

The latest artificial intelligence safety report

2. Deepfakes are improving and proliferating

The report describes the growth of deepfake pornography as a “particular concern”, citing a study showing that 15% of UK adults have seen such images. It adds that since the publication of the inaugural safety report in January 2025, AI-generated content has become “harder to distinguish from real content” and points to a study last year in which 77% of participants misidentified text generated by ChatGPT as being human-written.



The report says there is limited evidence of malicious actors using AI to manipulate people, or of internet users sharing such content widely – a key aim of any manipulation campaign.

The Guardian, Dan Milmo Global technology editor - Tue 3 Feb 2026

Useful vocabulary

Word	Definition
deepfake	fake video/image made by AI
to proliferate	to spread rapidly, increase
to distinguish	to see the difference between things
to misidentify	to identify incorrectly
malicious actor	person with bad intentions
to manipulate	to control or influence dishonestly
particular concern	special worry
authentic	real, genuine