Wednesday, June 26, 2024

9:00 AM - 9:20 AM

Generative AI for DCO: Why Timing and Signal are Key

Zachary Vaughn

Director, Federal Security Engineering

Vectra AI

Abstract:

The DoD's significant challenge in managing the immense volume of data it collects daily is complicated by the enormous effort needed for proper data labeling for training language models to augment defensive cyber operations. In order to better assist and automate the need for such labeling, Vectra AI is able to identify all assets and user identities operating in environments and all associated threat detection information in real time using 150+ algorithms in its security-led AI&ML platform. This information, along with the post-process, enriched metadata has all the necessary attributes and information necessary to properly train security-focused large language models to further assist in DCO efforts. Acting upon this highly effective output, we optimize accessing, cleansing, and interpreting data to ensure a continuous pipeline from diverse global and hybrid sources. Our strategy includes integrating various data streams such as Open Source Intelligence (OSINT), public datasets like Common Crawl and CICIDS with internally generated cybersecurity reports and logs into AI analytics platforms.

Clarity of signal is critical when considering AI model training approaches Dell Technologies has partnered with Vectra AI to provide a real-time network detection and response platform powered by true security-led AI comprised of supervised and unsupervised learning methodologies. Vectra AI's coverage is built to identify the behaviors associated with these consistent attacker activities, which enables confident identification of attackers without noise. Vectra aligns these detections to the associated entities observed on the network and their interactions, whether these are hosts, user identities or edge/IOT devices in real-time. This allows defensive cyber operators, analysts, and incident responders more upfront context regarding the detections and any of the participating components without adding voluminous amounts of purely anomalous indicators to investigate and consume time and attention. This type of signal being utilized at-scale within DoD and IC agencies is key to ensuring that models being trained to assist and augment DCO activities are automatically being provided the proper attributes and context via initial, real-time AI&ML processing which far outpaces human capacity By adopting the combined capabilities of Vectra's cybersecurity platform and Dell Technologies' infrastructure, we tackle the labeling challenges using supervised (e.g., classifying), unsupervised (e.g., clustering), and semi-supervised learning techniques (e.g., bootstrapping, co-training). This approach improves scalability and adaptability but also enhances cyber-threat detection, thereby protecting the integrity of the DoDIN and strengthening national security.