EDMO)." *Shaping Europe's Digital Future*, [digital-strategy.ec.europa.eu/en/policies/european-digital-media-observatory](https://digital-strategy.ec.europa.eu/en/policies/european-digital-media-observatory).