KCS Group Europe

‘Seeing is no longer believing’ – the concerns with Deepfake manipulation

July 2023

How sure are you that your business decisions are not based on AI-generated content that is infected and manipulated? According to The Dead Internet Theory, most of what we read on the internet today is the product of automatically generated content controlled by curated algorithms. ‘Content’ includes views, follows, likes and subscribers – but also comments left by alleged account holders and even complete blog posts and articles. The Dead Internet Theory claims the internet ‘died’ at some point in 2016, or maybe 2017. At the time, many would have refused to accept such a possibility. However, given what the world has witnessed in AI capabilities and advances, we know that this is and perhaps was indeed possible.

Given the number of cyber vulnerabilities that have occurred in recent years and the rise of AI intelligence, the DIT has gained momentum. Everyone – from researchers to futurists – is joining the debate about society’s dependence on the internet and for how long the internet and its content will be valid. The DIT believes that a vast amount of the content we consume on the internet is created by Large Language Models (LLMs) – including ChatGPT, OpenAI’s household name platform.

Fabricated human activity

In 2021, *The Atlantic* published an article, ‘Maybe You Missed It, but the Internet “Died” Five Years Ago’. It cites examples of the same profile photos and similar wording being used again and again, just slightly reconfigured, thousands of times over. This provides some evidence of bot-created chats – or else coincidence that must be considered extreme.

In another article from 2018, 'Everything That Once Seemed Definitively and Unquestionably Real Now Seems Slightly Fake', Max Read reported the fact that most web traffic originates from bots. The article opens with details of indictments against several people for a $36 million scam. In two of the largest Ad-Fraud operations in history, these criminals managed to infect 1.7 million computers. They used malware to direct traffic to their own websites where video content of client advertisers could be viewed. The sophisticated bots would then imitate human behaviour by clicking on and browsing content. Each of these 1.7 million bots also had forged identities and forged social media accounts that generated impressive (albeit false) performance statistics.

Such activities are thought to have been prolific as far back as 2013, with YouTube traffic, for example, being more than 50% bots. The situation was such that the Google-owned media giant feared 'inversion', where algorithms can no longer differentiate between bots and humans.

Max Read adds that social networking behemoth Facebook (FB) is also aware of over-inflated user numbers. In October 2018, advertisers filed a lawsuit against FB accusing it of significantly overstating the time its users watch videos – by as much at 60–80%. These are huge numbers. Given FB's enormous advertising revenue (nearly $85 billion in 2020 and $113.6 billion in 2022), one can imagine that they would not hurry to reveal the true scale of any fabricated popularity. According to reports, FB has twice admitted to misreporting its advertising influence.

Deepfake audio and video

The digital fabrication of online human activity has now evolved into deepfake software. This enables the video we see and the audio we hear to be replicated and manipulated with disturbingly convincing accuracy. A human's face can be replaced with another – for example, a politician's face placed on top of a porn star's. Advancing AI image-generating software like Midjourney and DALL-E allows users to type in a simple prompt requesting any photo-real image – or even real-life video output. This year, video and photo artist Boris Eldagsen famously submitted an AI-generated image to the Sony World Photography Awards contest and won, with judges none the wiser. As inversion increases – and as Eldagsen demonstrated – will humans (let alone algorithms) be able to differentiate real from fake?

In 2022, the World Economic Forum (WEF) reported the finding that 66% of IT Security professionals experienced deepfake attacks within their organisations. Examples of attacks include fake audio messages to accounts departments requesting money transfers or disclosure of restricted information. The WEF's 2022 Global Risks report described how a bank manager was tricked into transferring US$35 million to a criminal bank account in 2021.

Are our politics real?

The lack of firm foundations from which humans can set reference points to 'reality' is disconcerting. Reports of deepfake manipulation influencing public opinion, election outcomes, and social and national security create a sense that we humans are being relentlessly scammed, influenced and lied to by controlling elites.

In April this year, US public radio network NPR reported that AI-generated deepfake media creation is advancing so fast that policymakers can't create guidelines quickly enough to police them. In that same month, the Republican National Committee used AI deepfake video to create a 30-second ad that crystal-ball gazed into what President Joe Biden's second term could involve. The video, shown in the form of fake news reports, revealed issues ranging from a Chinese/Taiwan invasion to a complete shutdown of San Francisco. Some weeks before, deepfake images of former President Donald Trump fighting with police sent viewers into a frenzy. This prolific use of lifelike, deepfake content is now causing experts real problems. Talk of watermarking AI images and video, and restricting certain word use for text-to-image or text-to-video has already been made law in most US states, with particular restrictions on the names of those in office. However, this will have little to no effect on those bad actors looking to deceive. Conversely, the individuals that have been caught on real video committing real offences could cry 'deepfake' and the masses would have to be a bit sceptical, at the very least.

Summary

There is an ever-growing demographic engaging in content which promotes the belief that nothing we're presented with can be relied upon and that the real truth is out there somewhere, hidden under lock and key. Every year, businesses spend billions on acquisitions and investments. Some decision-makers are relying on risk reports that have been prepared using manipulated and infected data. Don't be one of them. We can no longer believe everything we see.