MediaWell

News Item

Facebook develops new method to reverse-engineer deepfakes and track their source | The Verge

By James Vincent
June 18, 2021

> Deepfakes aren't a big problem on Facebook right now, but the company continues to fund research into the technology to guard against future threats. Its latest work is a collaboration with academics from Michigan State University (MSU), with the combined team creating a method to reverse-engineer deepfakes: analyzing AI-generated imagery to reveal identifying characteristics of the machine learning model that created it.
>
> The work is useful as it could help Facebook track down bad actors spreading deepfakes on its various social networks. This content might include misinformation but also non-consensual pornography — a depressingly common application of deepfake technology. Right now, the work is still in the research stage and isn't ready to be deployed.
>
> Previous studies in this area have been able to determine which known AI model generated a deepfake, but this work, led by MSU's Vishal Asnani, goes a step further by identifying the architectural traits of unknown models.
>
> […]

Source: Facebook develops new method to reverse-engineer deepfakes and track their source | The Verge