36 CYBER BITES

Integrity of Content, Part II: Is Blockchain Technology the Answer to Fighting Disinformation and Deepfakes?

The extraordinary events of this year have created a heightened sense of anxiety, insecurity, and urgency on the issue of how humanity deals with disinformation and 'fake news'.

There is perhaps little surprise that tech giant Google took the decision last week to make all advertisers prove their identity so that internet users can see who is running the ad, and in which country the advertiser is located.

This is Google's answer to combatting the spread of fake content online, proliferated by the coronavirus pandemic. This is Google, perhaps anticipating the changes in legislation that will create and cement the duty of care of a tech company towards its online users in protecting them from harm. The UK government has already signalled these changes in its Online Harms White Paper.

This is Google, it could be argued, taking a responsible stance and embracing accountability for the content on its platforms.

Starting this summer, Google users will begin seeing important disclosure regarding advertisers, firstly in the US, and then gradually expanding worldwide.

Before running an ad on any Google platform, advertisers will need to undertake *"an identity verification process"* by providing personal identification and business incorporation documents or other documentation that proves their identity and country of origin. In short, proving they are legitimate. Users can then see the identity by clicking an *"about the advertiser"* option beside the promotion, so there's full transparency between the advertiser and the online public/users.

This is a salutary step forward in creating a safer online environment for consumers as well as brands by combatting "malvertising", "fake news", and disinformation.

It is noteworthy that advertising generates the vast majority of revenue in the search and social media portions of the Internet industry. Companies like Google and Yahoo rely heavily on income from advertising. According to Search Engine Watch, 47% - 64% of total website traffic last year came through search engines. Facebook and Twitter have become common mediums for communication and entertainment, resulting in heavy traffic and access to substantial user data. User volume and targeted advertising are lucrative tools, and social networks have exploited this advantage substantially. It would be interesting to see how this change in checking advertisers' identity is going to impact social networks, and other digital platforms. Will other tech giants follow suit?

One thing is clear, there has been an increase in threat 'actors' (criminals) buying traffic from ad networks and using "malvertising" as their primary method of attack. Leveraging user profiling from ad platforms, criminals are able to conduct successful drive-by attacks, infecting advertising campaigns with malware. Drive-by attacks often don't require any user interaction (so there's nothing to click on), and the infection is invisible.

This is what John Canfield, the ad integrity and product management director at Google was addressing in his statement:

"This change will make it easier for people to understand who the advertiser is behind the ads they see from Google and help them make more informed decisions when using our advertising controls. It will also help support the health of the digital advertising ecosystem by detecting bad actors and limiting their attempts to misrepresent themselves."

Moreover, with the development and proliferation of Deepfakes, protecting and preserving the integrity of visual and audio content online has become paramount. "Malvertising" and Deepfakes pose serious threats to society, individuals, and businesses alike, through the erosion of trust. To quote one of my clients: *"The thought of not being able to trust the integrity of the information I use to manage my business is a terrifying prospect."*

If erosion of trust and integrity of content are the central most important issues to be addressed, then what is the ideal structure by which they could be restored? Arguably, it would be decentralized (so no single arbiter of truth) and public (we can all check it), which is precisely what Bitcoin's blockchain provides for payments.

Because blockchain technology offers an immutable public record, an audit trail of each transaction, any falsified data stored on the chain should generally be easy to spot. For example, if an ad company says it will produce 500 blockchain-linked ads, but secretly produces 501, this extra entry will be apparent to anyone looking at the chain. The blockchain digital ledger can be the answer to storing information about the advertiser's identity, location, products/services advertised and to facilitate tracking payments.

This would highlight the use of blockchain technology with regard to stopping/reducing unauthorized ad traffic, combatting the spread of "malvertising", and, *"limiting the advertisers' attempts to misrepresent themselves"*, to repeat the words of Google's ad integrity director.

Magnifying the problem is the fact that, at least for the moment, most of the world's Internet users are unfamiliar with Deepfakes.

A tech-savvy 'actor' can make a very real-looking video and distribute it on a variety of platforms, thus going viral before anyone notices it is a fake; and with potentially devastating consequences.

Blockchain can help address this problem, even if it won't fully resolve it. Storing a video's original hash value as a "signature" on a public blockchain digital ledger, for example, could help authenticate originals and weed out fakes. Any video/audio whose hash value/signature doesn't match the original hash value on the blockchain would be identifiable as fake.

As the blockchain ledger can store and protect the title of ownership of intellectual property for example, so it can trace a video's/audio's source. If the source is untraceable, then one can assume it's fake.

If the source is clear, one can then check the hash value of the original video/audio against the value of the video/audio on government websites or major media outlets like the BBC, CNN, etc. If one finds that the video/audio in question has a different hash value from those on trusted platforms, then it is probably a fake one.

In the quest to prove authenticity, there may be the need for a device similar to crypto hardware wallets that one can use to confirm identity by linking it with the blockchain in some way. Ideally the challenge would be to develop a "multisignature" solution that will enhance and preserve security.

Blockchain technology, of course, does not and cannot solve the problem of human lies or falsification, but, arguably, it can go a long way to preserving the integrity of content and thus combatting the spread of disinformation and Deepfakes.

FLAVIA KENYON

36
COMMERCIAL