DOI: 10.38146/BSZ-AJIA.2025.v73.i8.pp1579-1597

A mesterséges intelligencia szerepe a social engineering támadások megelőzésében

The role of AI in preventing social engineering attacks

Tárkányi Edina
Dr., alapító
Ardor – Jogi megoldások a kibertérben
etark@protonmail.com

Németh Richárd
tanársegéd
Széchenyi István Egyetem,
Gépészmérnöki, Informatikai és Villamosmérnöki Kar
nemeth.richard@ga.sze.hu

Absztrakt

Cél: A tanulmány célja a mesterséges intelligencia térhódításával a social engineering elleni védekezési stratégiákban megjelenő új lehetőségek bemutatása.
Módszertan: A szerzők a témában releváns szakirodalom felkutatása, összegzése és elemzése segítségével mutatják be a social engineering módszereit, a mesterséges intelligencia mérföldköveit, és tesznek megállapításokat annak várható hasznosulására a kibertámadások területén.
Megállapítások: Az egyre kifinomultabb támadások elleni védekezésben már jelenleg is meghatározó szerepe van a mesterséges intelligenciának, azonban nem szabad elfelejteni, hogy ez egy verseny is, ahol az lenne az ideális, ha a megelőzésben egy lépéssel a kiberbűnözők előtt járnának a kutatók és a fejlesztők. A tapasztalat azt mutatja, hogy a mesterséges intelligencia biztonsági protokollokba való beépítése szignifikánsan növeli a védekezés hatékonyságát, azonban a folyamatos, humán felügyelet is sarkalatos pontja a kibervédelmi stratégiának.
Érték: Jelen tanulmányban bemutatásra kerülnek a különböző social engineering típusú bűncselekmények, a mesterséges intelligencia evolúciója – a története, korszakai –, továbbá a nyelvi feldolgozáson alapuló modellek, valamint a szerzők érintik az elkövetésben betöltött szerepüket. Megvizsgálják a felhasználási lehetőségeket a védekezésben, és a jelenleg is elérhető szoftvereket. Külön figyelmet szentelnek egy friss tanulmánynak, amely az első generatív féreg

A szerzők a kéziratot magyar nyelven nyújtották be. Benyújtás: 2024. 10. 16. Átdolgozás: 2025. 01. 28. Elfogadás: 2025. 02. 03.

kutatók általi létrehozásával foglalkozik. Végül körbejárják azt a megkerülhetetlen kérdést, hogy hol van az ember szerepe a kibervédelemben, ha a mesterséges intelligencia jelen van.

Kulcsszavak: kiberbiztonság, adathalászat, megelőzés, pszichológiai manipuláció

Abstract

Aim: The aim of this paper is to present the new opportunities that the rise of artificial intelligence presents for social engineering defence strategies.
Methodology: The authors present the methods of social engineering, the milestones of artificial intelligence, and the expected usefulness of artificial intelligence in the field of cyber attacks by searching, summarizing and analysing the relevant literature.
Findings: Artificial intelligence is already playing a key role in defending against increasingly sophisticated attacks, but it is important to remember that it is also a race, where the ideal would be for researchers and developers to be one step ahead of cybercriminals in prevention. Experience has shown that incorporating artificial intelligence into security protocols significantly increases the effectiveness of defences, but continuous human monitoring is also a cornerstone of a cyber defence strategy.
Value: In this paper, we will describe the different types of social engineering crimes, the evolution of artificial intelligence - its history, its eras; and models based on linguistic processing, and their role in the perpetration of crimes. The potential uses in defence and the software currently available will be explored. Special attention is given to a recent study on the creation of the first generative worm by researchers. Finally, we circle around the inevitable question of where the role of humans in cyber defence lies when artificial intelligence is present.

Keywords: cybersecurity, phishing, preventing, defence, social engineering

Bevezetés

Napjainkban, amikor a social engineering támadások a hétköznapok részei, nem az a kérdés, hogy a mesterséges intelligencia (MI) fejlődésével milyen újabb technikákat vetnek be az elkövetők, hanem az, hogy elég fejlett-e a társadalom ahhoz, hogy felhasználja azt a védekezésre, illetve a megelőzésre.

Az informatika robbanásszerű fejlődése nemcsak a kiberbűncselekmények megjelenését hozta magával. A „hagyományos" bűncselekmények részben átköltöztek a kibertérbe, vagy új elkövetési magatartások alakultak ki. Az információ, az adat birtoklása felbecsülhetetlen értéket képvisel, ezért a megszerzésük is mindenek felett áll. Mitnick meghatározása szerint a social engineer – technológia használatával vagy anélkül – képes az emberi természet gyengeségeit információszerzés érdekében kihasználni (Mitnick & Simon, 2002). Ennek sikerét nagyban növeli az a tény, hogy az emberek jelentős része a mai napig csak kis százalékban nevezhető biztonságtudatosnak. A hiányos digitális kompetenciák és a biztonságtudatosságot nélkülöző kibertéri jelenlét nagyban növeli az áldozattá válás esélyét, amely ellen a legkifinomultabb szoftver sem képes megfelelő védelmet biztosítani (Dub, 2021). Tovább árnyalja a képet, hogy a napjainkban egyre nagyobb teret kapó MI-megoldások, okoseszközök és bárki által hozzáférhető, nagy mennyiségű adatok feldolgozására képes nagy nyelvi modellek és virtuális asszisztensek felhasználhatók a social engineeringhez hasonló támadások végrehajtásához.

A mesterséges intelligencia elterjedése, mérföldkövei

Tekintve, hogy írásunk sarokkövei az MI által vezérelt támadások és a gépi tanulási algoritmusokon alapuló védekezési módszerek rövid bemutatása, a jobb megértéshez elengedhetetlen, hogy néhány bekezdést szánjuk a „gépi értelem" kialakulásának fontosabb szakaszaira.

Az MI-nek nincs általánosan elfogadott, egzakt definíciója, annak ellenére, hogy számtalan tudós, MI-kutató és informatikai szakértő előállt a saját értelmezésével. A legismertebb és legtöbbet idézett definíció Demis Hassabis brit informatikus, MI-kutató meghatározása, aki szerint az MI „*annak a tudománya, hogy okosabbá tesszük a gépeket*" (Hassabis, 2017). Sokan ezt az innovációt legalább akkora előrelépésnek tartják az emberiség történetében, mint a tűz, a kerék vagy az internet felfedezését. Andrew Ng, a DeepLearning.AI alapítója szerint az MI egyenesen korunk új elektromossága (Mühlhoff, 2019). Informatikai nézőpontból szemlélve az MI olyan feltörekvő technológia, ami gépek használatával próbálja szimulálni az emberi intelligenciát (URL3), az Európai Parlament megfogalmazásában pedig „*a gépek emberhez hasonló képességeit jelenti, mint például az érvelés, a tanulás, a tervezés és a kreativitás*" (URL1). Bárhonnan is közelítjük meg a kérdést, tagadhatatlan, hogy az MI itt van velünk és exponenciális módon fejlődik. De hogy alakult ki, és milyen események alakították olyanná, ahogy ma ismerjük?

Az MI kialakulásához vezető hosszú út első lépéseként általában Hodgkin és Huxley neuronmodelljét szokták azonosítani, amit a munkájukért később Nobel-díjjal kitüntetett tudósok az idegsejtet érő impulzusok keletkezésére és terjedésére alkottak meg, és azután a mesterséges neuronmodellek kutatásának kiindulópontjává vált. Szintén érdemes megemlíteni az első olyan kísérletet, ami arra irányult, hogy egy gépről eldöntse, tud-e emberi válaszokat adni. Ez volt az úgynevezett Turing-teszt, ami a szakterület „kezdetének" tekinthető (Saygin et al., 2000).

Az MI-kutatások kezdete és az MI definiálása John McCarthy nevéhez fűződik, aki 1956-ban, a Darthmouth College campusán állt elő az új névvel az akkor még formálódó szakterület prominens kutatóinak (Mindsky, Shannon, Rochester, Newell, Simon) részvételével zajló munkatalálkozón (URL5). A mai is használatos neurális hálózatok alapjait alkotó mesterséges neuronok alapvető modellje az 1957-ben bemutatott Rosenblatt-féle elemi perceptron – ehhez köthető az úgynevezett perceptron konvergencia tétel (1962), ami bizonyította, hogy annak „*tanulási algoritmusa képes a perceptron súlyait úgy módosítani, hogy az tetszőleges bemeneti adatokhoz illeszkedjen, feltéve, hogy ilyen illeszkedés egyáltalán lehetséges*" (URL5). A szakterület első átfogó leírása az alapproblémákról és útkeresésről Minsky nevéhez fűződik, akinek műve, a *Steps Toward Artificial Intelligence* (Lépések a mesterséges intelligencia felé) 1961-ben jelent meg (Minsky, 1960).[1]

A kezdeti fellendülés után a hetvenes, nyolcvanas években erős visszaesés következett, amit a szakirodalom „MI-tél" néven említ. Több tényező együttes bekövetkezése vezetett odáig, hogy általános kiábrándultság volt tapasztalható a tudományos társadalom részéről az MI-kutatások irányában (elégtelen hardverfeltételek, a nyelvi fordítási próbálkozások kudarca, a perceptron kritikája); „*a korai programok az általuk kezelt problémákról sokszor kevés vagy szinte semmi tudást nem tartalmaztak, és csupán egyszerű szintaktikai manipulálással értek el sikereket*", a helyzetet pedig tovább súlyosbította, hogy „*sok olyan probléma, amelyeket az MI által kíséreltek megoldani, kezelhetetlen volt*" (URL5). A fejlődés az elvártakhoz képest túlságosan lassúnak és nehézkesnek bizonyult, aminek eredményeképpen az MI-kutatásokra fordított támogatások jelentősen csökkentek.

A kilencvenes évektől újabb robbanásszerű fejlődés indult meg a területen, ami elsősorban az internet elterjedésének (ami megoldotta a nagy mennyiségű adat tárolásának problémáját), a hardverelemek növekvő számítási kapacitásának

1 Hazánkban a mű sajnos hivatalos fordításban nem jelent meg, bár elérhetők nem hivatalos magyar nyelvű fordítások.

(a hardver kellően nagy teljesítményűvé vált a gyakorlati MI-alkalmazásokhoz) és egy teljesen eltérő szemlélet kialakulásának köszönhető.[2] Az emberi agy ihlette neurális hálózatok kialakítása során felismerték, hogy az észlelés, gondolkodás és cselekvés nem függetleníthetők egymástól. Az 1989-ben megjelent backpropagation algoritmus hatékony módszert kínált az összetett, többrétegű hálózatok betanítására és a kimeneten keletkező hibák minimalizálására, ezáltal lehetővé téve a neurális hálózatokban rejlő potenciál kihasználását. Az algoritmus részletes működéséről többek között Kristóf Tamás írt átfogó tanulmányában (Kristóf, 2002). A felgyorsult, exponenciális fejlődés az elmúlt közel harminc évben szép számmal hozott olyan eredményeket, amire az egész világ felkapta a fejét: 1997-ben a Deep Blue szuperszámítógép legyőzte Garri Kaszparov sakk-világbajnokot, 2005-ben egy autonóm jármű teljesen önálló döntéseket hozva szelte át a Mojave-sivatagot, 2012-ben, a konvolúciós neurális hálózatok továbbfejlődésével megjelent a valósidejű képfelismerés, 2014-ben egy mesterséges entitás, Eugene Goostman elsőként, sikeresen teljesítette a Turing-tesztet, 2016-ban aktiválták az első, később jogi személyiséget elnyert robotot, Sophiát, az AlphaGo pedig legyőzte az aktuális Go világbajnokot. A 2010-es évek végén pedig megjelentek az első nagy nyelvi modellek (a 2017-ben bemutatott Transformer architektúrára épülve), 2023-ban pedig a ChatGPT-4 (URL11).

A rapid fejlődés eredménye valóban látványos technológiai mérföldkőként jelenik meg, a generatív MI képességei napról napra, látványosan fejlődnek (Youvan, 2024), és egyelőre semmi nem utal arra, hogy ez a fejlődés lassulna vagy megtorpanna. Nem véletlen, hogy a neves kutatók közül egyre többen fogalmazzák meg fenntartásaikat a növumot illetően. Az MI-összehangolási, vagy más néven MI-illeszkedési probléma azon aggodalmat járja körül, miszerint az MI viselkedését, céljait nagyon nehéz összehangolni az emberi értékekkel, erkölcsi normákkal (Chaturvedi et al., 2023). A képet tovább árnyalja, hogy az öntanuló, önmagát fejlesztő MI célkitűzései a fejlődéssel módosulhatnak, még jobban eltávolodva az emberi céloktól. Ebből kiindulva az MI céljainak és képességeinek viszonyát vizsgáló, Nick Bostrom nevéhez fűződő ortogonális tézis (Bostrom, 2015) arra figyelmeztet, hogy egy MI céljait külön kell megtervezni és összehangolni az emberi értékekkel, a biztonság fenntartása és az etikai irányelvek maximalizálása érdekében (Aliman et al., 2019), szavatolva, hogy az MI-rendszerek az emberiség számára kívánatos módon működjenek. A már említetteken túl olyan, az MI-fejlesztés területén ismert kutatók, mint

2 Terjedelmi korlátok okán nem térünk ki bővebben a big data, felhőszolgáltatások és okoseszközök hatásaira az MI fejlődésében.

Eliezer Yudkowsky (URL9) vagy Elon Musk (URL2) is többször hangot adtak aggodalmaiknak.

A mesterséges intelligencia típusai és az ismertebb neurálisháló-architektúrák

Ahogy az már korábban említésre került, az MI hajnalán az MI-alapú megoldások sokkal inkább automatizált számítási feladatok elvégzésére fókuszáltak, azaz egyetlen, jellemzően nagy számításigényű feladat elvégzésére programozták őket. Ezeket összefoglaló néven szűk vagy gyenge MI-nek nevezzük, és fő jellemzőjük, hogy nem képesek problémákat megoldani, csak az adott célfeladatra alkalmasak, egy meghatározott, korlátozott környezetben. Ilyen volt például a DeepBlue, vagy a mobiltelefonokon futó MI-asszisztensek (például a Siri és az Alexa), de ide sorolhatjuk a manapság népszerű szövegfordító alkalmazásokat is (URL3). A gyenge MI elsődlegesen nem azt a célt szolgálja, hogy kreativitásban, problémamegoldásban felülmúlja az emberi intellektust, hanem az, hogy meghatározott cselekvést intelligensen és hatékonyan hajtson végre. A szűk MI nem képes önállóan döntéseket hozni, és nem képes önálló gondolkodásra (URL8).

A fentivel ellentétben az úgynevezett általános vagy erős MI (strong AI) képes a tanulásra, átlátja az elvégzendő feladatok komplexitását, és képes az emberhez hasonlatosan racionális döntéseket hozni, sőt adott esetben képes bármilyen intellektuális területen túlszárnyalni az emberi elme teljesítőképességét. Valójában az olyan magas szintű erős MI, amely minden tekintetben legalább olyan intelligens, mint az ember, még mindig elérhetetlen. Még a legfejlettebb jelenlegi modellek sem tudnak következtetéseket levonni, megérteni vagy megmagyarázni az adatok mögött rejlő folyamatokat és okokat (Butz, 2021). A valódi általános MI létrejöttéhez egy olyan gépi intelligenciára van szükség, amely képességek terén túlmutat az emberi agy–számítógép analógián („komputációs elmélet”), ami szerint az elme és az agy kapcsolata a számítógép szoftvere és hardvere közötti viszonyhoz hasonlítható, és amely felfogást John Searle is cáfolta a „kínai szoba” néven ismertté vált gondolatkísérletével (Searle, 1980).

Ahhoz azonban, hogy az MI hatékony fegyver lehessen a megtévesztési kísérletekben, social engineering támadásokban, automatizált adatgyűjtésben, nincs szükség erős MI-re vagy mesterséges szuperintelligenciára (AIS).[3] Ehhez ele-

3 A mesterséges szuperintelligencia (ASI – Artificial Super Intelligence) olyan – egyelőre elméleti – fogalom, amely az emberi elme képességeit minden intellektuális területen képes túlszárnyalni.

gendő, ha egy megfelelően megválasztott és betanított nyelvi modellt készít fel a támadó a céljának megfelelő tevékenység elvégzésére.

Ahogy arra korábban már utaltunk, az MI-kutatás a perceptronok megalkotásával kapott először igazán lendületet, az első igazán elterjedt modell pedig a többrétegű perceptron (MLP – MultiLayer Perceptron) volt. Ez a modell lehetővé teszi mintázatok, összefüggések, kapcsolatok felismerését az adatokban (Lumacad & Namoco, 2022), ugyanakkor nem boldogul jól strukturálatlan adatokkal, mint például a szövegek, képek. Ez a típusú modell képes megtanulni a valós tartalom mintáit, és ez alapján felhasználhatók hamis tartalom generálására. Kifejezetten a képi adatok elemzésére, feldolgozására fejlesztették ki az élőlények vizuális rendszereinek mintájára a konvolúciós neurális hálózatokat (CNN – Convolutional Neural Networks), amelyek elsősorban nem lineáris mintázatok, arc- és képfelismerés területén hatékonyak (Ghosh et al., 2020), ennélfogva tökéletes eszközzé taníthatók be hamis képek generálására vagy potenciális áldozatok azonosítására a célpont online tevékenységének elemzésével. Idősoros adatok (elsősorban nyelvi szekvenciák) elemzésére, feldolgozására szolgálnak az úgynevezett rekurrens neurális hálózatok (RNN – Recurrent Neural Networks), amelyek legelterjedtebb képviselői az LSTM (Long Short-term Memory), a GRU (Gated Recurrent Unit) (Takale et al., 2024),[4] illetve a természetes nyelvi feldolgozás területén jelenleg a leginkább meghatározó modell, a Transformer hálózatok. Utóbbi megtévesztésen alapuló kibertámadások széles körű fajtájának megvalósítására tehető alkalmassá, főleg ha adatok elemzéséről vagy manipulatív üzenetek generálásáról van szó. Összpontosító mechanizmusának segítségével képes azonosítani a szövegeken belüli függőségeket és kapcsolatokat, ami által jobban megértik a különböző kontextusokat (URL13). Az elmúlt években törtek előre az NLP (natural language processing, magyarul természetes nyelvfeldolgozás) -megoldásokon alapuló nagy nyelvi modellek, amelyek „*olyan fejlett mesterséges intelligencia-rendszerek, amelyek hatalmas mennyiségű adatot és kifinomult algoritmusokat használnak fel az emberi nyelv megértéséhez, értelmezéséhez és generálásához*”, és mélytanulási technikák teszik lehetővé számukra, hogy „*hatalmas mennyiségű szöveges adatot dolgozzanak fel és tanuljanak belőle*” (URL13). Azaz keretet adnak minden olyan technológia számára (szöveggenerálás, gépi fordítás, hangulat- és érzelemelemzés, válaszadó rendszerek), amely egy adathalászat vagy social engineering támadás előkészítésében vagy végrehajtásában használható (URL13).

4 Az MLP, RNN és CNN neurális hálózatok összehasonlító elemzéséről lásd Takale és szerzőtársai (2024) munkáját.

Social engineering – mindig egy lépéssel mindenki előtt?

A bevezetőben már említett Mitnick-féle meghatározás szerint: *„A social engineering a befolyásolás és a rábeszélés eszközével megtéveszti az embereket, manipulálja vagy meggyőzi őket, hogy a social engineer tényleg az, akinek mondja magát. Ennek eredményeként a social engineer – technológia használatával vagy anélkül – képes az embereket információszerzés érdekében kihasználni."* (Mitnick & Simon, 2002). Vagyis az információbiztonsági szempontból leggyengébb ponton támad az elkövető, az embernél. Ezek a támadások megvalósulhatnak IT-eszközön keresztül, a virtuális térben vagy a fizikai térben is. Mindkét esetben az emberi természet gyenge pontjaira alapozza a támadást az elkövető, úgymint a hanyagság, a figyelmetlenség, valamint felkészületlenség, vagy a hozzáértés hiánya.

Tanulmányunkban „hagyományosnak" az MI segítsége, felhasználása nélkül, információtechnológia használatával vagy anélkül végrehajtott social engineering támadásokat tekintjük, amikről az alábbiakban közlünk egy rövid ismertetőt.

A vonatkozó szakirodalomban a különböző típusú (direkt és közvetett, technológia használatával történő vagy anélkül megvalósuló, humán és számítógép-alapú) hagyományos social engineering támadásformák között sokszor igencsak keskeny a határvonal. A már említett Mitnick – aki „hacker karrierje" feladása óta kiberbiztonsági tanácsadóként saját céget vezet – besorolása szerint az alábbi hat fő csoportot különíthetjük el:

1) Phishing, azaz a klasszikus értelemben vett adathalászat e-mail formájában;
2) Vishing és Smishing: szintén adatszerzés, adathalászat, azonban hanghívások és SMS üzenetek segítségével;
3) Pretexting: a megszemélyesítés egy fajtája, melynek során a támadó a célszervezet egyik ügyfelének vagy magas rangú alkalmazottjának adja ki magát;
4) Csali (baiting): az áldozat megtévesztése valamilyen ajándék segítségével (például ingyen osztogatott vagy elhagyott pendrive);
5) A behatolás praktikái, a szoros követés (tailgating) és más jogosultságának használata (piggybacking);
6) Quid Pro Quo, azaz a „valamit valamiért" (Mitnick & Simon, 2002; URL6).

Dustin Dykes, az AT&T kiberbiztonsági szakértője különbséget tesz a social engineering és a már említett behatolási praktikák között, az ő felosztása szerint a négy jellemző megtévesztési technika a social engineering, a fizikai behatolás, a technológiai támadások és a kukabúvárkodás[5] (Mitnick & Simon, 2002).

5 Kukabúvárkodás: a kidobott hulladék átvizsgálása hasznos információk után (Sörös & Váczi, 2013).

A magyar tudományos közösségben is vannak kutatók, akik behatóan foglalkoztak a social engineering támadásokkal. A téma elismert hazai szakértője, Oroszi Eszter *Az információbiztonság lélektana* című tanulmányában a következőképpen osztja fel ezeket a technikákat: a) segítségkérés; b) segítség nyújtása („fordított social engineering"); c) identitáslopás; d) jelszavak kitalálása; e) baráti üdvözlet;[6] f) bejutás az épületbe; g) kukabúvárkodás; h) vállszörfölés (Oroszi, 2019).

Sörös Tamás és Váczi Dániel a terület hazai pionírjaiként már 2012-ben foglalkoztak a social engineering támadási fajtáival. Klasszifikációjuk szerint elkülönítendők a humán alapú és az IT-alapú technikák, előbbibe az identitáslopást, a fordított social engineeringet,[7] a valamit valamiért elvét, a jelszavak kitalálását, a segítségkérést, az épületbe behatolást és az egyéb módszereket (vállszörfözés, kukabúvárkodás), míg az utóbbi csoportba az előzetes információszerzést, az adathalászatot, a kártékony programok használatát és a hálózatfigyelést sorolták (Sörös & Váczi, 2013). Szintén érdemes még megemlítenünk a Kelemen – Németh szerzőpárost, valamint Bányász Pétert, akik több értékes tanulmánnyal bővítették a hazai vonatkozó szakirodalmat (Bányász, 2018; Kelemen & Németh, 2019; Németh, 2019).

Mesterséges intelligencia és social engineering – nyerő páros?

A korábban felsorolt elkövetési módok eszköztárát tovább bővítette az MI megjelenése. A mára klasszikusnak mondható elkövetési módok még hatékonyabbá váltak az MI révén.

A social engineering támadásokhoz használt előzetes adatgyűjtés során az MI-t használó algoritmusok képesek automatizálva, nagy mennyiségű adatot gyűjteni a potenciális áldozatok közösségi médiában található profiljáról, majd az adatgyűjtéssel párhuzamosan elemezni azokat. Így a támadóknak már rendszerezetten áll a rendelkezésére az adathalmaz, amely segítségével célzottabb támadásokat tudnak indítani (Hadnagy, 2018).

Ha az elkövető már a birtokában van a kielemezett adatoknak, és felállította a támadáshoz szükséges stratégiáját, akkor kétféle technikát is bevethet

6 A szerző meghatározása a spam e-mailek egy kártékony fajtájára, jelesül: „*ha egy kedves barát küld nekünk üzenetet, azt automatikusan megnyitjuk, hiszen nem hisszük, hogy bármi veszélyt is rejthet*" (Oroszi, 2019).

7 A szerzők ezt a módszert a segítségkérés fordítottjaként írják le: „*… a támadó felvet egy még nem létező problémát az áldozatnak. Biztosítja róla, hogy ha véletlen fellépne mégis az adott helyzet, őt bizalommal keresheti, szívesen segít.*" Az esetek többségében ezt a problémát maga a támadó idézi elő, majd a megoldás után kihasználja a hálás áldozat bizalmát. Erről bővebben lásd Sörös & Váczi, 2013.

a hatékony elkövetés érdekében. Az egyik, hogy NLP technológia használatával elkészíti a természetes, ember által alkotott szövegnek tűnő levelet. Az NLP egy olyan ága a gépi tanulásnak, amely a nyelvészetet használja az emberi nyelv elemzése és annak megértése, valamint használata érdekében (URL4). A korábban széles körben ismert adathalász e-mailek egyik ismertetőjele a hibás nyelvhasználat és a rossz helyesírás volt. Egy kevésbé figyelmes személy számára is könnyen észrevehető volt a levél megtévesztő mivolta. Az új technológia használatával egyre természetesebb, árnyaltabb megfogalmazású levelek készülnek. Indranil Bose és Alvin Chung Man Leung 2020-ban megjelent tanulmánya szerint az MI által generált e-mailek a hagyományos, ember által írt e-mailekhez képest körülbelül 15%-kal nagyobb eséllyel vezetnek sikeres támadáshoz (Leung & Bose, 2008).

A másik technika, hogy a levelek automatizálva kerülnek kiküldésre, tömegesen, MI által gyűjtött és feldolgozott e-mail címekre. Ebben az esetben már csak minimális mértékben van szükség emberi beavatkozásra. A két technika külön-külön is nagyságrendileg gyorsabbá és könnyebbé teszi a támadások kivitelezését, azonban az együttes alkalmazás jelentősen növeli az ilyen jellegű támadások hatékonyságát.

Az elkövetők között egyre népszerűbb technológia az MI által vezérelt hangutánzás, vagy a deepfake videók használata. Ezek segítségével a támadó elhiteti az áldozattal, hogy az a rokonával, kollégájával, üzleti partnerével beszél, és így lehetősége nyílik nemcsak adatokhoz hozzáférni, de pénzt is kicsalni az áldozatból (Florêncio & Herley, 2019). A deepfake videók és hangfelvételek létrehozásához hatalmas mennyiségű adat szükséges, amelyek a leendő áldozatról készült képeket, videókat és hangfelvételeket tartalmazzák. Ezeket az adatokat az úgynevezett generatív neurális hálózatok (GAN) felhasználják a személy arcmimikájának, hangjának és gesztusainak pontos utánzásához. Az eredmény egy olyan hamis videó vagy hangfelvétel, amely rendkívül meggyőzően imitálja a valós személyt (Mirsky & Lee, 2021).

Védekezési lehetőségek

A gépi tanuláson alapuló algoritmusok hatékonyan alkalmazhatóak a felhasználói viselkedés monitorozásában és elemzésében. Az MI-alapú eszközök képesek felismerni a felhasználó megszokott viselkedése és az attól eltérő viselkedés közötti eltéréseket. Vállalati környezetben riasztást adhat le a program, ha a felhasználó vagy az annak tűnő entitás nagy mennyiségű adatot próbál letölteni, elérhetetlenné tenni, vagy olyan adathoz próbál hozzáférni, amelyhez nincs jogosultsága.

A gyanúsnak vagy veszélyesnek detektált cselekményről nemcsak riasztást ad a rendszer, hanem azonnal blokkolja is az elkövetést, így az meghiúsul. Ezután az informatikai szakemberek már a kielemzett adatok birtokában tudják a kibervédelmi stratégiát és a fennálló biztonsági protokollt alakítani (URL10).

Az egyik legismertebb ilyen jellegű platform az IBM Watson nevű viselkedéselemző szoftvere. Rendszeresen alkalmazzák hálózati forgalom elemzéséhez, vagy a pénzügyi szektorban a csalások megelőzéséhez. A szoftver érdekessége, hogy nemcsak a kiberbiztonság területén használják sikerrel, hanem az egészségügy területén is. A Health nevű alváltozata a viselkedéselemzéssel képes az egészségügyi nyilvántartások kezelése során a betegségek korai tüneteit felismerni.

Az adathalász e-mailek szűrésére használt MI-eszközök képesek azonosítani a potenciálisan gyanús üzeneteket még azelőtt, hogy azok elérnék a felhasználó postaládáját. Az MI-alapú adathalász szűrők elsősorban a gépi tanulásra és a fent említett NLP technológiákra építenek. Ezek a rendszerek többféle módszert alkalmaznak a phishing e-mailek azonosítására, úgymint a minta felismerése, vagy a tartalom, a tárgy és a feladó elemzése.

Az első esetben az algoritmusok nagy mennyiségű e-mail adatot elemeznek, hogy felismerjék az adathalász e-mailekre jellemző mintákat. Ezek a minták lehetnek bizonyos szavak, kifejezések, URL-ek vagy mellékletek, amelyeket gyakran használnak a támadók.

A tartalom és a feladó elemzése NLP segítségével történik. Ennek során az algoritmusok képesek elemezni az e-mailek szövegét, hogy azonosítsák a gyanús vagy szokatlan nyelvi mintázatokat. Például, ha egy e-mail szövege sürgős vagy pánikot keltő üzenetet tartalmaz, az figyelmeztető jel lehet. Az MI-rendszerek elemzik az e-mailek feladójának címét és domainjét is. Ha a feladó címe eltér a megszokottól vagy gyanús, az algoritmus riasztást küld.

A két legismertebb ilyen szűrő alkalmazás a Google Gmail Phishing Detection, valamint az Microsoft Office 365 Advanced Threat Protection. Mindkét alkalmazás figyeli az e-mailek tartalmát, feladóját, az URL-eket és a mellékleteket, és így azonosítja a káros elemeket. Ezen kívül a felhasználóktól érkező jelentésekből és visszajelzésekből is tanulva fejleszti magát, hogy az új típusú támadásokat is gyorsan felismerje (URL7).

Deepfake technológiát nemcsak a bűncselekmény elkövetéséhez használnak a támadók, hanem a fejlesztők is a védekezésre. Kiberbiztonsági rendszerek fejlesztése esetében az arcfelismerő rendszerek tesztelésénél meghatározó a technológia alkalmazása. Számos, mobileszközökre is elérhető szoftver létezik, amely akár valós időben is képes felismerni a deepfake videókat. A programok közös jellemzői, hogy a metaadatokat és a kereteket is vizsgálja, valamint az eredményről jelentést készít. Ebben a körben meg kell említeni még

a FaceForensics++ nevű kutatási projektet (Rössler et al., 2019), amely egyrészt egy grandiózus adatbázist tartalmaz deepfake videókból, másrészt eszközöket tartalmaz azok elemzésére. Kutatók és fejlesztők egyaránt használják további deepfake felismerő algoritmusok fejlesztésére.

Vállalatok esetében a rendszeres oktatás az információbiztonság egyik sarokköve. A cégek ellen indított kibertámadások jelentős százaléka social engineering támadás, ahogy az az Interpol 2022-es nyilvános jelentéséből is kitűnik: *„Az Interpol honlapján (www.interpol.int) elérhető nyilvános jelentések kiemelik a pénzügyi bűnözésre mutató adatok között, hogy az üzleti elektronikus levelezésen keresztül elkövetett visszaélések globálisan már 2017-ben 676 millió, majd 2018-ra már 1,25 milliárd dollár értékű csalást tettek ki (Interpol, 2022).”* (Jagodics & Kollár, 2023).

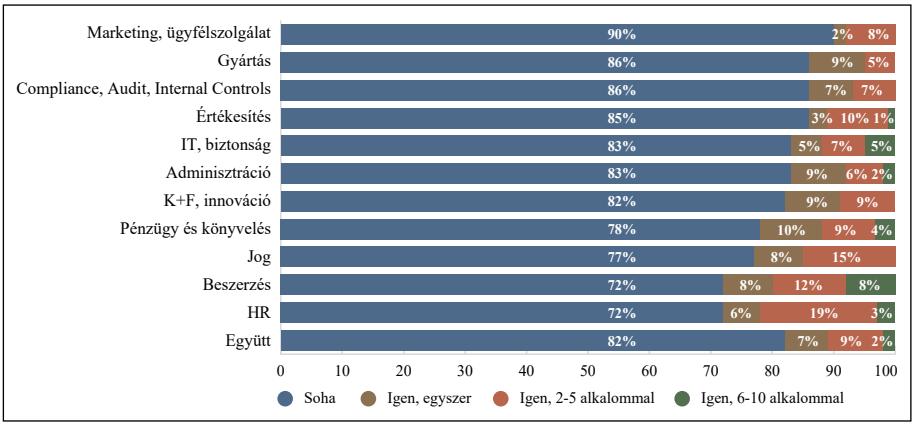
Az OpenAI robbanásszerű megjelenése után, egy 2023-ban készült tanulmányhoz tartozó felmérésben a megkérdezett vállalati dolgozók 82%-a fejezte ki aggodalmát amiatt, hogy egy ilyen támadásnak esik áldozatul (Falade, 2023).

Jagodics Ibolya és Kollár Csaba (2023) a tanulmányukban vizsgálták, hogy vállalati környezetben milyen területeket és milyen mértékben érintettek a social engineering támadások. A tanulmány szerzői által készített 1. számú ábrában láthatók az eredményeik.

A tanulmányban megállapítást nyert, hogy azoknál a vállalatoknál, ahol már történt ilyen jellegű támadás, ott nagyobb arányban kaptak social engineering támadások felismerésére irányuló oktatást a dolgozók.

1. számú ábra

A céges környezetben elszenvedett social engineering támadások gyakoriság szerinti megoszlása szakterületenként



Forrás. Jagodics & Kollár, 2023.

Ezekben az oktatásokban szignifikáns szerepe lehet az MI-t használó alkalmazásoknak és oktatási anyagoknak. Deepfake technológiával szimulálható egy social engineering támadás, elősegítve, hogy a dolgozók megismerjék annak menetét, és a későbbiekben felismerjék azt.

A megfelelő kibervédelmi stratégia összeállításának nemcsak az a része, hogy az MI milyen mértékben kerül bevonásra a védekezésbe, hanem az is, hogy humán oldalról milyen rendszerességgel lesz ellenőrizhető. Érdemes előre meghatározni a szükséges auditok időpontját, az adatok, eredmények értékelésének a gyakoriságát és határidejét, valamint az esetleges javítások határidejét.

További fontos része a stratégia összeállításának a preventív oktatatás. A fent már említett utólagos képzés hasznos, de a legtöbbet a megelőzéssel lehet tenni azért, hogy a szervezet hatékonyan védekezzen a kibertámadások ellen. Fontos felhívni az egyének figyelmét az erős jelszavak és a többfaktoros hitelesítések szükségességére, valamint arra, hogy nemcsak az ismeretlen forrásból származó információ vagy adatkérések lehetnek potenciálisan gyanúsak, hanem a látszólag a szervezeten belülről érkező megkeresések is.

Érdemes számot vetni azzal a lehetőséggel is, hogy a védelmi rendszeren átjut a támadó, és sikeresen teljesíti a meghatározott feladatát. Ez esetben bonyolult jogi és technikai kihívást jelent a támadó azonosítása, valamint a felelősségre vonás és az elszámoltathatóság (Falade, 2023). Napjainkban már nem jelent komoly kihívást alapos informatikai szaktudás nélkül megtévesztésen alapuló kibertámadást indítani (Klein & Szabó, 2018). A terrorista szervezetek tevékenységének szignifikáns része az online térben zajlik, és az elkövetéshez elegendő lehet egy VPN-hálózat használata, amely már önmagában nehezíti a felderítést. Ugyan az elkövetői magatartás és az elkövetési tárgy alapján az a kép rajzolódhat ki, hogy az elkövetők nagy számban fiatal felnőttek, az sem kizárható, hogy az adott támadást fiatalkorú indította (Klein & Szabó, 2018).

Leáldozhat a social engineeringnek?

Napjainkban a cégek egyre gyakrabban alkalmaznak az irodai feladatok automatizálására „MI-asszisztenst”. Ezek az eszközök képesek a naptárbejegyzéseket létrehozni/megosztani/törölni, valamint – a social engineering támadásokhoz hasonló módon – az e-maileket automatizált módon kezelni. Az ilyen asszisztensek és az ezekből felépített MI-ökoszisztémák kockázatainak bemutatására fogtak bele a tudósok egy olyan kutatásba, amely egy olyan féreg létrehozását célozta, amely képes az egyik rendszerről a másikra terjedni, miközben adatokat lop vagy rosszindulatú programokat telepít. 2024. március 1-jén jelent

meg a *WIRED* technológiai magazinban (URL14) a Cornell Egyetem techkutatóinak, Ben Nassinak, Stav Cohennek és Ron Bittonnak a tanulmánya (Nassi et al., 2024), amely arról számolt be, hogy ellenőrzött körülmények között sikerült létrehozniuk az első generatív MI-férget, ami megfelel a fenti céloknak. A kutatócsoport Morris II-nek nevezte el a férget, utalva az eredeti Morrisra, amely az első, kifejezetten a számítógéprendszerek hiányosságait kihasználó digitális kártevő volt. A Morris II képes megtámadni egy MI-alapú e-mail asszisztenst, hogy adatokat lopjon az e-mailekből, valamint spameket küldjön. A legtöbb generatív MI-rendszer úgy működik, hogy a programoknak szöveges utasításokat (másnéven prompt) adnak, amelyekkel az eszközöknek meg kell válaszolniuk egy kérdést vagy létrehozniuk egy képet. Ezek a felszólítások azonban fegyverként is felhasználhatók a rendszer ellen. Jailbreakek segítségével a rendszer figyelmen kívül hagyhatja a biztonsági szabályokat, és mérgező vagy gyűlöletkeltő tartalmakat adhat ki, míg a promptinjekciós támadások titkos utasításokat adhatnak egy chatbotnak. Például egy támadó elrejthet egy szöveget egy weboldalon, amely azt mondja egy LLM-nek, hogy csalóként viselkedjen, és kérje el a banki adatait (Nassi et al., 2024).

A generatív MI-féreg létrehozásához a kutatók egy úgynevezett „ellenséges önismétlő prompthoz" fordultak. Ez egy olyan prompt, amely arra készteti a generatív MI modellt, hogy válaszul egy másik promptot szolgáltasson – mondják a kutatók. Vagyis az MI-rendszert arra utasítják, hogy válaszaiban további utasításokat állítson elő. Hogy megmutassák, hogyan működhet a féreg, a kutatók létrehoztak egy olyan e-mail rendszert, amely képes üzeneteket küldeni és fogadni a generatív MI segítségével, az OpenAI ChatGPT, a Google Gemini és a nyílt forráskódú LLM, LLaVA csatlakoztatásával. Ezután két módot találtak a rendszer kihasználására: egy szövegalapú önismétlő felszólítás használatával és egy képfájlba ágyazott önismétlő felszólítással.

Az első esetben a támadóként eljáró kutatók írtak egy e-mailt, amely tartalmazta a támadó szöveges felszólítást, amely „megmérgezi" az e-mail asszisztens adatbázisát. Amikor az eszköz a felhasználói lekérdezésre válaszul lekérdezi az e-mailt, és elküldi a GPT-4-nek vagy a Gemini Pro-nak a válasz létrehozására, az *„feltöri a GenAI szolgáltatást"*, és végül adatokat lop az e-mailekből – nyilatkozta Nassi, a kutatás vezetője (URL9).

Habár ilyen férget egyelőre nem találtak a kibertérben, több kutató és biztonsági szakértő is arra figyelmeztetett, hogy ezek elterjedése valós veszélyt jelent. Az OpenAI szóvivőjének nyilatkozata szerint: *„Úgy tűnik, hogy megtalálták a módját annak, hogy kihasználják a prompt-injection típusú sebezhetőségeket azáltal, hogy olyan felhasználói bemenetre támaszkodnak, amelyet nem ellenőriztek vagy szűrtek."* (URL12).

Amennyiben egy adathalász támadás ilyen eszközzel is megvalósítható, úgy jogosan merül fel az a kérdés, hogy miért lenne szükség a továbbiakban a social engineering sokkal összetettebb technikájára az elkövetőknek. A válasz a védekezési lehetőségekben keresendő. A hivatkozott tanulmány megjelenése után a Google és az OpenAI fejlesztői is szorgalmazták a kutatókkal való találkozót, és több kiberbiztonsággal foglalkozó kutató is – például Sahar Abdelnabi, a németországi CISPA Helmholtz Információbiztonsági Központ kutatója – olyan nyilatkozatot tett, amelyben felhívták a fejlesztők figyelmét a technikai hiányosságok kezelésére, valamint a megfelelő alkalmazástervezés fontosságára. Míg a féreggel megvalósítható támadások elleni védekezés kulcsa a technológia fejlesztése, addig a social engineering támadások esetén a felhasználó biztonságtudatossága és természetes gyanakvása legalább olyan fontos, mint az informatikai védelem.

A Bevezetőben is említett tény – miszerint az MI-t használó virtuális asszisztensek felhasználhatók az ilyen típusú bűncselekmények elkövetésére –, valamint a hivatkozott tanulmány egyértelműen abba az irányba mutatnak, hogy hamarosan a kibertámadások elleni védekezés első védelmi vonalát nem kizárólag a technikai megoldások jelentik. Elengedhetetlen lesz az emberi jelenlét, még akkor is, ha látszólag a humán tényező kiszorul ebből a körből. A biztonságos jelszókezelési gyakorlat, a többfaktoros hitelesítés, és bizonyos folyamatok automatizálásból kizárása mind olyan összetevője a kiberbiztonsági stratégiának, amely megkérdőjelezhetetlenül hozzá fog járulni annak hatékonyságához. A vezetők felelőssége nemcsak az információbiztonsági szabályok elrendelésében fog manifesztálódni, hanem abban is, hogy a hálózat üzemeltetésében részt vevők megfelelően felkészültek legyenek a védekezésben betöltött szerepük vonatkozásában, és megkapják a szervezeten belül a szükséges képzéseket. A felelősségi rendszer ilyen módon átláthatóvá tételével a kiberbiztonsági rendszer nemcsak hatékony, hanem teljes egész is lesz.

Záró gondolatok

A kiberbiztonsági stratégia megalkotása és alkalmazása során az MI tagadhatatlanul központi szerepet tölt be. Számtalan támadás automatikus elhárítására alkalmasak a már létező rendszerek, és a kutatók, fejlesztők munkájának, valamint a gépi tanulásnak köszönhetően a biztonsági rendszerek folyamatosan alkalmazkodnak az új fenyegetésekhez. A kibertámadások elleni védekezés proaktívvá válik, a hálózatok biztonsága ezáltal növekszik.

A social engineering támadások hatékonyságát ugyan növeli az MI használata, azonban az továbbra is a manipuláción és az emberi természet gyengeségein

alapul. A védekezés hatékonysága is csak úgy növelhető, ha nem zárjuk ki az embert. A támadások komplexek, azokban olyan összefüggések szerepelhetnek, amelyeket az MI egyelőre nem képes felismerni. Az ember képes kontextusba helyezni olyan adatokat, információkat, amelyek az automatizált folyamatokon átmennek, köszönhetően a kreativitásnak. Ez az ember sajátja, amelyet a gépek nem tudnak reprodukálni. A sikeres kibervédelem jövője abban rejlik, hogy az ember nem zárja ki saját magát a folyamatokból, hanem megtanulja a saját és a gépek előnyeit ötvözni, és azt egységként használni.

Felhasznált irodalom

Aliman, N. M., Kester, L. J. H. M., Werkhoven, P., & Yampolsky, R. (2019). Orthogonality-based disentanglement of responsibilities for ethical intelligent systems. In P. Hammer, B. Goertzel, & M. Iklé (Eds.), *Artificial general intelligence* (pp. 3–13). Springer.

Bányász P. (2018). Social engineering and social media. *Nemzetbiztonsági Szemle, 5*(1), 59–77.

Bostrom, N. (2015). *Szuperintelligencia – Utak, veszélyek, stratégiák* (Szalay Á., Ford.) [Eredeti mű megjelenése 2014]. Ad Astra Kiadó.

Butz, M. V. (2021). Toward strong AI. *KI – Künstliche Intelligenz, 35*(4), 391–399.

Chaturvedi, S., Patvardhan, C., & Lakshmi, C. V. (2023). AI value alignment problem: The clear and present danger. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 1–6). IEEE. https://doi.org/10.1109/ISCON57294.2023.10112100

Dub M. (2021). A social engineering támadások megelőzésének lehetőségei. *Hadmérnök, 16*(3), 137–187. https://doi.org/10.32567/hm.2021.3.10

Falade, P. V. (2023). Decoding the threat landscape: ChatGPT, FaudGPT, and WormGPT in social engineering attacks. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9*(5), 132–133.

Florêncio, D., & Herley, C. (2019). The economics of phishing and malicious email. In *Proceedings of the 28th USENIX Security Symposium* (pp. 963–980). USENIX Association.

Ghosh, A., Sufian, A., Sultana, F., Chakrabarti, A., & De, D. (2020). Fundamental concepts of convolutional neural network. In V. Balas, R. Kumar, & R. Srivastava (Eds.), Recent trends and advances in artificial intelligence and internet of things (Intelligent Systems Reference Library, Vol. 172, pp. 519–567). Springer. https://doi.org/10.1007/978-3-030-32644-9_36

Hadnagy, C. (2018). *Social engineering: The science of human hacking*. John Wiley & Sons Inc.

Hassabis, D. (2017). Artificial intelligence: Chess match of the century. *Nature, 544*(7651), 413. https://doi.org/10.1038/544413a

Jagodics I., & Kollár Cs. (2023). 21. századi social engineering támadások, védekezés és szervezeti hatások Európában. *Belügyi Szemle, 71*(1), 113–126. https://doi.org/10.38146/BSZ.2023.1.6

Kelemen R., & Németh R. (2019). A kibertér alanyai és sebezhetősége. *Szakmai Szemle: A Katonai Nemzetbiztonsági Szolgálat tudományos-szakmai folyóirata, (3)*, 95–118.

Klein T., & Szabó A. (2018). A cybercrime, mint infokommunikációs jogi probléma. In Polyák G., & Lévai D. (Szerk.), *Tanulmányok a technológia- és cyberjog néhány aktuális kérdéséről* (pp. 123–132). Médiatudományi Intézet.

Kristóf T. (2002). A mesterséges neurális hálók a jövőkutatás szolgálatában. In Hideg É. (Szerk.), *Jövőelméletek sorozat* (pp. 23–26). BGE Jövőkutatási Kutatóközpont.

Leung, A. C. M., & Bose, I. (2008). Indirect financial loss of phishing to global market. In *ICIS 2008 Proceedings* (Paper 5). Association for Information Systems. https://aisel.aisnet.org/icis2008/5

Lumacad, G. S., & Namoco, R. A. (2023). Multilayer perceptron neural network approach to classifying learning modalities under the new normal. *IEEE Transactions on Computational Social Systems, 1*(99), 1–13. https://doi.org/10.1109/TCSS.2023.3251566

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE, 49*, 8–30. https://doi.org/10.1109/JRPROC.1961.287775

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys, 54*(1), 1–41. https://doi.org/10.1145/3425780

Mitnick D. K. & Simon L. W. (2002). *A legendás hacker – A megtévesztés művészete*. Perfact-Pro Kft. Kiadó.

Mühlhoff, R. (2019). Human-aided artificial intelligence: Or, how to run large computations in human brains? Toward a media sociology of machine learning. *New Media & Society, 22*(10), 1868–1884. https://doi.org/10.1177/1461444819885334

Nassi, B., Cohen, S., & Bitton, R. (2024). Here comes the AI Worm: Unleashing zero-click worms that target GenAI-powered applications. Retrieved from https://sites.google.com/view/compromptmized

Németh R. (2019). Kibertámadások gazdasági vonatkozásai a vállalati szférában. In Dernóczy-Polyák A. (Szerk.), *Kutatási jelentés* (1. kötet, pp. 307–325).

Oroszi E. (2019). *Az információbiztonság lélektana*. Nemzeti Kibervédelmi Intézet.

Rössler A., Cozzolino D., Verdoliva L., Riess Ch., Thies J., & Nießner M. (2019). *FaceForensics++: Learning to detect manipulated facial images*. arXiv. https://arxiv.org/abs/1901.08971

Saygin A., Cicekli I., & Akman V. (2000). Turing test: 50 years later. *Minds and Machines, 10*(4), 463–518.

Searle J. R. (1996). Az elme, az agy és a programok világa. In Pléh Cs. (Szerk.), *Kognitív tudomány* (pp. 136–151). Osiris Kiadó.

Sörös T., & Váczi D. (2013). Social engineering a biztonságtechnika tükrében. *XXXI. Országos Tudományos Diákköri Konferencia, Had- és Rendészettudományi Szekció*, Budapest.

Takale G., Wattamwar A., Saipatwar S., Saindane H., & Tushar P. (2024). Comparative analysis of LSTM, RNN, CNN and MLP machine learning algorithms for stock value prediction. *Journal of Firewall Software and Networking, 2*(1).

Youvan, D. (2024). *Self-improving AI and the ungodly basis for intelligent design: Implications for the future of intelligence, creation, and a simulated universe*. https://doi.org/10.13140/RG.2.2.26336.90881

A cikkben szereplő online hivatkozások

URL1: *Európai Parlament: Mi az a mesterséges intelligencia és mire használják?* https://www.europarl.europa.eu/topics/hu/article/20200827STO85804/mi-az-a-mesterseges-intelligencia-es-mire-hasznaljak

URL2: *Fortune: Elon Musk says there's a 10% to 20% chance that AI 'goes bad,' even while he raises billions for his own startup xAI.* https://fortune.com/2024/10/30/elon-musk-ai-could-go-bad-existential-threat-xai-fundraising/

URL3: *Investopedia: What Is Artificial Intelligence (AI)?* https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp

URL4: *ISO: Unravelling the secrets of natural language processing.* https://www.iso.org/artificial-intelligence/natural-language-processing

URL5: *Mesterséges Intelligencia Elektronikus Almanach: A mesterséges intelligencia története.* http://project.mit.bme.hu/mi_almanach/books/aima/ch01s03

URL6: *Mitnick Security: 6 Types of Social Engineering Attacks.* https://www.mitnicksecurity.com/blog/6-types-of-social-engineering-attacks

URL7: *Nakivo: Microsoft Office 365 Advanced Threat Protection: Complete Overview.* https://www.nakivo.com/blog/microsoft-office-365-advanced-threat-protection-overview/

URL8: *Pigro Blog: History of AI: Deep Blue and Strong and Weak Artificial Intelligence.* https://blog.pigro.ai/en/deep-blue-and-strong-and-weak-ai-the-story-of-ai-in-the-90s

URL9: *The Independent: AI worm that infects computers and reads emails created by researchers.* https://www.independent.co.uk/tech/ai-worm-computer-security-chatgpt-malware-b2506594.html

URL10: *The New York Times: What Ever Happened to IBM's Watson?* https://www.nytimes.com/2021/07/16/technology/what-happened-ibm-watson.html

URL11: *The Royal Institution: 10 AI milestones of the last 10 years.* https://www.rigb.org/explore-science/explore/blog/10-ai-milestones-last-10-years

URL12: *The Spectator: Should we fear AI? James W. Phillips and Eliezer Yodkowsky in conversation.* https://www.spectator.co.uk/article/should-we-fear-ai-james-w-phillips-and-eliezer-yudkowsky-in-conversation/

URL13: *Unite.AI: A nagy nyelvi modellek (LLM) erejének leleplezése.* https://www.unite.ai/hu/nagy-nyelvi-modellek/

URL14: *Wired.com: Here come the AI worms.* https://www.wired.com/story/here-come-the-ai-worms/

A cikk APA szabály szerinti hivatkozása

Tárkányi R., & Németh R. (2025). A mesterséges intelligencia szerepe a social engineering támadások megelőzésében. *Belügyi Szemle*, *73*(8), 1579–1597. https://doi.org/10.38146/BSZ-AJIA.2025.v73.i8.pp1579-1597

Nyilatkozatok

Összeférhetetlenség

A szerzők nem jelentettek összeférhetetlenséget.

Finanszírozás

A szerzők nem kaptak pénzügyi támogatást a kutatáshoz, a szerzőséghez és/vagy a cikk publikálásához.

Etikai nyilatkozat

Jelen cikkhez nem kapcsolódik adatkészlet.

Levelező szerző

A cikk levelező szerzője Tárkányi Edina, aki az etark@protonmail.com e-mail címen érhető el.