Tomorrow's Affairs
Analysis of today
Assessment of tomorrow

By: Tomorrow's Affairs Staff

The elusive advantage of the dark side of AI technologies for the production of media content

Perhaps you were fooled by a series of photos in which New York police officers arrest and try to subdue Donald Trump while he resists.

Perhaps you even thought for a moment that it was possible and authentic. No one can blame you for that, as the public has been almost obsessively anticipating the former US president's first appearance before police and prosecutors in New York.

Or perhaps you thought, even for a second, that the photo of Vladimir Putin kneeling and kissing the hand of Chinese leader Xi Jinping during their recent talks in the Kremlin was real.

Like all misinformation, these fake photos have an air of authenticity, or they were created in a context that will make you think that what you see is possible.

Although the creators of the photo showing Trump's arrest and Putin's humility before Xi wrote that it was done with the use of AI, millions of people who shared it on social media did not mention that.

None of them thought it was necessary, not even as a warning. Many did not even read the information.

Everyone thought the scene was possible, or they cynically wanted something like that to be authentic.

Disinformation is a primary threat

GCHQ director Jeremy Fleming recently informed the UK cabinet that disinformation had been one of the primary threats from artificial intelligence, warning of the importance of people being aware of that.

UK Prime Minister Rishi Sunak announced that AI policy could be one of the most important in the next few years, given its impact on the economy and national security, as his spokesperson said.

After the initial enthusiasm for the potential of AI in creating text, audio, or video content, and particularly the possibilities of text generators like ChatGPT4, concern about the dark side of our AI assistants has been growing.

“This tool is going to be the most powerful tool for spreading misinformation that has ever been on the internet. Crafting a new false narrative can now be done at dramatic scale”, said Gordon Crovitz from NewsGuard, a company that tracks online misinformation.

Eight years of advantage in favour of fake-news

Playing and outsmarting ChatBot platforms is over. Things get serious when all the perceived flaws are put into the context of real events: next year's US presidential elections, for example.

If Russia's malignant digital meddling in the 2016 elections was "to an industrial scale”, as assessed by a Columbia University study last year, what can we expect eight years later?

We have dramatically advanced content creation technology on digital platforms, with the dominant participation of AI, to the point where human participation is sometimes reduced to zero.

People's ability to separate the real from the fake has not improved nearly as much since 2016, despite all the attempts to raise awareness and strengthen the mechanisms that could help it.

In the collision of advanced technology that produces fake content and narratives and people's unchanged ability to defend themselves, the winner is unfortunately already known.

Contamination of the online sphere

Our experiences with the latest releases of AI-powered tools, like ChatGPT, show that their imperviousness to misinformation comes from the very foundation on which they were built.

The content we receive from our AI assistants comes from the existing online "knowledge" which has been sufficiently contaminated with misinformation that purification on a technological level is currently impossible.

We are entering a new era of spreading misinformation and fake news, but with familiar strategies, which rely above all on fact-checking.

On that basis, more digital applications are being developed using AI in the opposite direction to expose the mistakes AI made in its first steps.

One such platform, developed by experts in Italy, recognises technological steps in creating some manipulative audio, photo, or video content. Through the reverse engineering process, it removes those changes and returns the content to its original.

These and similar tools will help journalists and fact-checkers avoid falling into the trap of fake content. But the field of public perception they care about is still much smaller than the digital expansion of social media where there is no such protection.

General distrust in information sources

Despite some progress last year, news sources have failed to fix their trust problem. This has been the conclusion of the annual Edelman Trust barometer for 2023, an index that measures global trust in state institutions, business, and media.

This also means that a significant part of the global audience does not trust sources of information that responsibly approach cleaning them from contamination with fake content.

Their trust in the media, whether traditional or social networks, still hovers around 50%.

The only effective defence against misinformation still comes down to the human factor and its ability to recognise the authenticity problem, and warn others about it.

The production of new technical tools to combat fake content has been adapted to this.

But the effect remains small due to general distrust in almost all institutions as sources of

information, particularly governments, and the media.

The technology of producing fake reality in this context has an unattainable advantage over the defenders of the authentic.

Thanks to all of us who, out of ignorance, amusement, or malice, has at least once shared fake news or a photo on social media.