Culture Club/Human Risk Management SIG
Managing Human Risk from GenAI

Meeting Summary
December 07, 2023

December's Culture Club meeting attracted the largest number of attendees for this SIG all year. Following introductions, participants were asked if they have a GenAI policy within their organization, and almost 80% said they had one. A few mentioned that their policy was in draft form, and others commented that they were in the process of constructing procedures for the use of ChatGPT. A few mentioned that their policy was "no authorized use of ChatGPT." Following this, the meeting's agenda included a hot topics discussion and a discussion around the ways generative AI is impacting the HR side of our organizations.

Hot Topics Session: Security Champions and HRM Metrics

On attendee's list of hot topics today was the discussion of security champions and human risk measurements. The first discussion asked the question about how others are creating successful security champion programs. Actionable insights included creating a feedback loop from champions to the security teams to amplify security messages among non-security employees, setting annual objectives so successes are visible, and using company recognition programs and electronic 'badges' to reward enterprising team members. One participant noted that they had a program, but when the key person left, the program didn't survive. They revamped the program so it was not dependent on one key leader. Members were asked to share employee engagement programs with other SIG members.

The second discussion asked "how do we measure human risk to cyberscurity?" Several attendees shared stories about using results from simulated phishing tests, reporting statistics on failed tests along with additional individual measures (such as completion of required training, position in the company, performance on previous tests, etc). In one company, realizing the lowest click rates from phishing test reports created a healthy competition between departments. Another member reported that they conducted interviews to better understand why phishing tests were failed, and that created a very successful way to share experiences.

Human Risk Management and GenAI

The primary topic for discussion at this meeting was understading the new risks GenAI introduces and identifying managerial actions to take to reduce the risk. Several examples were shared, like deepfake and how it is easy to leverage GenAI tools for a new generation of phishing vectors. Also discussed were hallucinations and inaccurate responses from GenAI tools: Tools sound authoritative but can still be wrong. "The biggest human risk at this point is people taking the information from one of these tools as 'absolutely correct' and taking action without verifying the results," commented one attendee. Members shared their vision that a 'second wave' of GenAI is on the horizon as vendors incorporate AI into their offerings; even as we build guardrails for use within our organizations, our we must prepare for our vendors to use this technology in their offerings.

Not all participants come from organizations embracing GenAI. "It's still not allowed in our network. We are extremely risk adverse and we are still studying how to best bring GenAI tools in. It's a work in progress," said one attendee.

Some actionable insights to manage the human risk from the growing adoption of GenAI included **creating an AI committee** (of tech, cyber, and business representatives) to establish the best controls, policies and guardrails for our culture; **creating multiple avenues of training**-- courses, short videos, just in time training modules-- to help users understand the risks and devise appropriate actions to mitigate the risks; and **executive prioritization** through blogs and other outreach activities to show that GenAI tools can be beneficial but must be used wisely to best fit into the organization.

About Cybersecurity at MIT Sloan

Cybersecurity at MIT Sloan brings together thought leaders from industry, academia, and government with MIT faculty, researchers and students to address strategy, management, governance and organization of cybersecurity of critical infrastructure using an interdisciplinary approach.

For more information, visit https://cams.mit.edu