Detecting Deepfake Creation

3 min

Deepfake content, whether audio, video, or image-based, is becoming more realistic. AI models continue to self-improve as well as their outputs. Despite this, researchers have worked hard to create tools to help detect this content.

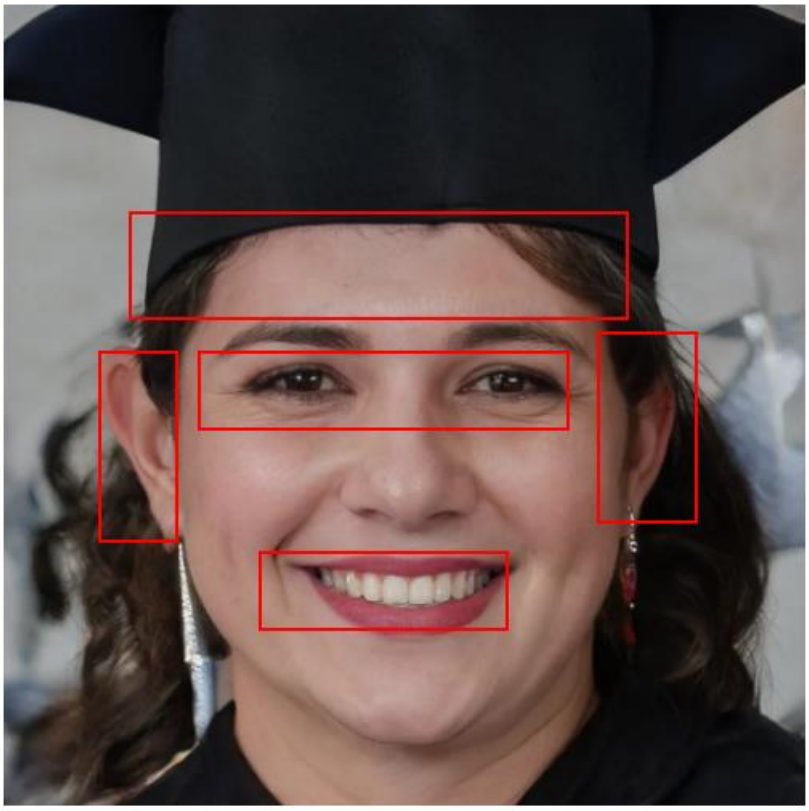
While ironic, one of the most effective ways to identify AI-created content is to employ AI. Instead of using AI algorithms for generation, some researchers have begun to use it detect this content. A great example of this is Intel's "FakeCatcher." Instead of employing systems designed to identify fake content, Intel has taken a slightly different approach, training its FakeCatcher model to identify real content. Intel can detect fake content by looking for unique biological factors or a lack thereof! Things like blood flow create minute but measurable changes that AI algorithms do not capture. Other groups have taken a similar approach to tracking biological factors such as eye movement, which is difficult for AI algorithms to generate accurately.

Similarly, there is a huge interest in identifying deep fake audio. Organizations such as ASVspoof have created biannual challenges to attract talented researchers and engineers to demonstrate their detection solutions. These challenges highlight the need for effective detection methods and the ingenuity and creativity of the developers of these experimental systems.

Of course, while AI is a powerful tool for identifying deepfake content, another equally powerful solution is human ingenuity. While deepfakes may be convincing, there are generally many telling signs. Faces features may be off, certain aspects may sound off, or the messaging within a video/audio snippet may seem unreal. Nowadays, one of the most important things we can do when interacting with content is cross reference and verify the messaging ourselves. If something, even negligible, seems off, we must look to other sources for supporting evidence.

Task: The image in the browser is an AI-generated image. Click in the rectangles to see why this AI-generated image is out of place.

Click on the image below:



Mouse Coordinates: (263, 272.9749984741211)