HADMÉRNÖK
VÉDELEMINFORMATIKA

20. évfolyam (2025) 2. szám 129–142.
DOI: 10.32567/hm.2025.2.8

Szabó Gergő[1]

A generatív mesterséges intelligencia használata a kibertudatosítási képzésekben, 1. rész

The Use of Generative Artificial Intelligence in Cybersecurity Awareness Training, Part 1

Absztrakt

A tanulmány célja annak vizsgálata, hogy miként alkalmazható a generatív mesterséges intelligencia (GenMI) a kibertudatossági képzésekben, különös tekintettel az oktatás személyreszabhatóságára, hatékonyságára és elfogadottságára. A kutatás elméleti része az ADDIE-modell mentén tárgyalja a GenMI szerepét az oktatástervezésben, bemutatva annak lehetőségeit a tananyagfejlesztés, a személyre szabás és a visszacsatolás területén. Részletesen ismerteti a RAG-technológia szerepét is, különösen a tartalmi pontosság és az adatbiztonság aspektusaira fókuszálva. Az empirikus kutatás egy 109 fő részvételével végzett kérdőíves vizsgálatra épül, amelyet a Technology Acceptance Model (TAM) keretrendszere szerint szerkesztettem. A kérdőív többek között phishing szimulációkat, egy generált szcenáriót, valamint attitűdkérdéseket tartalmazott.

Kulcsszavak: mesterséges intelligencia, kibertér, kibertámadás

Abstract

This study examines how generative artificial intelligence (GenAI) can be applied in cybersecurity awareness training, with particular emphasis on personalisation, effectiveness, and user acceptance. The theoretical section follows the ADDIE instructional design model and explores the role of GenAI in educational planning, highlighting its potential in content development, adaptive learning, and feedback mechanisms. Special attention is given to

[1] Hallgató, Nemzeti Közszolgálati Egyetem, e-mail: sz.gergo628@gmail.com

Retrieval-Augmented Generation (RAG) technologies, focusing on content accuracy and data protection considerations.

The empirical part of the research is based on a questionnaire survey conducted with 109 participants and designed according to the Technology Acceptance Model (TAM). The questionnaire included phishing simulations, an AI-generated cybersecurity scenario, and attitude-based questions. The aim of the study is to assess how non-expert users perceive and evaluate cybersecurity training materials generated by artificial intelligence, and to identify key factors influencing their acceptance and perceived usefulness.

Keywords: artificial intelligence, cybersecurity, awareness training, phishing

Bevezetés

A digitális technológiák fokozódó jelenléte a munkahelyi és a magánélet területén egyre inkább indokolttá teszi a kiberbiztonsági tudatosság fejlesztését. Az olyan fenyegetések, mint az adathalászat, a célzott támadások vagy a social engineering,[2] gyakran nem technikai sebezhetőségeken, hanem emberi tényezőkön keresztül érik el céljukat.

A SlashNext 2023-ról készült jelentése szerint a ChatGPT megjelenését követő évben 1265%-kal nőtt a rosszindulatú adathalász-e-mailek száma, különösen a hitelesítő adatokat célzó támadásokban. Ugyanebben az időszakban a magyar nyelv védőhatása is megszűnni látszik, a generatív mesterséges intelligencia (GenMI) eszközeinek segítségével már anyanyelvi szintű adathalász-üzeneteket is lehet készíteni, megszüntetve ezzel a korábbi figyelmeztető nyelvi hibákat.[3]

A leghatékonyabb védekezést ebben az esetben a megfelelő ismeretanyag és felkészültség biztosítja, amelyet célzott oktatási programok révén lehet fejleszteni.

A GenMI, különösen a nagy nyelvi modellek (Large Language Modells, LLM) megjelenésével új távlatok nyíltak az oktatás személyre szabása terén. Az ilyen rendszerek képesek az egyéni tudásszinthez, tanulási stílushoz vagy motivációhoz illeszkedő szcenáriók, példák és magyarázatok generálására, így támogatva az adaptív tanulást. Ez a megközelítés különösen ígéretes a kiberbiztonsághoz hasonló, dinamikusan változó területeken, ahol az oktatási tartalmak gyors elavulása állandó kihívást jelent.

Ugyanakkor a GenMI által generált tartalmak használata etikai, minőségi és elfogadottsági kérdéseket is felvet: vajon megbízhatók-e ezek az anyagok? Mennyire tanulnak belőlük az emberek? És milyen mértékben fogadják el az ilyen típusú oktatást azok, akik nem rendelkeznek mélyebb technológiai ismeretekkel?

A kutatás ezekre a kérdésekre keresi a választ, az elméleti háttér áttekintése mellett pedig egy kérdőíves vizsgálatot is tartalmaz. A vizsgálat célja, hogy felmérje a generatív mesterséges intelligenciával létrehozott kiberbiztonsági

[2] Olyan manipulációs technikák összessége, amelyeket arra használnak, hogy az embereket bizalmas információk, például jelszavak, pénzügyi adatok vagy hozzáférési jogosultságok önkéntes kiadására vegyék rá, gyakran azzal, hogy megbízható forrásnak álcázzák magukat a támadók. Lásd: Shaw et al. 2009.

[3] SlashNext 2023.

oktatási szcenáriók elfogadottságát és értékelését nem szakmabeli felhasználók körében. A cikksorozat első részében az elméleti háttér és a rendelkezésre álló technológiai megoldások áttekintéséről lesz szó, míg a második rész a kutatási kérdéseket és a kérdésekhez kapcsolódó hipotézisek tesztelését mutatja be.

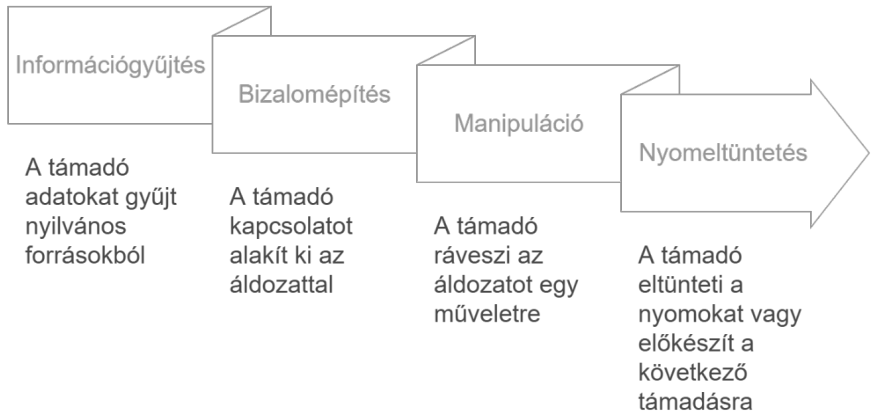
A generatív mesterséges intelligencia szerepe a kiberbiztonsági oktatásban

A fejezet tárgyalása előtt fontos röviden tisztázni néhány gyakran egymás helyett használt, de tartalmilag eltérő fogalmat. Az információvédelem az információk védelmét jelenti a jogosulatlan hozzáféréstől, függetlenül attól, hogy azok digitális vagy fizikai formában léteznek. Ennek egy szűkebb, technológiai fókuszú része az informatikai biztonság, amely az információkat kezelő rendszerek, hálózatok és szerverek védelmére koncentrál. Ezzel szemben a kiberbiztonság már kifejezetten a hálózati környezetben megjelenő fenyegetések, például adathalászat és zsarolóvírusok elleni védelmet helyezi előtérbe. Végül a kibervédelem elsősorban a nemzeti vagy kritikus infrastruktúrák védelmére használt fogalom, amely magában foglalja a technikai mellett a stratégiai, jogi és szervezeti védekezési mechanizmusokat.[4]

A kibervédelem napjainkban dinamikusan fejlődő szakterület, hiszen folyamatosan jelennek meg újabb támadási vektorok és kritikus sérülékenységek. Az egyik legrégebb óta jelen lévő, ugyanakkor máig kiemelten veszélyes támadási forma a social engineering, ezen belül is az adathalászat. A social engineering célja, hogy a támadó a felhasználó jóhiszeműségét kihasználva szerezzen hozzáférést védett rendszerekhez, vagy szivárogtasson ki érzékeny személyes, üzleti vagy piaci adatokat, jellemzően haszonszerzés céljából.

A támadások sokszor egy jól felépített manipulációs folyamatra épülnek. Ezt a folyamatot jól modellezi a Social Engineering Attack Framework (social engineering támadási keretrendszer, SEAF),[5] amely négy fő fázisból áll. Első lépés az információgyűjtés, amikor a támadó a célpontról adatokat gyűjt nyilvános forrásokból vagy közösségi médiából. Második lépésként a támadó interakcióba lép az áldozattal, és bizalmi kapcsolatot alakít ki. Harmadik lépésként a támadó a megszerzett bizalom révén rábírja az áldozatot valamilyen műveletre, például linkre kattintásra vagy adatok átadására. Végezetül a támadás befejeződik, a támadó eltünteti a nyomokat, vagy további támadásokat készít elő (1. ábra).

[4] Merritt 2024.
[5] Mouton et al. 2014.



1. ábra: A SEAF fázisai
Forrás: a szerző szerkesztése

Ez a támadási módszer elsősorban az alkalmazottakat célozza, hiszen minden munkavállalónak szüksége van valamilyen fokú jogosultságra feladatai ellátásához. Az ő megtévesztésük révén a támadók képesek megkerülni a technikai védelmi rendszereket is, ezért terjedt el a kibervédelmi szakmában az a mondás, hogy „a kiberbiztonság leggyengébb láncszeme maga az ember".

A kiberfenyegetések változó természete és az oktatás szükségessége

A Verizon adatai szerint az adatszivárgások több mint 74%-a közvetlenül az emberi tényezőre vezethető vissza,[6] beleértve a hibás döntéshozatalt, a gyenge jelszavak használatát vagy a figyelmetlen linkkattintást. Ezzel párhuzamosan a Proofpoint kutatása arra figyelmeztet, hogy a szervezetek 90%-a nem technikai áttörések, hanem social engineering és célzott adathalászat révén kerül veszélybe.[7] Ezt a képet erősíti az ENISA Threat Landscape 2024 is, amely alapján a válaszadók 31%-a szerint a felhasználói hiba volt az incidens kiváltó oka, míg 17% a többlépcsős hitelesítés (multi-factor authentication, MFA) hiányát nevezte meg a támadás fő tényezőjeként, tehát az incidensek legalább 48%-a emberi mulasztásra vagy hiányosságra vezethető vissza.[8]

A helyzetet tovább súlyosbítja a generatív mesterséges intelligencia gyors térnyerése, amely a támadók kezében is hatékony eszközzé vált. A SlashNext jelentése szerint a ChatGPT nyilvános megjelenését követő egy éven belül az MI által támogatott adathalászkampányok száma 1265%-kal nőtt. Ez különösen az e-mailek nyelvi hitelességét és személyre szabását érinti: a támadók ma már képesek anyanyelvi

[6] Verizon 2024.
[7] Proofpoint 2023.
[8] European Union Agency for Cybersecurity 2024.

szintű, kontextushoz illesztett és pszichológiailag célzott üzeneteket generálni, ez pedig a felhasználók számára rendkívül megnehezíti a megtévesztés felismerését.[9]

Mindez világosan rámutat arra, hogy a hagyományos, évente egyszer tartott kibertudatossági képzés már nem elegendő a mai fenyegetettségi környezetben. Az oktatásnak alkalmazkodnia kell a gyorsan változó támadási technikákhoz, figyelembe kell vennie a résztvevők eltérő előismereteit, és folyamatos, valós idejű visszacsatolással kell támogatnia a tanulási folyamatot.

A generatív mesterséges intelligencia e téren új lehetőségeket kínál, képes valósághű, egyénre szabott szcenáriók generálására, amelyek lehetővé teszik, hogy a tanulók biztonságos környezetben tapasztalják meg a különböző támadások működését és következményeit. Ezáltal nemcsak az elméleti tudás mélyül, hanem a gyakorlati felkészültség és a helyzetfelismerési készség is fejlődik.

Ez a trend azonban túlmutat a technológiai újításokon, stratégiai jelentőségű. A képzések személyre szabása, különösen a kiberbiztonság területén, a szervezeti reziliencia egyik kulcsfontosságú tényezőjévé válik. A generatív mesterséges intelligencia ebben a kontextusban nem csupán egy új eszközként, hanem új oktatási paradigma előfutáraként jelenik meg.

A generatív mesterséges intelligencia támogatási lehetőségei a NIST CPLP keretrendszerben

A kibertudatossági oktatási tervet minden szervezetnek saját kockázatelemzése és értékelése alapján kell kialakítania. Nem létezik univerzálisan alkalmazható sablon, amely minden környezetben módosítás nélkül használható lenne. A kockázatértékelés eredményei határozzák meg azokat a stratégiai szempontokat és prioritásokat, amelyek mentén az oktatási célok kijelölhetők. Ezek figyelembevételével alakítható ki egy olyan tanulási program, amely illeszkedik a szervezet fenyegetettségi profiljához és kockázattűrő képességéhez.

Ezt a megközelítést a National Institute of Standards and Technology Cybersecurity and Privacy Learning Program (Szabványügyi és Technológiai Nemzeti Intézet Kiberbiztonsági és Adatvédelmi Tanulási Programja, NIST CPLP) szerzői is hangsúlyozzák, a kiberbiztonság nem kizárólag technológiai vagy IT-feladat, hanem szervezeti szintű felelősség, amelyben minden szereplőnek, az első számú vezetőtől a beosztott munkatársakig, van szerepe.[10]

Ami minden jól működő kibertudatossági oktatási terv közös vonása, az a biztonsági kultúra kialakítása iránti igény. A szervezeten belül minden szinten meg kell hogy jelenjen az a szemlélet, amely támogatja, ösztönzi és fenntartja a biztonságtudatos viselkedést. A biztonsági kultúra kialakításának igénye azonban nem alulról, hanem felülről kell hogy induljon, a felső vezetés részéről, stratégiai szintről kell érkeznie annak az elköteleződésnek, amely megalapozza a célok megfogalmazását, és biztosítja az oktatási program végrehajtásához szükséges jogosítványokat.

[9] SlashNext 2023.
[10] Merritt et al. 2024.

A stratégiai tervben a vállalat vezetésének világosan meg kell határoznia azokat az alapelveket és irányvonalakat, amelyek mentén a kibertudatossági képzés kialakítható. Idetartozik a szervezet víziója és küldetése, a stratégiai célok és feladatok, az oktatás megközelítése és résztervei, az alkalmazandó eljárásrendek, valamint azok a mérőszámok és jelentési rendszerek, amelyekkel a program hatékonysága mérhető.

A stratégiai irányvonalak részét képezi a szervezet kockázatkezelési eljárásrendje is. A kialakított biztonsági kultúra lehetőséget teremt arra, hogy az alkalmazottak a saját kockázatértékelésük alapján hozhassanak döntéseket.[11]

A képzések elsődleges célja a tudásbeli és készségszintű hiányosságok azonosítása és pótlása. A NIST CPLP ebben a kontextusban kifejezetten javasolja a generatív mesterséges intelligencián alapuló tananyagok és szerepkör-specifikus képzések alkalmazását. Ennek köszönhetően minden alkalmazott olyan, a saját munkakörére szabott biztonsági ismereteket szerezhet, amelyek közvetlenül hozzájárulnak a szervezet átfogó kiberbiztonságához és kockázatkezelési stratégiájához.

A képzési programoknak szorosan illeszkedniük kell a stratégiai tervben rögzített vállalati célkitűzésekhez, valamint a kockázatkezelési elvekhez és a hatályos adatvédelmi és megfelelőségi *(compliance)* előírásokhoz. A képzések hatékonyságát folyamatosan monitorozni szükséges, például az alkalmazottak viselkedésének változása vagy az incidensek számának alakulása alapján, így lehetőség nyílik az oktatási program folyamatos fejlesztésére.

A kibertudatossági oktatás másik alappillérét a vállalati szabályzatok és eljárásrendek képezik. Ezek világosan megfogalmazzák a szervezet elvárásait az alkalmazottak felé: mit és hogyan kell végrehajtani. A szabályoknak egyszerű, közérthető állításoknak kell lenniük, amelyek az elvárt viselkedést határozzák meg. Az eljárásrendek rögzítik, hogyan történik a szabályok gyakorlati végrehajtása, meghatározzák az egyes szereplők felelősségi köreit, a végrehajtás lépéseit és fázisait, valamint a kapcsolódó dokumentációkat.

Fontos, hogy ezek a szabályzatok és eljárásrendek összhangban legyenek a vállalat más, nem kiberbiztonsági területeken alkalmazott belső irányelveivel. Ez garantálja, hogy az alkalmazottak számára érthető, elfogadható és teljesíthető napi elvárások szülessenek. Idetartozhat például az üzletfolytonossági terv vagy a katasztrófaelhárítási protokoll is.

Ezek összessége adja meg azt a stabil alapot, amelyre egy hatékony kibertudatossági oktatási terv épülhet. Az alkalmazottaknak éppen ezeket a szabályokat és eljárásrendeket kell megismerniük, és ezek végrehajtására kell őket felkészíteni a képzés során. Ezután kezdődhet meg az oktatási terv kidolgozása az ADDIE-modell mentén, amelyet a következőkben kiberbiztonsági szempontból elemzek.

Elemzés

A generatív mesterséges intelligencia ebben a szakaszban jelentős támogatást nyújthat az oktatási tervet készítők számára. A GenMI képes gyorsan elvégezni az adatgyűjtést és a trendanalízist, például a legfrissebb közismert sebezhetőségek és kitettségek

[11] Merritt et al. 2024.

adatbázisának (Common Vulnerabilities and Exposures, CVE[12]) feldolgozásával. Mivel ezek az adatbázisok a legkritikusabb sebezhetőségeket listázzák, figyelmen kívül hagyásuk jelentősen növeli a szervezeti kockázatot. A GenMI ezenkívül képes szakcikkek, iparági jelentések és kiberbiztonsági kutatások gyors áttekintésére is, és az ezekből kinyert tudást képes összevetni a szervezet korábbi biztonsági eseményeivel. Ennek eredményeképp azonosíthatók azok a kihívások és fenyegetések, amelyekre az oktatási tervnek reagálnia kell.

A fenti adatokra építve a GenMI segíthet a tananyag összeállításában, biztosítva annak naprakészségét, különösen akkor, ha új támadási vektorok jelennek meg. A vállalat stratégiai célkitűzéseinek figyelembevételével az MI javaslatokat tehet a fenyegetések rangsorolására is, így a képzés a legrelevánsabb és legsürgetőbb problémákra összpontosíthat.

Amennyiben a vállalat korábban már tartott kibertudatossági tréningeket, a mesterséges intelligencia képes feldolgozni az ezekből származó visszacsatolásokat, beleértve a nyílt szöveges válaszokat is. A természetesnyelv-feldolgozás technológiájával az MI képes elemezni a kulcskifejezéseket, problémás tartalmakat, és ezáltal segítheti a tréningfejlesztőket abban, hogy a hibásan értelmezett vagy alulteljesített területeket javítsák.

Tervezés és fejlesztés

A GenMI képes a munkavállalók korábbi tréningjeiből vagy előzetes felmérésekből származó eredmények alapján azonosítani az egyéni hiányosságokat és erősségeket. Ezek alapján tanulói profilokat alakít ki, amelyekre építve különböző nehézségi szintű oktatási tartalmak állíthatók össze. A kezdők számára az alapfogalmak és a gyakorlati példák kapnak nagyobb hangsúlyt, míg a haladó szintű felhasználók számára technikai elemzéseket, esettanulmányokat és szimulációkat kínál a rendszer.

A generatív mesterséges intelligencia az oktatás gyakorlati dimenziójában is kulcsszerepet játszik. Lehetővé teszi interaktív, valósághű szimulációk generálását, amelyek során a dolgozók a megszerzett kiberbiztonsági ismereteket biztonságos környezetben alkalmazhatják. Ilyen például a HacWare által fejlesztett MI-eszköz, amely minden munkavállaló számára egyedi adathalász-e-maileket generál a valós támadási mintázatok alapján. Nem egységes kampányok futnak tehát, hanem személyre szabott, dinamikusan változó tartalmakat küld ki a rendszer.[13] Fontos ugyanakkor megjegyezni, hogy kormányzati vagy magasabb biztonsági osztályba sorolt rendszerek esetében az ilyen „teszttámadások" generálása különösen körültekintő tervezést, valamint előzetes engedélyezést igényelhet.

A GenMI emellett képes valós incidensek elemzésére, és azok alapján olyan szimulációs forgatókönyvek előállítására, amelyek tükrözik az adott szervezet valóságos működését. Ezek a forgatókönyvek könnyen paraméterezhetők, így a különböző

[12] CVE: Egy nyilvános, szabványosított elnevezést és azonosítót biztosító adatbázis az ismert szoftveres és hardveres sebezhetőségek számára. A CVE-azonosító célja, hogy a különböző biztonsági eszközök és szolgáltatások ugyanarra a sebezhetőségre egységes hivatkozással tudjanak utalni. Lásd CVE [é. n.].

[13] AI-Generated Phishing Simulations Are a Game Changer for MSPs 2023.

szervezeti egységek, például az IT, a kiberbiztonsági részleg vagy a marketingosztály tudásszintjéhez és feladataihoz igazodó szituációk generálhatók. A szimulációk csoportos formában is megvalósíthatók, ezzel is szemléltetve, hogyan képes egy ártalmatlannak tűnő incidens gyorsan eszkalálódni.

A multimédiás tartalmak integrálása szintén hozzájárul az élményszerű tanuláshoz. A GenMI képes infografikákat, oktatóvideókat vagy interaktív vizuális elemeket létrehozni, amelyek például bemutatják az adathalász-támadások evolúcióját. A deepfake technológia, amelyet a kiberbűnözők is egyre gyakrabban használnak, visszafordítható az oktatás szolgálatába. A Breacher.ai például deepfake videókkal és hangfelvételekkel támogatja a szimulációs tréningeket: ezek révén a dolgozók valósághű, megtévesztő MI-tartalmakkal találkozhatnak, amelyeket biztonságos környezetben tanulnak felismerni.[14]

Egy empirikus kutatásban Brito és munkatársai egy rövid, elméleti és gyakorlati elemeket is tartalmazó, mesterséges intelligenciát bemutató modult illesztettek be hat kiberbiztonsági kurzusba, amelyet kérdőíves módszerrel mértek előtte és utána. A résztvevők 30%-os tudásszint-növekedésről számoltak be, a válaszok 84%-os egyezést mutattak a tananyag tartalmi elemeivel, az általános vélekedés pozitív volt (átlagos sentiment score: +0,50).[15] Ez a példa jól mutatja, hogy a GenMI-alapú tartalomfejlesztés nemcsak technológiai, hanem pedagógiai szempontból is képes hozzáadott értéket teremteni a képzések során. A tanulmány megerősíti, hogy a generatív mesterséges intelligencia szerepe nem csupán elméletben, hanem a gyakorlatban is releváns és hatékony a kiberbiztonsági tananyagok fejlesztésében.[16]

Fontos azonban hangsúlyozni, hogy a GenMI által készített oktatási tartalmakat minden esetben szakmai validációnak kell alávetni. A mesterséges intelligencia nem válthatja ki teljes mértékben az emberi szereplőket: az oktatási célok és a vállalat belső szabályozási környezetének összehangolása továbbra is az oktatásszervező feladata marad. A GenMI tehát nem autonóm megoldás, hanem egy intelligens támogató eszköz, amely a megfelelő kontroll mellett jelentősen növelheti a képzések hatékonyságát és relevanciáját.

Megvalósítás

A GenMI-chatbotok szimulált támadási szituációkban is bevethetők, például szerepjáték-alapú gyakorlatokon keresztül. Egy ilyen rendszer képes „támadóként" viselkedni: például egy technikai támogatást végző munkatárs szerepében próbálja rávenni a felhasználót arra, hogy bizalmas információkat osszon meg vele, például jelszót. A gyakorlat végén azonban a chatbot „átalakul" oktatóvá, és visszajelzést ad arra vonatkozóan, hogy milyen jelek utalhattak volna a megtévesztő szándékra. Ilyen módon működött a ChatGPT is a New Jersey állami kiberbiztonsági központ belső

[14] Breacher.ai 2025.
[15] Brito et al. 2025.
[16] Brito et al. 2025.

képzési folyamatában, ahol kérdezz-felelek típusú támogatóként segítette a dolgozók tájékozódását.[17]

A képzés részeként a GenMI automatizált módon képes különböző típusú tesztek és vizsgák előállítására. Ezek a felmérések lefedhetik az egyszerű feleletválasztós kérdésektől kezdve a nyílt végű kérdések értékelésén át az interaktív, multimédiás tesztekig terjedő spektrumot. A tanulók egyéni előrehaladásához igazodva a rendszer képes adaptív módon nehezíteni vagy egyszerűsíteni a feladatokat, ezzel is támogatva a hatékony tanulási folyamatot.

A tanulási élmény javítása érdekében a generatív mesterséges intelligencia gamifikációs elemek integrálására is alkalmas. Valós idejű szimulációkkal, pontgyűjtő rendszerekkel vagy versenyszerű kihívásokkal növelhető a tanulói elköteleződés és motiváció. A tanulási folyamat ezáltal nemcsak hatékonyabbá, hanem élményszerűbbé is válik, így hosszabb távon is hozzájárul a megszerzett tudás mélyebb rögzüléséhez.

Értékelés

Az oktatási programok hatékonyságának mérése és értékelése elengedhetetlen a folyamatos fejlesztés érdekében. A GenMI-alapú rendszerek ebben a folyamatban kulcsszerepet játszanak, nem csupán objektív teljesítményértékelést tesznek lehetővé, hanem személyre szabott, azonnali visszacsatolást is biztosítanak a képzésben részt vevő alkalmazottak számára. Ezáltal a tanulók gyorsabban felismerhetik saját hiányosságaikat, és célzottabban fejleszthetik készségeiket.

A GenMI-rendszerek nem csupán értékelnek, hanem pedagógiai támogatást is nyújtanak, a helyes megoldás mellett azonnali, közérthető magyarázatot is képesek adni. Ez különösen fontos a hibák értelmezése és a tanulási folyamat elmélyítése szempontjából. A visszacsatolás tartalma ráadásul összekapcsolható a tanuló egyéni profiljával, például korábbi eredményeivel vagy tanulási preferenciáival, így a képzési anyag még jobban személyre szabható.

Összességében az értékelés GenMI által támogatott módja nem csupán hatékonyabbá, hanem adaptívabbá és még inkább tanulóközpontúvá is teszi a képzési folyamatot. A visszacsatolás azonnalisága és relevanciája lehetőséget teremt arra, hogy a kiberbiztonsági oktatás valóban dinamikusan reagáljon a tanulók fejlődési szükségleteire és a folyamatosan változó fenyegetettségi környezetre. A hatékonyság mérése során az önértékelésen túl objektív metrikák, például a szimulált támadásokra adott reakcióidő, a hibaarány vagy az incidensek számának változása is fontos visszajelzést adhatnak.

Etikai kihívások a generatív mesterséges intelligencia használata során

Fontos azonban hangsúlyozni, hogy az ingyenes verziók hatékonysága erősen függ a felhasználó által megfogalmazott kérdések (promptok) minőségétől. Ahhoz, hogy

[17] New Jersey Cybersecurity & Communications Integration Cell [é. n.].

ezek az eszközök valóban figyelembe vegyék a szervezet specifikus szabályzatait és eljárásrendjeit, a promptoknak pontosan és szakszerűen kell megfogalmazniuk az elvárt kontextust, ami a nem szakmai felhasználók számára kihívást jelenthet.

A fizetős, vállalati verziók, például a ChatGPT Enterprise, lényegesen alkalmasabbak a teljes oktatási életciklus támogatására. Ezek a megoldások lehetővé teszik a vállalatspecifikus tudásbázis integrálását, így az MI relevánsabb és kontextusérzékenyebb válaszokat tud adni. Ugyanakkor a költségek is magasabbak, a ChatGPT vállalati verziója például alkalmazottanként havi 20 dolláros előfizetési díjjal jár. E megoldások azonban nem kizárólag a kiberbiztonsági képzésekben alkalmazhatók, hanem számos más belső tudásmenedzsment-feladatban is értékes támogatást nyújthatnak.

A személyre szabott tanulási utak megvalósításához nemcsak pedagógiai, hanem technológiai eszközökre is szükség van. A Cognitive CyTutor nevű rendszer[18] jól illusztrálja, hogyan támogatható a NIST CPLP tanulási ciklusának személyre szabott komponense egy federált, megerősítéses tanulást alkalmazó MI-platformmal. A tanulói viselkedést és előrehaladást elemző modellek lehetővé teszik, hogy a rendszer automatikusan módosítsa az oktatási stratégiákat, miközben a tanulók adatai helyben maradnak.[19]

Egy jó példája a GenMI NIST CPLP keretrendszer szerinti alkalmazásának a SENSAI (scalable, embedded, natural-language system for AI instruction, skálázható, beágyazott, természetes nyelvű mesterségesintelligencia-alapú oktatórendszer), amelyet egy amerikai egyetemen fejlesztettek ki és alkalmaztak. A rendszer bevezetése több ezer hallgató bevonásával történt, és a visszajelzések alapján jelentősen javult a problémamegoldási hatékonyság és a tanulói elégedettség. A rendszer egyetlen tanársegéd költségéhez mérhető anyagi ráfordítással volt működtethető, miközben lényegesen nagyobb tanulói kört szolgált ki.[20]

Mindezek mellett elengedhetetlen számolni a GenMI-rendszerek inherens kockázataival is. A nagy nyelvi modellek, mint a GPT, központi tanítási adatokon alapulnak, amelyek torzításokat hordozhatnak. Emellett a hallucináció jelensége is fennáll, a modell akkor is generálhat magabiztosnak tűnő válaszokat, ha nem rendelkezik megbízható háttértudással az adott kérdésről. Ez különösen kritikus lehet olyan érzékeny területeken, mint a védelemre vagy biztonsági eljárásokra vonatkozó tanácsadás, ahol a téves információ akár súlyos következményekkel is járhat.

A problémák kezelése érdekében elengedhetetlen a rendszeres emberi jóváhagyás, a GenMI által generált tananyagokat és válaszokat az oktatók részéről folyamatosan ellenőrizni kell. Emellett célszerű a belső szabályzatok integrálása az MI-rendszerek működésébe, hogy a generált tartalom illeszkedjen a szervezet elvárásaihoz. Az irányelvek és a határok világos meghatározásával jelentősen csökkenthető a félrevezető vagy téves válaszok esélye. E kockázat csökkentésére szolgálhatnak a Retrieval-Augmented Generation (visszakereséssel támogatott szövegalkotás, RAG-) rendszerek, amelyek a válaszokat megbízható, külső forrásokkal egészítik ki, növelve ezzel a tartalmi pontosságot.

[18] Yogarajah–Hossain 2025.
[19] Yogarajah–Hossain 2025.
[20] Nelson–Doupé–Shoshitaishvili 2025.

Külön figyelmet érdemel a hitelesség és az aktualitás kérdése is. A GenMI-modellek, még ha technológiailag fejlettek is, nem feltétlenül rendelkeznek a legfrissebb információkkal, például a naprakész CVE-adatbázisok tartalmával. Ezért az ilyen típusú kritikus információk értelmezésében és alkalmazásában az MI nem helyettesítheti az emberi szakértelmet.

Emellett fontos szerepe van a felhasználók felkészítésének is, a mesterséges intelligencia akkor működik igazán hatékonyan, ha a munkavállalók képesek strukturált és célzott kérdéseket megfogalmazni. A prompt engineering, azaz a helyes kérdésfeltevés elsajátítása tehát önálló tanulási céllá válik, nemcsak a kiberbiztonsági oktatás, hanem más vállalati folyamatok szempontjából is hasznos készségként. A hatékony promptolás része a célirányos utasításadás, a kontextus világos meghatározása, valamint a kívánt választípusok előzetes rögzítése, ami különösen fontos a kiberbiztonsági fogalmak torzulásmentes átadásához.

Nem elhanyagolható szempont az adatvédelem sem. A vállalati belső képzések során alkalmazott GenMI-megoldások komoly adatbiztonsági kockázatokat hordozhatnak, különösen, ha azok nyílt, internetes platformokon futnak. A Samsung 2023-as esete intő példa: több mérnök véletlenül szenzitív céges adatokat, köztük félvezetőgyártó eszközök forráskódját töltötte fel a ChatGPT-be, ami súlyos adatvédelmi incidenst eredményezett.[21]

További támadástípusok, mint a *prompt injection* vagy a *data poisoning,* szintén torzíthatják a GenMI-rendszerek válaszait, különösen, ha azok nem megfelelően védett vagy nyílt környezetben működnek. 2023 márciusában az OpenAI szolgáltatásában egy technikai hiba következtében a felhasználók idegen beszélgetések címeit és egyes számlázási adatait is láthatták, ami jól mutatja: az MI-rendszerek nem csupán adatkezelők, hanem potenciális adatforrások is a támadók számára.[22]

Ez különösen problémás egy olyan vállalati oktatási platform vonatkozásában, ahol a rendszer a cég belső tudásbázisából merít (például a NIST által említett eljárásrendek és szabályzatok), vagy esetleg a vállalat maga kormányzati entitás, és minősített adatokkal dolgozik. Amennyiben a GenMI-t szabályozás és felügyelet nélkül használjuk, érzékeny üzleti információk, minősített vagy személyes adatok is kiszivároghatnak a generált tartalmakon keresztül.

Számos vállalat felhőszolgáltatásként veszi igénybe a GenMI segítségével támogatott oktatási rendszereket, ami további adatvédelmi aggályokat vet fel. A harmadik fél által üzemeltetett, felhőben tárolt adatok potenciálisan hozzáférhetők illetéktelen személyek vagy támadók számára, ha a felhőszolgáltató kiberbiztonsága sérül. További problémát jelenthet a nyílt API-ok használata, elképzelhető, hogy a vállalati rendszerek túl sok adatot osztanak meg az MI-szolgáltatóval. Továbbá a felhőalapú MI-szolgáltatók is naplózhatják a felhasználói interakciókat, amelyeknek a kiszivárgása esetén bizalmas adatok tömegesen kerülnek az internetre.[23]

Vállalati környezetben kritikus, hogy a generatív MI használata megfeleljen a GDPR és a hamarosan várható EU AI Act előírásainak. Ez főként azt jelenti, hogy a rendszerbe bevitt érzékeny adatok ne kerüljenek illetéktelen kezekbe, és ne használják fel azokat

[21] Petkauskas 2023.
[22] Petkauskas 2023.
[23] Solving the Security Challenges of Retrieval-Augmented Generation (RAG) 2025.

a szolgáltatók más célra, például a nyelvi modell újratanítására a cég engedélye nélkül. A nagy felhős MI-szolgáltatók reagáltak erre, az OpenAI ChatGPT Enterprise verziója kifejezetten hangsúlyozza, hogy „az ügyfél adata az ügyfélé marad, nem használjuk a beszélgetéseket a modell tanítására", továbbá a szolgáltatás SOC 2 megfelelőséggel rendelkezik, és a kommunikáció titkosított.[24] Hasonlóképpen a Microsoft Security Copilot is úgy lett kialakítva, hogy a vállalati adatok kontrollja az ügyfélnél maradjon, a Microsoft nyilatkozata szerint a Security Copilot által feldolgozott adatok nem kerülnek be az OpenAI alapmodell további tanításába, és a rendszer teljesen zárt, megfelel a Microsoft vállalati biztonsági és megfelelőségi követelményeinek.[25] Ezek fontos szempontok például a pénzügyi vagy az egészségügyi szektorban, ahol tiltott lehet külső felhőbe személyes adatot küldeni.

Összességében a GenMI számos ponton képes támogatni a kiberbiztonsági oktatási folyamatokat, az elemzéstől kezdve a személyre szabott tananyagfejlesztésen át a szimulációk és értékelések automatizálásáig. Ugyanakkor mind a tartalmi hitelesség, mind az adatvédelmi kockázatok szempontjából világossá vált, hogy a generatív rendszerek használata csak megfelelő kontroll és szabályozás mellett lehet biztonságos. Különösen érzékeny környezetben, például a kormányzati, a pénzügyi vagy az egészségügyi szektorban, olyan megoldásokra van szükség, amelyek képesek egyesíteni a GenMI előnyeit a belső tudásbázisok védelmével. A felelős mesterséges intelligencia (Responsible AI) elvei, mint az átláthatóság, méltányosság, elszámoltathatóság, szintén alkalmazhatók a GenMI-rendszerekre, különösen érzékeny környezetekben, mint a pénzügyi vagy a kormányzati szektor.

Összegzés

A tanulmány első részében áttekintettem a GenMI szerepét a kibertudatossági képzések oktatástervezési folyamatában. Bemutattam, hogyan képes a GenMI támogatni az oktatókat az oktatási tartalom létrehozásának különböző fázisaiban, az elemzéstől az értékelésig. A tanulásszervezés klasszikus modelljein túl (például ADDIE) a viselkedésbefolyásoló elméleteket is megvizsgáltam, különös tekintettel a tudatosságnövelési célokat szolgáló keretrendszerekre, mint a NIST CPLP modell.

A tanulmány második részében a GenMI-rendszerek által generált kihívásokra kifejlesztett RAG modell bemutatásán felül, már egy kvantitatív kutatás eredményeit ismertettem, amelynek célja, hogy feltérképezze, miként viszonyulnak a felhasználók a generatív mesterséges intelligencia által létrehozott kiberbiztonsági oktatási tartalmakhoz.

[24] Introducing ChatGPT Enterprise 2023.
[25] Jakkal 2023.

Felhasznált irodalom

AI-Generated Phishing Simulations Are a Game Changer for MSPs (2023). ConnectWise Marketplace, 2023. július 11. Online: https://marketplace.connectwise.com/blogs/ai-generated-phishing-simulations-are-a-game-changer-for-msps/

Breacher.ai (2025): *Deepfake Archives.* Online: https://breacher.ai/solutions/deepfake-simulation/

Brito, Fernando et al. (2025): Enhancing Cybersecurity Education With Artificial Intelligence Content. *In Proceedings of the 56th ACM Technical Symposium on Computer Science Education 1.* New York: Association for Computing Machinery, 158–164. Online: https://doi.org/10.1145/3641554.3701958

CVE [é. n.]: Frequently Asked Questions (FAQs). Online: https://www.cve.org/ResourcesSupport/FAQs

EU Artificial Intelligence Act. Online: https://artificialintelligenceact.eu/ai-act-explorer/

European Union Agency for Cybersecurity (2024): *ENISA Threat Landscape 2024: July 2023 to June 2024.* Online: https://data.europa.eu/doi/10.2824/0710888

Jakkal, Vasu (2023): Introducing Microsoft Security Copilot: Empowering Defenders at the Speed of AI. *Official Microsoft Blog,* 2023. március 28. Online: https://blogs.microsoft.com/blog/2023/03/28/introducing-microsoft-security-copilot-empowering-defenders-at-the-speed-of-ai/

Merritt, Marian et al. (2024): *Building a Cybersecurity and Privacy Learning Program. NIST SP 800-50r1.* Gaithersburg: National Institute of Standards and Technology. Online: https://doi.org/10.6028/NIST.SP.800-50r1

Mouton, Francios et al. (2014): Social Engineering Attack Framework. In *2014 Information Security for South Africa.* 1–9. Online: https://doi.org/10.1109/ISSA.2014.6950510

Nelson, Connor – Doupé, Adam – Shoshitaishvili, Yan (2025). SENSAI: Large Language Models as Applied Cybersecurity Tutors. In *Proceedings of the 56th ACM Technical Symposium on Computer Science Education 1.* New York: Association for Computing Machinery, 833–839. Online: https://doi.org/10.1145/3641554.3701801

New Jersey Cybersecurity & Communications Integration Cell [é. n.]: *ChatGPT and Its Impact on Cybersecurity.* Online: https://www.cyber.nj.gov/guidance-and-best-practices/artificial-intelligence/chatgpt-and-its-impact-on-cybersecurity

Introducing ChatGPT Enterprise. (2023). *OpenAI,* 2023. augusztus 28. Online: https://openai.com/index/introducing-chatgpt-enterprise/

Petkauskas, Vilius (2023): Lessons Learned from ChatGPT's Samsung Leak. *Cybernews,* 2023. május 8. Online: https://cybernews.com/security/chatgpt-samsung-leak-explained-lessons/

Proofpoint (2023): *2023 Security Awareness Study. How Effective Are Your Awareness Programs – and Do Your Employees Agree?* [H. n.]: Information Security Media Group. Online: https://www.proofpoint.com/us/resources/analyst-reports/ismg-security-awareness-training-perception-study

Shaw, R. S. et al. (2009): The Impact of Information Richness on Information Security Awareness Training Effectiveness. *Computers & Education,* 52(1), 92–100. Online: https://doi.org/10.1016/j.compedu.2008.06.011

SlashNext (2023): *State of Phishing.* Online: https://www.prnewswire.com/news-releases/slashnexts-2023-state-of-phishing-report-reveals-a-1-265-increase-in-phishing-emails-since-the-launch-of-chatgpt-in-november-2022--signaling-a-new-era-of-cybercrime-fueled-by-generative-ai-301971557.html

Solving the Security Challenges of Retrieval-Augmented Generation (RAG) (2025). *Polymer,* 2025. január 31. Online: https://www.polymerhq.io/blog/ai/solving-the-security-challenges-of-retrieval-augmented-generation-rag/

Verizon (2024): *2023 Data Breach Investigations Report: Frequency and Cost of Social Engineering Attacks Skyrocket.* Online: https://www.verizon.com/business/resources/reports/2023-data-breach-investigations-report-dbir.pdf

Yogarajah, Jesudhas – Hossain, Gahangir (2025): The Concept of Cognitive CyTutor Utilizing Federated Learning for Cybersecurity Education. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC),* 681–690. Online: https://doi.org/10.1109/CCWC62904.2025.10903809