beazley

Article

Emerging Risk Report: AI and Sexual Misconduct Liability

Nisar Siddiqui • October 05, 2023

Generative artificial intelligence (AI) has been a hot topic for much of 2023, and the insurance community has been active in the conversation about AI's potential impact on the risk landscape for a wide range of industries. But while most of these conversations are speculative in nature, centering on future risks and "down the line" concerns, many people would be shocked to learn that the use of AI-generated content in child sexual abuse already presents a significant and growing threat today.

A lot of people incorrectly believe that AI-generated content cannot cause "real" physical harm, but this is unfortunately not the case. Images that were initially posted online for innocent purposes, from parents' social media posts to schools' publicity shots and team photos, can be manipulated using generative AI to create pornographic versions that may then be leveraged by schoolyard bullies and potential offenders alike.

There has been quite a bit of press commentary and legislative and regulatory concern around AI in child sexual abuse image proliferation of late, and in a **recent case in Spain**, AI was used to create naked images of underage girls, which were then distributed around the local region where the victims lived[1].

This case demonstrates how AI can make one of the most horrible situations an organization may face all the more complicated to navigate. Brokers and their clients should educate themselves on this emerging risk, its potential to alter or complicate their risk management oversight and response obligations, and the risk prevention resources that are available today to help them address this emerging threat.

Deep fakes present an emerging risk that needs awareness

There are two primary types of AI deep fakes operating in this

space: **AI-generated fictional images** and videos and **image or video alteration that is accomplished through face swapping.** Both kinds of images have the potential to lead to child sexual abuse image proliferation, worsening the already troubling epidemic of abuse, **as seen in the Spanish case mentioned above .**

In many cases, AI could act as an accelerator for abuse, as the accessibility of AI tools facilitates ease of use, providing increased accessibility for pedophiles seeking pornographic content. This in turn is leading to fears that AI generated content could encourage potential offenders resulting in real life abuse. It is also likely to open the door to a rise in related student-on-student bullying or pranking risk in educational settings. Organizations may find themselves facing a claim when such behavior takes place among its constituents, as in the case of student-on-student bullying. The lack of guardrails in some of these applications creates an added burden for many organizations and in particular education, healthcare, transportation, and leisure entities, who could find themselves at the epicenter of a claim. Given the volume, speed and sophistication of AI generated content and the struggles that online platforms have in policing content, there is increased pressure on these organizations to be vigilant themselves.

Increased vigilance is essential

Similar to the social media policies commonplace in most educational institutions today, organizations need to be mindful of how to respond to an incident. Among the questions that should be addressed immediately is whether the offending image is real, or AI generated – and if it's the latter, who created it, how it was created, and how it has been distributed and shared.

While initially working out what has happened, it's essential to handle this kind of situation with sensitivity and **compassion**. Perceptions of insensitivity or mishandling can easily compound the error in the eyes of victims, their families, the public and, potentially, the courts. When the initial response to such an incident is poorly handled, people may be more inclined to sue as retribution and to force change, increasing the likelihood of a full-blown lawsuit and greater negative attention. A thoughtful and measured response will help mitigate the likelihood of this kind of response.

Key resources to manage these risks are included with every Safeguard policy at no additional charge

Beazley's Safeguard offering is designed not just to help clients minimize the risk of abuse within their organization, but also to respond effectively if an event occurs. Our trusted risk prevention partner, **Praesidium**, provides resources and advice to policyholders at no additional charge.

Now is the time for organizations to evaluate:

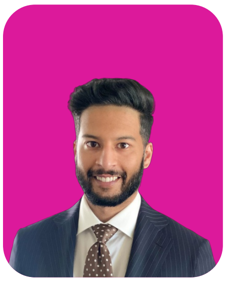
- Do we have defined expectations addressing appropriate boundaries between our team and the consumers served or between consumers?

- Do we have a defined policy addressing social media and electronic communications? Does it address the taking and posting of photos involving program participants?
- Do we have a defined policy addressing technology and internet usage within the organization?
- Do we have screening procedures designed to gather information about potential employees and volunteers, including their criminal history and past behavioral interactions with vulnerable individuals?
- Do we provide information to consumers and/or their families addressing bullying, the organization’s policies, or how to report concerns of any nature?

The Praesidium resources available through Beazley’s Safeguard offering are designed to help organizations navigate these questions and in the event of an allegation, ensure the organizational response process operates smoothly, strategically, and integrates a survivor-focus.

As with many emerging risks, no one has all of the answers today – but organizations also cannot afford to wait for them. As AI’s capabilities and implications continue to evolve at pace, organizations are well-advised to take advantage of all available information and resources.

We are here to help. Find out more about our Safeguard solution **here.**



Nisar Siddiqui

Underwriter - Safeguard

1- Outcry in Spain as artificial intelligence used to create fake naked images of underage girls | CNN

beazley