New results in AI research: Humans barely able to recognize AI-generated media

May 21 2024, by Felix Koltermann

Visualization to the Paper "A Representative Study on Human Detection of Artificially Generated Media Across Countries." Credit: CISPA

AI-generated images, texts and audio files are so convincing that people are no longer able to distinguish them from human-generated content. This is the result of an online survey of about 3,000 participants from Germany, China, and the U.S. This is the first time that a large transnational study has examined this particular form of media literacy.

CISPA-Faculty Dr. Lea Schönherr and Professor Dr. Thorsten Holz presented the results this week at the [45th IEEE Symposium on Security and Privacy](#) held in San Francisco May 20–23. The [study](#) was conducted in cooperation with Ruhr University Bochum, Leibniz University Hanover, and TU Berlin and is posted to the *arXiv* preprint server.

Due to the [rapid developments](#) in the field of artificial intelligence, masses of images, text and audio files can now be generated with just a few clicks.

Professor Dr. Thorsten Holz explains the risks that are associated with this in his view, "Artificially generated content can be misused in many ways. We have important elections coming up this year, such as the elections to the EU Parliament or the presidential election in the U.S. AI-generated media can be used very easily to influence political opinion. I see this as a major threat to our democracy."

Against this background, the automated recognition of AI-generated media is an important research challenge. "But this is a race against time," CISPA-Faculty Dr. Lea Schönherr explains. "Media created with newly developed AI generation methods are becoming increasingly difficult to recognize using automatic methods. That's why it ultimately depends on whether a human can make appropriate assessments."

This consideration was the starting point for investigating whether humans are able to identify AI-generated media at all.

Most participants classify AI-generated media as man-made

The results of their transnational cross-media study are astonishing. "We are already at the point where it is difficult, although not yet impossible,

for people to tell whether something is real or AI-generated. And this applies to all types of media: text, audio, and images," Holz explains.

Across all countries and media types, the majority of study participants classified AI-generated media as man-made.

"We were surprised that there are very few factors that can be used to explain whether humans are better at recognizing AI-generated media or not. Even across different age groups and factors such as educational background, political attitudes or media literacy, the differences are not very significant," Holz elaborates.

Study includes socio-biographical data

Between June 2022 and September 2022, the quantitative study was conducted as an online survey in China, Germany, and the U.S. Respondents were randomly assigned to one of the three media groups "text," "image," or "audio" and were shown 50% real and 50% AI-generated media. In addition, socio-biographical data, knowledge of AI-generated media as well as factors such as media literacy, holistic thinking, general trust, cognitive reflection, and political orientation were collected.

After data cleansing, 2,609 data sets remained (822 U.S., 875 Germany, 922 China), which informed the analysis. The AI-generated audio and text files used in the study were generated by the researchers themselves, while the AI-generated images were taken from an existing study. The images they used were photorealistic portraits, the texts were news items, while the audio files were excerpts from literature.

Starting points for further research

The study results provide important takeaways for cybersecurity research. "There is a risk that AI-generated texts and audio files will be used for social engineering attacks. It is conceivable that the next generation of phishing e-mails will be personalized to me and that the text will match me perfectly," Schönherr explains. She believes that developing defense mechanisms for precisely such attack scenarios is an important task for the future.

However, further research desiderata also emerge from the study. "Firstly, we need to better understand how people can still recognize AI-generated media at all. We are planning a laboratory study where participants will have to explain to us how they recognize whether something is AI-generated or not. On the other hand, we need to consider how we can support this technically, for example through automated fact-checking processes," Schönherr concludes.

More information: Joel Frank et al, A Representative Study on Human Detection of Artificially Generated Media Across Countries, *arXiv* (2023). DOI: 10.48550/arxiv.2312.05976

Provided by CISPA Helmholtz Center for Information Security