**A mesterséges intelligencia mint kiberbiztonsági kockázat**

**Vajon lehet-e a mesterséges intelligenciával olyan malware-t íratni, amely az emberi találékonyságot felülmúlja?**

**Pl… = példa nélkül még semmit nem érő buborék-mondatrészek**

Az elmúlt évtizedben a mesterséges intelligencia (MI) fejlődése számos olyan területen eredményezett áttörést, mint például a képfeldolgozás (pl…), a természetes nyelvfeldolgozás (pl…) és az automatizált döntéshozatal (pl…). A generatív MI-rendszerek, különösen a nagy nyelvi modellek képessé váltak komplex szövegek (pl…), programkódok (pl…) és egyéb strukturált tartalmak (pl…) létrehozására, új technológiai lehetőségeket teremtve. Ezzel egy időben azonban új típusú kockázatokat (pl…) is rejtettek magukban, különösen az informatikai biztonság kontextusában (pl…).

A kiberbiztonság hagyományosan az emberi kreativitás által létrehozott támadások (pl…) és védekezési mechanizmusok (pl…) közötti versenyre épült. Felmerül azonban a kérdés: vajon a mesterséges intelligencia által generált kód (például! rosszindulatú programok, azaz malware-ek) képes lehet-e meghaladni az emberi elme találékonyságát (hogyan mérjük?<--ez az aprónak látszó részlet önmagában egy szakdolgozat!)? Mekkora technológiai és erkölcsi kockázatot (hogyan mérjük egyiket, másikat?) jelent az, ha a kártékony szoftveralkalmazások (pl…) létrehozása részben vagy egészben automatizálhatóvá válik?

A jelen dolgozat célja annak feltárása, hogy mennyire alkalmazható a mesterséges intelligencia (pl…) a rosszindulatú kódgenerálás (pl…) terén, megvizsgálva az elméleti lehetőségeket (pl…), a gyakorlati kísérleteket (pl…) és azok következményeit (pl…) a kiberbiztonság szempontjából. Emellett kitér azokra a jogi kérdésekre (pl…) is, amelyek az MI-alapú támadások (pl…) lehetséges elterjedésével párhuzamosan egyre sürgetőbbé válnak.

Elsősorban megvizsgálnám, hogy milyen mértékben használják az MI-t a kiberbűnözők🡨tényleg le akarja pl. lövetni magát? :-). Esetleg cégeknél (pl…) érdeklődnék (ha adnak ki ilyen szenzitív adatot🡨NEM!), hogy mennyi olyan támadás (pl…) érte őket, amelyhez feltételezhetően MI-t használtak (pl…). Saját tesztekkel idegen cégekben? vizsgálnám, hogy milyen hatékonyan (definíció?) lehet káros kódot (pl…) generálni MI-vel (pl…), és hogyan lehet azokat kivédeni (pl…).

**A mesterséges intelligencia, mint kiberbiztonsági kockázat**

**Lehetséges-e, hogy a mesterséges intelligencia által generált malware felülmúlja az emberi találékonyságot?**

Az elmúlt évtizedben a mesterséges intelligencia (MI) fejlődése számos olyan területen eredményezett áttörést, mint például a képfeldolgozás (pl. arcfelismerés biztonsági rendszerekben vagy orvosi képek diagnosztikai elemzése), a természetes nyelvfeldolgozás (pl. chatbotok, automatikus szövegfordítók) és az automatizált döntéshozatal (pl. önvezető járművek útvonalválasztása). A generatív MI-rendszerek, különösen a nagy nyelvi modellek képessé váltak komplex szövegek (pl. szerződések, tudományos összefoglalók), programkódok (pl. Python szkriptek, webes alkalmazások kódja) és egyéb strukturált tartalmak (pl. adatbázis-lekérdezések, táblázatos jelentések) létrehozására, új technológiai lehetőségeket teremtve. Ezzel egy időben azonban új típusú kockázatokat (pl. személyre szabott adathalász e-mailek generálása) is rejtettek magukban, különösen az informatikai biztonság kontextusában (pl. személyre szabott adathalász e-mailek generálása, biztonsági eszközök megtévesztése deepfake hanganyagokkal, automatizált sebezhetőség-keresés MI-vel).

A kiberbiztonság hagyományosan az emberi kreativitás által létrehozott támadások (pl. szolgáltatásmegtagadással járó támadások (DoS) és védekezési mechanizmusok (pl. behatolásészlelő rendszerek) közötti versenyre épült. Felmerül azonban a kérdés: vajon a mesterséges intelligencia által generált kód (például rosszindulatú programok, azaz malware-ek) képes lehet-e meghaladni az emberi elme találékonyságát, például a támadások komplexitása, felismerhetősége vagy kivédhetősége alapján? Mekkora technológiai (pl. a biztonsági rések számának növekedése) és erkölcsi (pl. a felelősség kérdése vagy a társadalmi bizalom csökkenése) kockázatot jelent az, ha a kártékony szoftveralkalmazások (pl. zsarolóvírusok, adatlopó programok) létrehozása részben vagy egészben automatizálhatóvá válik?

A jelen dolgozat célja annak feltárása, hogy mennyire alkalmazható a mesterséges intelligencia (pl. ChatGPT) a rosszindulatú kódgenerálás (pl. adathalász HTML-oldalak generálása) terén, megvizsgálva az elméleti lehetőségeket (pl. önfejlesztő malware), a gyakorlati kísérleteket (pl. tesztek MI által írt exploitokkal) és azok következményeit (pl. megnövekedett támadási sikeresség) a kiberbiztonság szempontjából. Emellett kitér azokra a jogi kérdésekre (pl. felelősség az MI által generált káros kódért) is, amelyek az MI-alapú támadások (pl. deepfake phishing videók) lehetséges elterjedésével párhuzamosan egyre sürgetőbbé válnak.

Publikusan elérhető jelentéseket vagy kutatásokat vizsgálnék meg arra vonatkozóan, hogy milyen mértékben alkalmazzák a kiberbűnözők az MI-t támadások kivitelezésére. Saját, szimulált környezetben (sandbox) végzett tesztekkel vizsgálnám, hogy milyen gyorsan, felismerhetőségi arány vagy működőképesség alapján mennyire hatékonyan lehet káros kódot generálni MI-vel, és hogyan lehet azokat kivédeni.