Editorial: Are we ready for auditing AI-based systems?

Organizations increasingly emphasize compliance with information security policies in conformance to a set of laws and regulations. Audits are the preferred mechanism to evaluate claimed compliance by an organization. Information system audits are conducted to ensure the integrity of corporate information and operations, and confidentiality of corporate assets. With the advent of artificial intelligence (AI) systems, the audit approach needs to shift considerably – while traditional systems require a set of controls to realize the objectives, systems based on AI need to be treated differently. It is not only important to ensure the integrity of the AI system but also to ensure that such system was originally designed to be fair and accurate.

AI systems must be evaluated to ensure that there are no biases in the system based on the data that are used to train these systems. These biases appear in the data based on historical or social inequities or based on over or underrepresentation of specific demographic groups. AI systems have inherent limitations stemming from the limits of the data that are used to train these systems, such as inaccuracies in prediction (e.g. falsely accusing someone of crime based on misrecognition), biases (e.g. denying a student of color admission to a prestigious college since historic data had few students of color being admitted) or sabotage (e.g. tampering with algorithm or providing malformed data). It is considerably hard to conduct audit of such systems as these are often black boxes with little insight into the analytic process and lack of explainability.

The challenge facing us is to figure out how to audit AI systems to test for biases and inaccuracies. Testing the AI systems for biases requires a technical definition of fairness which itself varies. One common way is to use counterfactuals for testing systems whereby manipulating a sensitive variable (with limited influence on decision) should not result in changing the algorithm's decision. For instance, if the gender of the applicant changes from male to female in the input to an admission AI system, the resulting admission decision should remain the same. Testing is not perfect though since the counterfactual test may work with the test set, but there may be differences in different datasets. Generally, audits are performed by independent third parties to engender trust in the audit. The trained auditors typically follow checklists based on a standard to evaluate and benchmark the system against the adopted standard. However, the auditors are limited by their training to understand the subtle nuances of AI systems.

To further compound the problem, the standards to audit AI systems have not yet been established. As AI-based systems are expected to get increasingly deployed, we need to invest in developing effective standards for auditing AI systems along with creating legislation based on harmonizing societal expectations to ensure that such systems are developed to rigorous standards while minimizing biases and inaccuracy. The methods of conducting audit, whether control-based or continuous, takes on an entirely different

Organizational Cybersecurity Journal: Practice, Process and People
Vol. 2 No. 2, 2022
pp. 77-78
Emerald Publishing Limited
e-ISSN: 2635-0289
p-ISSN: 2635-0270
DOI 10.1108/OCJ-08-2022-037

meaning in the context of AI systems. We may need to reconfigure our approach to audit emphasizing improvements in processes rather than excessive focus on compliance alone, overall, addressing the value of audit in improving cybersecurity posture of an organization. At the same time, we cannot ignore to enhance skills and competencies of the next generation of auditors enabling them to effectively audit AI systems.

The papers appearing in this issue nudge us to confront realities faced by the organizations in the light of complexities presented by technologies. Bennet Simon von Skarczinski, Arne Dreißigacker and Frank Teuteberg survey 5,000 German organizations contributing towards improving the information base on the costs of cyber incidents. The findings indicate that most organizations suffer little to no costs, whereas only a small proportion suffers high costs. Marcia Combs and Casey Hazelwood examine the loss of personal privacy from the use of wake words in smart home devices. The findings of this study suggest users can better protect their privacy by selecting "Alexa" as their wake word and selecting smart speakers with the highest number of microphones in the far-field array.

Allen Johnston provides a realistic perspective on extant research identifying the leading organizational cybersecurity research topics of interest. The topic modeling analysis of the organizational level studies indicate the key areas to be governance and strategic level decision-making in shaping organizational security successes and failures, organizations' ability to detect cybercrime and the cost of security negligence. Deborah Richards and Salma Banu Nazeer Khan emphasize the role of ethical principles and policy awareness priming on information and communication technology policy compliance. The results indicate that targeted priming might offer a nonresource intensive approach to cybersecurity training. This study reminds us of the role ethics may play in addressing some of the challenges from emergent technologies.

We sincerely hope our readers will find these research papers to be invigorating.

Sanjay Goel and Gurvirender Tejay