THREAT
INTELLIGENCE
FORECAST

The impact of
GENERATIVEAI
on the Malvertising
Landscape

November 2023

Advancements in generative artificial intelligence (AI) offer threat actors exploiting digital advertising access to new tools that can augment the scale and scope of their malvertising campaigns.

EXECUTIVE SUMMARY

Cybercriminals are all but certain to leverage generative AI to enable malvertising campaigns in the coming year. Generative AI is easy to adapt and can enhance the quantity and quality of malvertising creatives and ad copywriting. We expect these campaigns to initially target English speakers as most early-stage large language models (LLMs) are currently most effective in English. Additionally, AI identification tools are prone to inaccuracy, and, in the background of all of this, AI regulatory efforts are generally centered around consumer data protection, with little to no mention on the criminal use of AI. Finally, there are examples of nation-state actors leveraging AI in disinformation campaigns that could serve as a proof-of-concept for further illegal exploitation. We have high confidence in our assessment based on reliable and corroborated open sources, including academic research, government assessments, qualified subject matter experts, and proprietary TAG Threat Exchange data.

AI GENERATED CONTENT

AI-generated content is easy to create and cost-effective, which makes a malvertising campaign more accessible. Malvertisers can leverage AI-generated content to augment both their ad quantity and quality. Apart from crafting ad creatives, malvertisers can also utilize AI to code the malware associated with malvertising.

- AI content creation tools like ChatGPT and Dall-E can transform a simple prompt into human-like prose or a stylized image, which saves time and money for malvertisers. These tools allow malvertisers who need more technical advertising skills and content creation skills, such as grammar and graphic design, to launch their malvertising campaign quickly.[i], [ii]
- Malwarebytes found multiple ad campaigns that mimicked big brands using search engine optimization (SEO) poisoning in May. AI content generators could enhance this method of attack because AI can manipulate SEO with keywords to position ad content higher in a user's search results.[iii], [iv], [v] Furthermore, Malwarebytes has also identified malicious sponsored ads alongside AI-assisted search results.[vi]
- In July, hackers released WormGPT, ChatGPT's dark web antithesis. Designed by hackers for hackers, WormGPT codes malware and aids criminal requests. [vii]
- A combination of dark web forums and collegiate research indicates commercial and open source, not just dark web, LLMs can be used to write malware code.[viii], [ix]

OVERCOMING LANGUAGE BARRIERS

The developers of most commercially available LLMs initially trained them with the English-speaking internet. Consequently, these LLMs work more effectively in English or when translated into English. This bias towards English will likely evaporate over time as more geographically diverse models come online; however, AI-enabled exploitation in English will remain more common over the next year.

- Generative AI will be especially useful in creating more effective scam landing pages, evident in the research published by cybersecurity vendor Darktrace, which indicates threat actors

utilize AI-powered bots to write more effective phishing emails, boosting their previous attack quality and output.[x]
- Based on past malvertising campaigns and historical data from the TAG Threat Exchange, malvertisers tend to target the United States, United Kingdom, Canada, and other countries in the European Union, with most reported ad creatives seen in English, followed by German and Polish.[xi]

LACK OF PUBLISHER SAFEGUARDS AND AI REGULATIONS

Current safeguards against misusing AI-generated content are inadequate to address emerging threats, including AI-enabled malvertising. Moreover, AI regulation primarily concerns consumer data protection and overlooks curbing criminal use. Furthermore, the key tech players in AI's creation cannot reach a consensus on regulatory guidelines.

- AI detection tools still need to be prepared for widespread adoption, with many showing inconsistencies in identifying inaccuracies, biases, and discrimination, making them unusable for detecting AI-generated content and ad creatives.[xii]
- Attempts at AI regulation generally focus on protecting consumer data, not curbing criminal use. The United States' AI Bill of Rights and the European Union's AI Act are in the drafting phase, with language prioritizing consumer data protection.[xiii], [xiv] Notably, Italy banned then unbanned ChatGPT after it met data protection requirements, while countries including Spain, Canada, and others continue to investigate OpenAI's data protection processes.
- While the major technology companies behind many recent advancements publicly discuss AI regulation proposals, establishing a global consensus on AI regulation remains challenging as the landscape is vast.[xv], [xvi]

AI-ENABLED DISINFORMATION AS PROOF-OF-CONCEPT FOR MALVERTISERS

State-sponsored, AI-enabled disinformation campaigns appeared in two recent disclosures by U.S. intelligence and threat researchers. Both broadly illustrate the risks associated with this new technology. While neither incident leverages advertising, their existence is a testament to the malicious applications of even basic generative AI capabilities. Criminal actors are likely to take note of these tactics, follow the lessons learned from their identification, and plausibly apply these techniques to novel applications, including AI-enabled malvertising campaigns.

- Suspected Chinese operatives used AI-generated images in a months-long campaign focused on divisive U.S. political topics, like gun violence, and "denigrating U.S. political figures and symbols" since at least March 2023. According to researchers from Microsoft, Recorded Future, the RAND Corporation, NewsGuard, and the University of Maryland, Beijing's influence campaign using artificial intelligence is a rapid change in tactics.[xvii], [xviii]
- Typically, China-linked influence operations have yet to generate significant authentic engagement on social media; however, recent attempts appear more successful as they increasingly leverage generative AI and real people to create tailored, authentic content.[xix]
- U.S. intelligence says Moscow intends to try to end U.S. and European support for Ukraine by using Russia's spy agencies to push propaganda supporting pro-Russian political parties and stoking conspiracy theories with AI and other new technologies.[xx]

OUTLOOK

Technological disruptions like generative AI put the advertising industry on precarious footing. The advantages to its adoption are myriad, but examples of its misuse are sobering. While AI will enable ad tech to automate complex tasks to better serve relevant and tailored ads to consumers, cybercriminals will leverage these same applications to enhance the effectiveness of their illegal activity. However, the next year presents an opportunity—however fleeting—for the digital advertising industry to invest in the collective defense of its supply chain.

GLOSSARY

Generative Artificial Intelligence (AI) – a type of artificial intelligence capable of producing unique outputs such as texts, images, videos, and more

Large Language Models (LLMs) – an advanced deep learning model that is trained with a vast amount of data, giving it the ability to understand and generate language that is human-like

Search engine optimization (SEO) poisoning – a technique used by cybercriminals to boost the overall ranking of their malicious website to appear higher on search result pages, leading unsuspecting users to click on the site and potentially download malware or other malicious content.

Disinformation – the deliberate creation and dissemination of false or misleading information with the intent to deceive or mislead and is typically used to manipulate public opinion, sow discord, discredit opponents, or create confusion.

AUTHORED BY TAG THREAT INTELLIGENCE

Contact TAG/TI at info@tagtoday.net.

TAG is the Information Sharing and Analysis Organization (ISAO) for the digital advertising industry, a U.S. Department of Homeland Security designation making TAG the primary forum for sharing cyber threat intelligence in our industry.

The TAG Malvertising Threat Exchange (MTX) enables the TAG Community to share real-time intelligence about threats they see, stay abreast of new and emerging threats that could affect their operations, and protect the digital advertising supply chain as a whole.

The MTX enables companies to:
- Leverage a centralized intelligence platform to collaborate within your company, with other companies working to combat the same threat, or with the TAG Community as a whole;
- Share and receive timely, actionable and highly relevant threat intelligence between trusted parties in the TAG Community;
- Enrich, enhance, and shorten your own investigations with high-fidelity intel.

What We Mean When We Say: An Explanation of TAG's Estimative Language

We use phrases such as we judge, we assess, and we estimate—and probabilistic terms such as probably and likely—to convey analytical assessments and judgments. Such statements are not facts, proof, or knowledge. These assessments and judgments generally are based on collected information, which often is incomplete or fragmentary. Some assessments are built on previous judgments. In all cases, assessments and judgments are not intended to imply that we have "proof" that shows something to be a fact or that definitively links two items or issues.

In addition to conveying judgments rather than certainty, our estimative language also often conveys 1) our assessed likelihood or probability of an event; and 2) the level of confidence we ascribe to the judgment.

Estimates of Likelihood. Because analytical judgments are not certain, we use probabilistic language to reflect the Community's estimates of the likelihood of developments or events. Terms such as probably, likely, very likely, or almost certainly indicate a greater than even chance. The terms unlikely and remote indicate a less than even chance that an event will occur; they do not imply that an event will not occur. Terms such as might or may reflect situations in which we are unable to assess the likelihood, generally because relevant information is unavailable, sketchy, or fragmented. Terms such as we cannot dismiss, we cannot rule out, or we cannot discount reflect an unlikely, improbable, or remote event whose consequences are such that it warrants mentioning. The chart provides a rough idea of the relationship of some of these terms to each other.

Remote	Very unlikely	Unlikely	Even chance	Probably/ likely	Very likely	Almost certainly

Confidence in Assessments. Our assessments and estimates are supported by information that varies in scope, quality, and sourcing. Consequently, we ascribe high, moderate, or low levels of confidence to our assessments, as follows:
High confidence generally means good quality of information, evidence from multiple collection capabilities, and possible to make a clear judgment. A "high confidence" judgment is not a fact or a certainty, however, and such judgments still carry a risk of being wrong.

Moderate confidence generally means evidence is open to several interpretations or is credible and plausible but lacks correlation.
Low confidence generally means the assessment is based on fragmentary information, or from collection capabilities of dubious reliability.

Sources

[i] European Union law enforcement agency; Europol; ChatGPT The impact of Large Language Models on Law Enforcement; retrieved 15 August 2023; https://www.europol.europa.eu/cms/sites/default/files/documents/Tech%20Watch%20Flash%20-%20The%20Impact%20of%20Large%20Language%20Models%20on%20Law%20Enforcement.pdf (source reliability – high)

[ii] Global media company; Forbes; Beyond ChatGPT: 14 Mind-Blowing AI Tools Everyone Should Be Trying Out Now; retrieved 15 August 2023; https://www.forbes.com/sites/bernardmarr/2023/02/28/beyond-chatgpt-14-mind-blowing-ai-tools-everyone-should-be-trying-out-now/?sh=61df2f287a1b (source reliability – medium)

[iii] Subject matter expert; Jérôme Segura, Sr. Director of Threat Intelligence at Malwarebytes; Malvertising via brand impersonation is back again; retrieved 15 August 2023; https://www.malwarebytes.com/blog/threat-intelligence/2023/05/malvertising-its-a-jungle-out-there#:~:text=One%20particularly%20devious%20kind%20of,website%20within%20the%20ad%20snippet (source reliability – high)

[iv] Technology marketing company; TechTarget; Pros and cons of AI-generated content; retrieved 15 August 2023; https://www.techtarget.com/whatis/feature/Pros-and-cons-of-AI-generated-content?Offer=abt_pubpro_AI-Insider (source reliability – medium)

[v] Global business newspaper; Financial Times; Google to deploy generative AI to create sophisticated ad campaigns; retrieved 15 August 2023; https://www.ft.com/content/36d09d32-8735-466a-97a6-868dfa34bdd5 (source reliability - medium)

[vi] Subject matter expert; Jérôme Segura, Sr. Director of Threat Intelligence at Malwarebytes; Malicious ad served inside Bing's AI chatbot; retrieved 06 October 2023; https://www.malwarebytes.com/blog/threat-intelligence/2023/09/malicious-ad-served-inside-bing-ai-chatbot#:~:text=Users%20looking%20for%20software%20downloads,powered%20by%20OpenAI%27s%20GPT%2D4. (source reliability – high)

[vii] Computer and network security company; SlashNext; WormGPT – the Generative AI Tool Cybercriminals are using to launch Business Email Compromise Attacks; retrieved 15 August 2023; https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks/ (source reliability – medium)

[viii] Cyber threat intelligence provider; Check Point Research; OPWNAI: Cybercriminals starting to use ChatGPT; retrieved 15 August 2023; https://research.checkpoint.com/2023/opwnai-cybercriminals-starting-to-use-chatgpt/ (Source reliability – medium)

[ix] Academic research; Carnegie Mellon University, Center for AI Safety, and Bosch Center for AI; Universal and Transferable Adversarial Attacks on Aligned Language Models; retrieved 15 August 2023; https://arxiv.org/pdf/2307.15043.pdf (Source reliability - high)

[x] Global news organization; The Guardian; AI chatbots making it harder to spot phishing emails, say experts; retrieved 15 August 2023; https://www.theguardian.com/technology/2023/mar/29/ai-chatbots-making-it-harder-to-spot-phishing-emails-say-experts (source reliability - medium)

[xi] Information Sharing and Analysis Organization (ISAO); Trustworthy Accountability Group (TAG) Malvertising Threat Exchange (MTX) proprietary data from 1 January 2022 – 31 July 2023; retrieved 15 August 2023 (source reliability – high)

[xii] U.S. government; Federal Trade Commission; FTC Report Warns About Using Artificial Intelligence to Combat Online Problems; retrieved 15 August 2023; https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems (Source reliability - medium)

[xiii] U.S. Government; White House; Blueprint for an AI Bill of Rights Making Automated Systems work for the American People; retrieved 15 August 2023; https://www.whitehouse.gov/ostp/ai-bill-of-rights/ (Source reliability – high)

[xiv] Legislative Body of the European Union; European Parliament; AI Act: a step closer to the first rules on Artificial Intelligence; retrieved 15 August 2023; https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence (Source reliability - high)

[xv] Global media company; Forbes; Regulating Artificial Intelligence – Is Global Consensus Possible?; retrieved 15 August 2023; https://www.forbes.com/sites/hecparis/2022/09/09/regulating-artificial-intelligence--is-global-consensus-possible/?sh=1880913d7035 (Source reliability – medium)

[xvi] Business and financial news; CNBC; White House secures voluntary pledges from Microsoft, Google to ensure A.I. tools are secure; retrieved 15 August 2023; https://www.cnbc.com/2023/07/21/white-house-secures-voluntary-pledges-from-microsoft-google-on-ai.html (Source reliability - medium)

[xvii] Global News Organization; The New York Times; China Sows Disinformation About Hawaii Fires Using New Techniques; retrieved 06 October 2023; https://www.nytimes.com/2023/09/11/us/politics/china-disinformation-ai.html (source reliability - medium)

[xviii] Computer Software Vendor; Microsoft; Sophistication, scope, and scale: Digital threats from East Asia increase in breadth and effectiveness; retrieved 06 October 2023; https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW1aFyW (source reliability - high)

[xix] U.S.-based News Site; Axios; Meta takes down largest online influence network in history; retrieved 06 October 2023; https://www.axios.com/2023/08/30/meta-largest-influence-network-china (source reliability – medium)

[xx] Global News Organization; The New York Times; Putin's Next Target: U.S. Support for Ukraine, Officials Say; retrieved 06 October 2023; https://www.nytimes.com/2023/10/02/us/politics/putin-ukraine-spy-united-states.html (source reliability - medium)

tag

tagtoday.net