

PRESS RELEASE

New UNESCO report warns that Generative AI threatens Holocaust memory

Paris, 18 June 2024 – A UNESCO report published today warns that unless decisive action is taken to integrate ethical principles, AI could distort the historical record of the Holocaust and fuel antisemitism. The report cautions that not only can Generative AI enable malicious actors to seed disinformation and hate-fueled narratives, but it can also inadvertently invent false or misleading content about the Holocaust.

Download the UNESCO report: AI and the Holocaust: rewriting history

"If we allow the horrific facts of the Holocaust to be diluted, distorted or falsified through the irresponsible use of AI, we risk the explosive spread of antisemitism and the gradual diminution of our understanding about the causes and consequences of these atrocities. Implementing UNESCO's Recommendation on the Ethics of AI is urgent so that younger generations grow up with facts, not fabrications," warned Audrey Azoulay, Director-General of UNESCO.

Published in partnership with the World Jewish Congress, the UNESCO report explains that as learners increasingly use Generative AI to complete assignments and find information online, they risk exposure to distorted information about the Holocaust, which has found new ways to spread through AI-generated content. With four in five (80%) of young people between the ages of 10 and 24 now using AI several times a day for education, entertainment and other purposes, action to ethically guide these new technologies must be taken quickly.

Al can absorb and amplify society's biases

Generative AI must be trained using vast amounts of data. This data is often mined from the Internet and may include misleading or harmful content. AI systems therefore inherit human biases, potentially misrepresenting information about specific events, reinforcing prejudices. This is particularly true in the context of Holocaust, because of the prevalence disinformation about this event. The report notes that due to lack of supervision, guidance and moderation by AI developers, generative AI tools may also be trained on data from Holocaust denial websites.

Al has also been documented to enable bad actors to distort Holocaust-related content, creating fabricated testimonies, and even altering historical records. Deepfake images and audio content created using Generative AI are particularly convincing for young people, who may encounter them on social media platforms. The Historical Figures App allowed users to chat with prominent Nazis such as Adolf Hitler and Joseph Goebbels, and falsely claimed that individuals such as Goebbels were not intentionally involved in the Holocaust and had tried to prevent violence against Jews.

"Hallucinating" history due to missing data

Generative AI models are prone to inventing or "hallucinating" events, personalities and even historical phenomena when they do not have access to sufficient data. The report underscores that ChatGPT and Google's Bard, have both produced content detailing Holocaust-related events which never took place. ChatGPT entirely fabricated the concept of 'Holocaust by drowning' campaigns in which the Nazis

drowned Jews in rivers and lakes, and Bard generated fake quotes from witnesses to support distorted narratives of Holocaust massacres.

In addition to the risk of manipulations and "hallucinations" undermining established facts and trust in experts, the report also underscores the dangers of AI's tendency to oversimplify complex history, privileging a narrow range of sources and a small selection of events. On search engines, which are powered by AI, 60-80 % of the top image results are of a single Holocaust site, Auschwitz-Birkenau.

A solution: act decisively to implement UNESCO's ethical principles

UNESCO calls on governments to accelerate the implementation of the Organization's <u>Recommendation</u> on the Ethics of AI, the first and only global standard in this area, unanimously adopted by its Member States in 2021. This Recommendation is already being integrated into the legislation of over fifty countries.

UNESCO urges tech companies too to implement its standards, and to shoulder their responsibilities and ensure that principles such as fairness, transparency, human rights and due diligence are built into applications at the design stage. In February 2024, <u>eight tech companies were the first to sign a commitment</u> to the Director General of UNESCO to apply an ethical approach to the development of new AI tools.

Tech companies must work closely with the Jewish community, Holocaust survivors and their descendants, educators, experts in antisemitism and historians when developing new Al tools.

UNESCO also appeals to education systems to help safeguard the facts of the Holocaust by equipping young people with digital literacy and critical thinking skills, as well as a sound understanding of the history of this genocide.

Further Reading

- UNESCO report: AI and the Holocaust: rewriting history
- UNESCO report: Bias Against Women & Girls in Large Language Models
- UNESCO Recommendation on Ethics and Artificial Intelligence
- UNESCO's Guidance for generative AI in education and research
- <u>Addressing antisemitism through education</u>
- What UNESCO does to address hate speech

About UNESCO

With 194 Member States, the United Nations Educational, Scientific and Cultural Organization contributes to peace and security by leading multilateral cooperation on education, science, culture, communication and information. Headquartered in Paris, UNESCO has offices in 54 countries and employs over 2300 people. UNESCO oversees more than 2000 World Heritage sites, Biosphere Reserves and Global Geoparks; networks of Creative, Learning, Inclusive and Sustainable Cities; and over 13 000 associated schools, university chairs, training and research institutions. Its Director-General is Audrey Azoulay.

"Since wars begin in the minds of men, it is in the minds of men that the defenses of peace must be constructed" – UNESCO Constitution, 1945.

More information: www.unesco.org

MEDIA CONTACT

Clare O'HAGAN +33 (0) 145 68 17 29 c.o-hagan@unesco.org

