

SYNTHETIC MEDIA, GENERATIVE AI AND DEEPFAKES

WITNESS' RECOMMENDATIONS FOR ACTION

GUIDING PRINCIPLES

- Center those protecting human rights and democracy at the frontlines, the harms they experience, and the solutions they prioritize.
- Place firm responsibility across the AI, tech, and information pipeline.
- Push for human rights-based technical infrastructure, norms, platform policies, laws, and regulations.

RECOMMENDATIONS

- 1** Technology companies, legislators and policymakers should identify risks, create threat models, and prioritize solutions centering global perspectives that are grounded in the lived, practical and expert experience of vulnerable and excluded people facing similar harms.
- 2** Media and digital literacy are not enough, but are still as necessary as ever. While detection and authentication technologies can provide signals to help verify content, the media has a critical role in not promoting naive 'spot with your eyes' solutions that place all responsibility on the consumers of information.
- 3** Human rights and accessibility concerns should shape the design and implementation of watermarks, labels, and other solutions that can help verify provenance (e.g., C2PA and CAI).
- 4** Tools for detecting synthetic media and generative AI content should be accessible to those that need them most. Equity in access to detection tools and capacities is critical to ensure that civil society and media can have tools designed with their needs in mind, as well as the relevant skills needed to use them. For deepfakes that may have a significant impact on society, access to forensic expertise is critical.

- 5** Platforms and messaging apps should support research, development, and deployment of more accessible tools to explain and contextualize 'shallowfakes' (e.g., mis-contextualized images and video), including intuitive reverse image and video search and other forms of easily providing context to visual content.
- 6** Platform policies, regulations, and laws on content moderation (particularly around satire and other gray areas) should incorporate internationally recognized human rights standards for freedom of expression.
- 7** 'Fortify the Truth': Improve journalistic and human rights reporting and advocacy practices to respond to a changed reality.

Building upon years of WITNESS' foundational research and global advocacy on synthetic media, our **Technology Threats and Opportunities Team** has been preparing for the impact of AI on our ability to discern the truth.

In consultation with human rights defenders, journalists, and technologists on four continents, we've identified the most pressing concerns and recommendations on what we must do now.

witnessgenai.global

WITNESS
SEE IT FILM IT
CHANGE IT