Důvěřuj AI prověřuj!

MIROSLAV NEČAS

Podíl automatizace narůstá ve všech odvětvích a stejně tak je tomu i v oblasti bezpečnosti. Systémy využívající umělou inteligenci (AI) se v segmentu bezpečnosti objevily již před několika lety. Nyní již téměř každý detekční systém uvádí, že využívá AI, typicky jde o strojové učení. S příchodem umělé inteligence by se mohlo zdát, že lidé již brzy nebudou potřeba. Je tomu opravdu tak?

Díky rozšířenosti AI se začínají objevovat útoky, které využívají zranitelností umělé inteligence. Tyto zranitelnosti vyplývají ze způsobu, jakým se AI učí dělat rozhodnutí na základě vstupních dat.

Zjednodušeně řečeno, vstupní data promítá do mnohorozměrného prostoru, a následně se snaží určit, co je „správné“ a co „špatné“, na základě redukce tohoto prostoru.

AI tedy ve skutečnosti nerozumí podstatě problému, ale rozhoduje se z hlediska kvantifikace jeho vybraných vnějších projevů.

Útoky na umělou inteligenci spočívají v odhalení a předstírání těchto symptomů, nebo v ovlivnění dat, na kterých se AI učí.

Miroslav Nečas,
Bussines Development Manager
TOVEK

AI potřebuje kompetentní dohled

Při využívání AI je důležité vědět, na základě čeho se rozhoduje, a průběžně kontrolovat, zda se rozhoduje správně. Nejde jen o to, čelit případným útokům. Když se povaha vstupních dat významně změní, AI přestane dělat správná rozhodnutí. Je to stejné, jako nechat čtyřleté dítě řídit auto v plném provozu. Většina komerčních AI řešení funguje jako black box i pro své výrobce, a to nás staví do nelehké pozice.

Podporujme lidský vhled, rozvahu a kreativitu

Systém Tovek Cyber ARMS umožňuje člověku sledovat dění v IT infrastruktuře v kontextu celé organizace i jejího okolí. Využívá pokročilé techniky analýzy a vizualizace velkých dat, tak aby je člověk byl schopný vnímat a vyhodnocovat. Zdrojem těchto dat jsou například SIEM systémy, databáze zranitelností, IoT senzory, SCADA, konfigurační databáze, systémy docházky a ERP systémy, webové zdroje či databáze ekonomických subjektů. Díky vhledu do dění v organizaci i mimo ni je člověk schopný kvalifikovaně posoudit, zda systémy fungují správně, identifikovat nesrovnalosti, jež prošly filtrem AI, a zároveň posoudit, které poplachy jsou relevantní a které nikoliv.