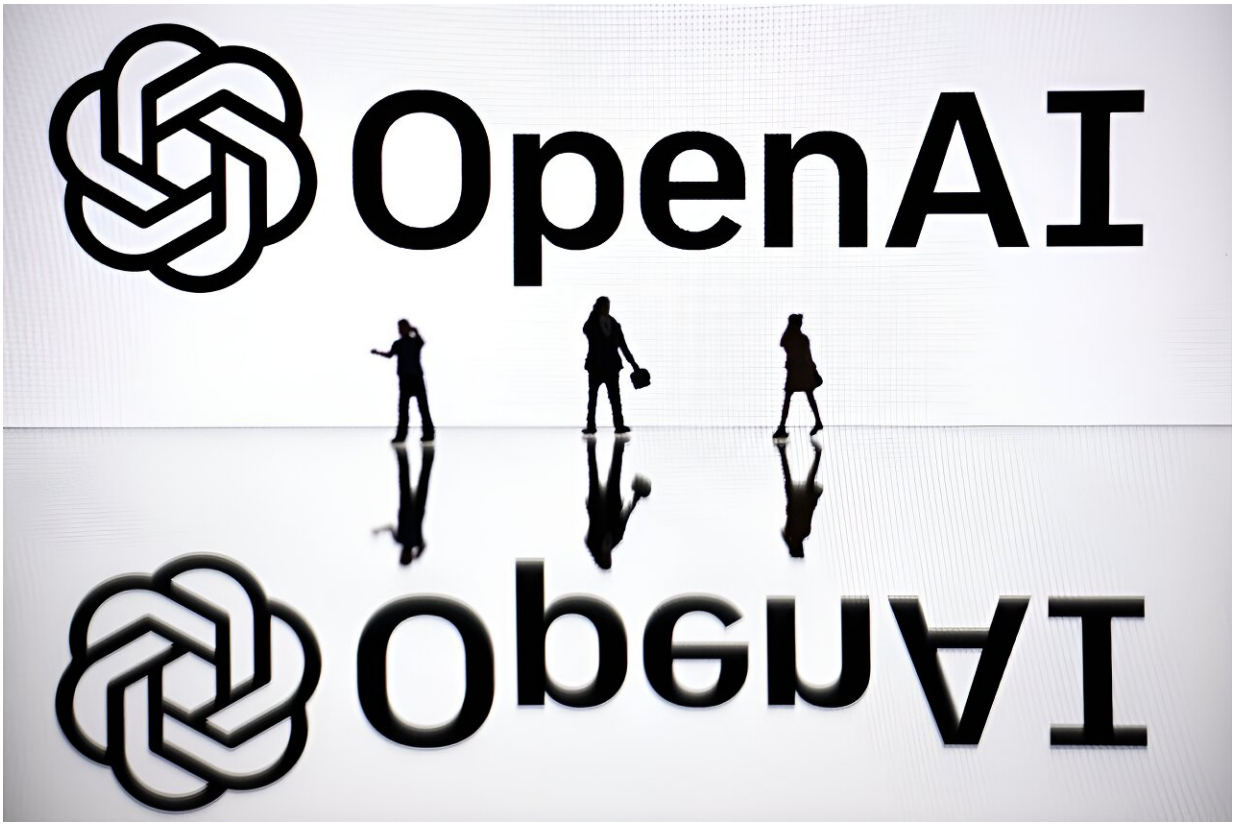
OpenAI unveils tool to detect DALL-E images

May 8 2024



OpenAI has announced the launch of a new tool aimed at detecting whether digital images have been created by artificial intelligence.

OpenAI, the Microsoft-backed artificial intelligence company behind the popular image generator DALL-E, on Tuesday announced the launch

of a new tool aimed at detecting whether digital images have been created by AI.

Authentication has become a major concern in the fast development of AI, with authorities worried about the proliferation of deep fakes that could disrupt society.

According to the company, OpenAI's image detection classifier, which is currently under test, can assess the likelihood that a given image originated from one of the company's generative AI models like DALL-E 3.

OpenAI said that during internal testing on an earlier version, the tool accurately detected around 98 percent of DALL-E 3 images while incorrectly flagging less than 0.5 percent of non-AI images.

However, the company warned that modified DALL-E 3 images were harder to identify, and that the tool currently flags only about five to 10 percent of images generated by other AI models.

OpenAI also said that it would now add watermarks to AI image metadata as more companies sign up to meet the standards from the Coalition for Content Provenance and Authenticity (C2PA).

The C2PA is a tech industry initiative that sets a technical standard to determine the provenance and authenticity of digital content, in a process known as watermarking.

Facebook giant Meta last month said it would begin labeling AI-generated media beginning in May using the C2PA standard. Google, another AI giant, has also joined the initiative.