COLLECTIVE INTELLIGENCE

Generative AI and Misinformation: Implications for Business

Bénédicte Earl and Tom McGivan
March 2024

Misinformation stemming from generative AI could cost organizations millions. Companies should prepare now to mitigate, monitor and respond to potential threats that can arise from bad actors using generative AI to create visual misinformation.

Considerations for business:

While misinformation has long been an issue for organizations, generative AI has made the threat more acute, due to specific points around quality, accessibility, limitations in online monitoring and regulation, governance and platform control.

Quality and accessibility of generative AI

The quality of generative AI has dramatically improved in the past year, owing in large part to the availability of data, improvements in AI models and increases in computing power for training models. With limited tools to "prove" an image's veracity, it poses a significant threat to business if users cannot easily identify visual misinformation. Some experts are concerned about the capacity of newer models to generate more persuasive misinformation at scale, tailored to individual targets.

Increasingly, companies are building into their crisis plans how to spot an AI-generated visual, but these mechanics are not foolproof.

Limitations in online monitoring

There are limitations in monitoring generative AI online, including so-called deepfakes. While many social-monitoring platforms can quickly pick up conversation related to misinformation, most tools are limited in visual-monitoring capabilities and rely on the language around the imagery as they are unable to process images.

Regulation, governance and platform control

On some social media platforms, AI-generated content runs unrestricted. On others, watermarks (both visible and not) and metadata tags are being used to label AI-generated images. In the EU, the Digital Services Act and recently approved AI Act both take strides towards combatting this trend. Nonetheless, for the most part AI technology is accelerating faster than platforms can keep up. In fact, most social media platforms' algorithms can actually reward AI-generated images with in-feed algorithmic promotion, as they tend to be sensationalized and drive engagement.

Ultimately, recognizing AI on social media will require a combination of regulation of content and transparency, new tools for detection, as well as improvements in users' media literacy.

Is your organization ready?

With all this in mind, the threat that generative-AI-enabled misinformation poses to organizations goes beyond reputational – it can be existential if the content is taken to be true by investors, regulators or customers. There have been numerous recent incidents where companies have faced crisis when trust in the organization and their reputation comes under fire as the result of generative AI.

To prepare, businesses should develop an action plan with the following questions in mind:

Enhanced monitoring, escalation and response

- Does your organization have comprehensive, real-time online-monitoring systems in place that include individuals who may be targets for generative AI misinformation?
- Are the escalation processes from these alerts efficient enough that a reaction and potential response could quickly be approved and, if needed, shared?
- Do current crisis-response plans and frameworks include misinformation-specific tools and resources?
- Do crisis plans include scenario planning for AI-generated misinformation, including deepfakes? Is there a response framework in place to understand if, when and how to respond to this online activity, and when to bring in other voices, such as C-suite or third-party influencers?

Effective preemption

- Is your C-suite active online so their identity can be confirmed, and they can be mobilized if needed on social media?
- Can your digital and web team be activated quickly to create a defense landing page in extreme situations?

Internal regulation

- Does your organization have internal policies in place around the use of generative AI, and AI more broadly?

Knowledge and awareness

- How is your organization staying abreast of developments in technological trends, specifically around AI?
- Are you training your organization to be on the lookout for AI-generated misinformation?

To continue the conversation

Bénédicte Earl
Director, London
bearl@brunswickgroup.com

Bénédicte is a specialist in digital strategy, with experience in creating and executing campaigns across geographies, political landscapes and corporate settings. Bénédicte began her career as a journalist at The Sunday Times. She is currently focused on building companies' resilience to mitigate, react and respond to online mis- and dis-information.

Tom McGivan

Associate, London

tmcgivan@brunswickgroup.com

Tom is a specialist in ESG strategy, with a focus on technology ethics, governance and policy. Tom joined Brunswick from Virgin Group, where he worked as a researcher and advisor on behalf of Sir Richard Branson and the senior leadership of the Group's 40-plus companies. He is currently studying for an MSc in data and AI ethics at The University of Edinburgh.