

Google admitted its AI created deepfakes with terrorism and illegal child-related content

These findings are included in the company's report submitted to the Australian eSafety Commission.

From April 2023 to February 2024, Google recorded 258 cases of generating content linked to terrorism and extremism, as well as 86 instances of creating materials exploiting children.

The company stated that it strictly combats the latter type of content using hash-matching for immediate removal but does not yet apply such mechanisms to fight extremist deepfakes.

The Australian commission praised Google's transparency, calling the report "unique in the world."

At the same time, other companies faced penalties: X and Telegram were fined for insufficient measures in combating harmful content.