

37. Országos Tudományos Diákköri Konferencia
Had- és Rendészettudományi Szekció
Budapest, 2025. április 22–24.
Nemzeti Közsolgálati Egyetem
Hadtudományi és Honvédtisztképző Kar

E-MAIL BIZTONSÁG NÖVELÉSÉNEK LEHETŐSÉGEI TER- MÉSZETES NYELVFELDOLGOZÁSSAL

Jelige: *Krabb*

Pályamunka lezárásának időpontja: 2023. 11. 09.

2025.

TARTALOMJEGYZÉK

Tartalom

1.ABSZTRAKT.....	3
2.ACM KLASSZIFIKÁCIÓ.....	5
3.BEVEZETÉS.....	6
3.1. BEVEZETŐ GONDOLATOK	6
3.2. MOTIVÁCIÓ	6
3.3. EDDIGI KUTATÁSI EREDMÉNYEK ISMERTETÉSE.....	6
3.4. SOCIAL ENGINEERING TERMINOLÓGIÁK	8
3.5. TECHNOLÓGIAI ALAPOK	10
4.A TÉMA AKTUALITÁSA.....	13
4.1. TÉMA AKTUALITÁSA ONLINE FORRÁSOK ALAPJÁN	13
4.2. TÉMA AKTUALITÁSA SZAKÉRTŐI MEGKÉRDEZÉS ALAPJÁN	13
4.3. TUDOMÁNYREPREZENTÁCIÓ MTMT-BEN 2023 TAVASZ	13
4.4. TUDOMÁNYREPREZENTÁCIÓ MTMT-BEN 2023 ŐSZ.....	14
4.5. TUDOMÁNYREPREZENTÁCIÓ IEEEEXPLORE-BAN 2023 TAVASZ	15
4.6. TUDOMÁNYREPREZENTÁCIÓ IEEEEXPLORE-BAN 2023 ŐSZ.....	16
4.7. TUDOMÁNYREPREZENTÁCIÓ JSTORE-BAN 2023 ŐSZ	17
4.8. TUDOMÁNYREPREZENTÁCIÓ SCOPUS-BAN 2023 ŐSZ.....	18
4.9. SCOPUS AI – TÉMA FELDERÍTÉSE.....	19
5.PROBLÉMÁK MEGFOGALMAZÁSA.....	21
5.1. ÉSZREVÉTELEK.....	21
5.2. HIPOTÉZISEK	22
6.SZEKUNDER KUTATÁS.....	23
6.1. IRODALOMKUTATÁS EREDMÉNYEI.....	23
6.1.1. Technológiai 11.05.....	23
6.1.2. Nyelvi	24
7.PRIMER KUTATÁS.....	26
7.1. SZAKÉRTŐI MEGKÉRDEZÉSEK	26
7.2. LAKOSSÁGI MEGKÉRDEZÉS (KÉRDŐÍVVEL).....	27
7.2.1. Módszertan.....	27
7.2.2. Eredmények.....	27
7.3. KORÁBBI MÉRÉSEK BEMUTATÁSA	29
7.3.1. 1. Modell	30
7.3.2. 2. Modell	31
7.3.3. 3. Modell	31
7.3.4. További modellek sikertelensége.....	31
7.4. EMAIL EXPORT-IMPORT ELŐRE PONTOZÁS	32
7.5. ÚJ TÁMADÁSI OSZTÁLYOK LÉTREHOZÁSA	33
7.6. MÉRÉSEK SZABADON HASZNÁLHATÓ NAGY NYELVI MODELLEKKEL	33
7.6.1. Bard.....	33
7.6.2. ChatGPT.....	34
7.7. SAJÁT MÓDSZER KIDOLGOZÁSA	34
8.KONKLÚZIÓ - TÉZISEK.....	36
9.MUNKAMEGOSZTÁS.....	38
10.IRODALJEGYZÉK.....	39

1. ABSZTRAKT

1. Probléma megfogalmazása

Nagyon sok olyan műszaki/informatikai megoldással találkoztunk, amik nem voltak képesek megfelelő hatékonysággal detektálni és reagálni a Social Engineering támadásokra. Pedig amennyiben megvizsgáljuk a folyamatot egy támadó által küldött kártékony emailt sok helyen meg lehetne fogni típustól függően. Példaképpen kipróbáltunk egy átlagos védelemmel rendelkező üzleti laptopra (reaktív vírusirtó, tűzfal, Email Biztonsági Megoldás a levelező szerver előtt) egy kártékony emailt küldeni és meglepődve tapasztaltuk, hogy semmilyen riasztás nem jelzett.

2. Hipotézisek

- H1: A jelenleg alkalmazott Email Biztonsági Megoldások nem képesek kiszűrni a célzott Social Engineering támadásokat.
- H2: A Phising és Smishing támadások hasonló digitális lenyomattal rendelkeznek, emiatt a védelmi fejlesztéseknél a szövegtörzsek ekvivalensnek tekinthetők.
- H3: A jelenleg alkalmazott Email Biztonsági Megoldások természetes nyelvfeldolgozással ötvözve (NLP) képesek lehetnek olyan tipikus Social Engineering jeleket felismerni az e-mailekben (rávétel módszere, bizalom keltése, sürgetés stb...) amelyekkel javítható lenne az Email Biztonsági Megoldások hatékonysága.
- H4: Meg lehet határozni enumerációval egy generális neurális háló esetén, hogy milyen Social Engineering támadásokat ismer fel a szűrés jobban egy hasonló méretű tanítóhalmaz esetén.
- H5: A H3-as pontban meghatározott Social Engineering támadásokat tovább bontva meghatározhatóak azok az elemek az emailekben, amiket nem képesek jelenleg a mélytanuló algoritmusunk felismerni.
- H6: Az Email Biztonsági Megoldások Social Engineering támadások elleni hatékonysága javítható azzal, hogyha a konvencionális és a mesterséges intelligencia alapú védelmi megoldások mellé rendelkezik még egy komplexebb kockázatkezelő modullal.

3. Primer és szekunder kutatás

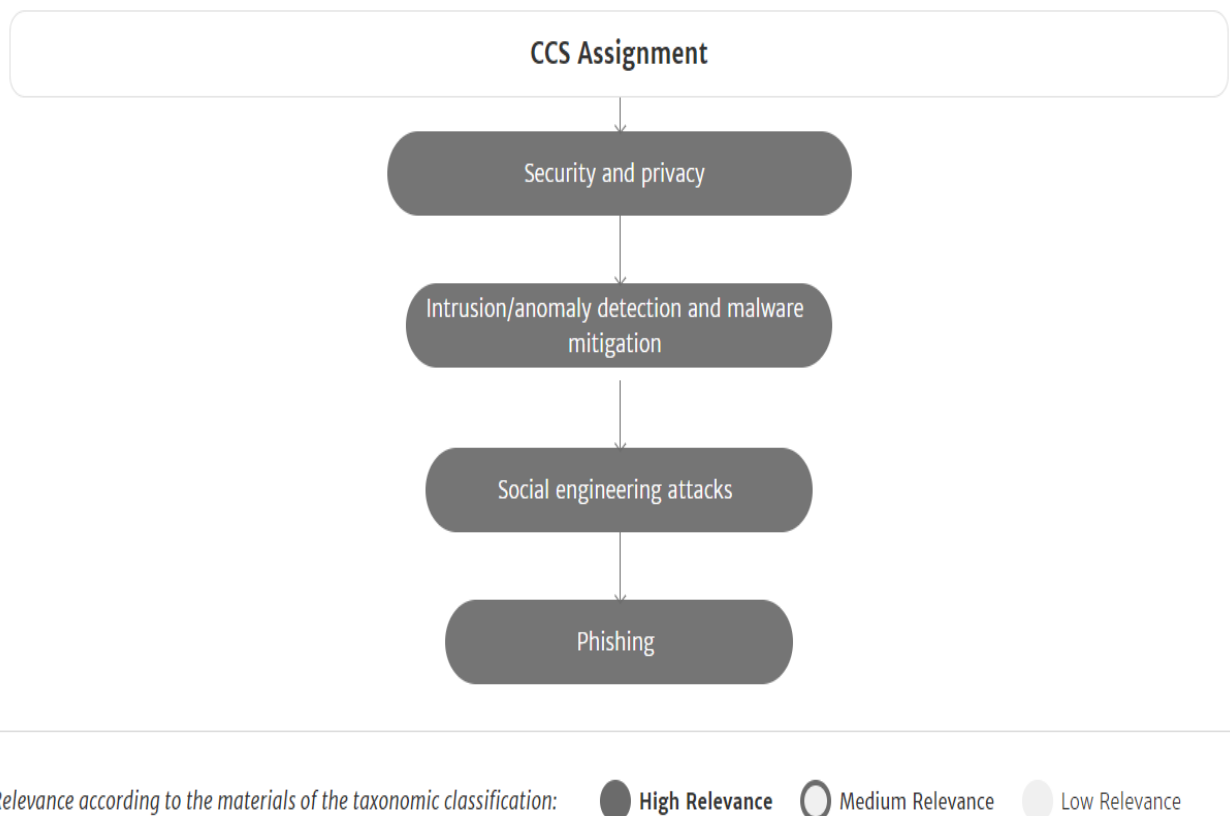
Szekunder kutatásunkban megismertük a szakirodalomban a Social Engineering mélytanulás alapú detektálásával kapcsolatos tudományos folyóirat cikkeket és a technológiák pontos és mély megértése után kidolgoztunk egy olyan elméleti módszert mellyel ki tudnánk nyerni a

mély tanuló rendszerből, hogy mi miatt döntött úgy, hogy gyanúsna vagy úgy, hogy biztonságosnak ítéli.

Primer kutatásunkként interjúkat folytattunk a téma aktualitásáról és az interjúk során megpróbáltuk feltérképezni a jelenlegi hazai email biztonsági helyzetet. Valamint a számunkra gyanúsan magas pontosságú modelleken végeztünk méréseket, hogy megbizonyosodjunk az eredmények pontosságáról.

2. ACM KLASSZIFIKÁCIÓ

Az Országos Tudományos Diákköri (OTDT) Tanács javaslata alapján elkészítettük az ACM (Association for Computing Machinery) a számítási klasszifikációs rendszerben (CCS - Computing Classification System) a kutatásunk polihierarchikus ontológiáját.



3. BEVEZETÉS

3.1. BEVEZETŐ GONDOLATOK

Az IPAR 4.0 közepén és az IPAR 5.0 közeledtével egyre rohamosabban fejlődik az információ technológia, de vajon az általános biztonságunk is olyan exponenciális mértékben fejlődött mint, ahogy a technológiák egy új ipari forradalom kezdetén?

Sajnos erre az esetek nagy részében nem tud válaszolni a tudomány. Rengeteg szakirodalom, valamint saját kutatásaink és tapasztalataink is bizonyítják azt, hogy nem vagyunk képesek a biztonság macska egér játékában az egér sebességét megközelíteni. Különösen olyan esetekben, amikor a támadók már nem csak informatikai támadási vektorokat alkalmaznak, hanem pszichológiával ötvözik támadásaikat, akkor az eddig csak informatikai biztonságra fókuszáló védekező kollégáknak esélyük sem lesz arra, hogy felvegyék a harcot a modern szemléletmódú támadásokkal. [23]

A gépi tanulási technológiák rohamos fejlődése, különösen a nagy nyelvi modellek (pl.: OpenAI ChatGPT, Google Bard) áttörést jelentettek a természetes nyelvfeldolgozás (NLP) tudományterületen, ami most a biztonság Blue Team oldalán számunkra is hasznos lehet, hogy megpróbáljuk felvenni a harcot az új képességekkel felvértezett kiber-támadókkal szemben [23]

3.2. MOTIVÁCIÓ

Egyszerű tesztjeinkből hamar kiderült, hogy nem tudjuk műszakilag magas hatékonysággal kivédeni a Social Engineering támadásokat és csakis kizárólag a humán alapú biztonságtudatosági képzéseket tudjuk erre a célra használni. Úgy gondoltuk, hogy mérnök hallgatóként elkezdünk megoldást találni erre a kihívásra a mai legmodernebb nyelvi modellek, a pszichológia, valamint a nyelvészet ötvöztetésével.

3.3. EDDIGI KUTATÁSI EREDMÉNYEK ISMERTETÉSE

- Krasnyánszki Brúnó: Gimnazista koromban a Kutató Diákok Országos Szövetségénél végeztem kutatást Social Engineering témában, ahol azt vizsgáltam milyen típusú támadások hogyan érintik a KKV-kat és az állami szereplőket. Valamint, hogyan lehet az ilyen támadások ellen védekezni. A fő konklúziója ennek a majdnem egy éves kutatásnak az lett, hogy jelenleg biztonságtudatosági oktatásokkal és ezen oktatások hatékonyságának visszamérésével lehet egyedül megküzdenni a Social Engineering támadásokkal

szemben. A Kutató Diákok Országos Szövetsége által szervezett 'TUDOMÁNYOS DIÁKKÖRÖK XXII. KÁRPÁT-MEDENCEI KONFERENCIÁJA TUDOK 2022 Műszaki- és reáltudományi Tematikus Konferencia'-án, a 'TUDOMÁNYOS DIÁKKÖRÖK XXII. KÁRPÁT-MEDENCEI KONFERENCIÁJA (TUDOK)' c. konferencián, valamint a 'X. KUTDIÁK TUDOMÁNYOS HÉTVÉGE - XX. KutDiák Tudományos Esszépályázat'-on volt alkalmam bemutatni eredményeimet 2022 tavaszán. Ebből a kutatásból témavezetőm felkérésére cikket írtam a Biztonságtudományi Szemlében' Hogyan változtatta meg a XXI-ik századi kibervédelmet a Social Engineering?' címmel, ahol összefoglaltam eddigi kutatási eredményeimet a nyilvánosság számára is. 2022-ban mikor egyetemista lettem tovább folytattam kutatásomat, immár olyan műszaki megoldás kutatását tűztem ki célnak, amit az eddigi kutatásom szerint még nem létezett vagy legalábbis nem terjedt el. Az Óbudai Egyetem 57. Tudományos Diákköri Konferenciáján prezentáltam továbbá' Mesterséges intelligencia alapú email biztonsági megoldások alacsony hatékonyságának gyökér okainak vizsgálata Social Engineering támadások során' címmel. Ez a TDK dolgozat alapozta meg a mostani kutatásomat, így a továbbiakban részben vagy egészben azon részeket, amik még most is relevánsak idézni fogok belőle. Mivel korábbi TDK munkámból is vettem át részeket ezért ezeket minden bekezdés végén jelöltem az önplágium elkerülése érdekében. Az irodalomjegyzékben 23-as számmal [23] jelölt korábbi TDK dolgozatom.

- Csizmadia Réka: Motivációm a jelenlegi kutatáshoz és a TDK-dolgozat megírásához több tényezőtől fakad. Korábbi tanulmányaimat az ELTE Bölcsészettudományi Karán végeztem, ahol nyelvészeti és társadalompszichológiai kutatásokat folytattam. Ezek a kutatások mélyebb megértést nyújtottak számomra az emberi kommunikáció és viselkedés mechanizmusairól, amely szorosan kapcsolódik a társadalmi mérnöki támadások jellegéhez. Az első egyetemi félévem alatt már részt vettem TDK-munkában, amely lehetőséget adott arra, hogy kutatási képességeimet és módszertanomat fejlesszem. Brúnóval való közös munkánk további inspirációt adott számomra, mivel az együttműködésünk során tapasztaltak tovább mélyítették érdeklődésemet az információbiztonság és a társadalmi mérnökség területén. Fordítói tapasztalatom, amely tolmács munkából és szakfordításokból származik, különösen hasznosnak bizonyult a szakirodalom feldolgozásában és a kutatás nemzetközi kontextusba helyezésében. Mindezek az előzmények és tapasztalatok motiválnak arra, hogy hozzájáruljak az email biztonsági megoldások fejlesztéséhez, és új perspektívákat nyújtsak a társadalmi mérnöki támadások elleni védekezés terén.

3.4. SOCIAL ENGINEERING TERMINOLÓGIÁK

A Social Engineering kifejezés definíciója kapcsán egy jól ismert szakértő Christopher Hadnagy¹ „The Art of Human Hacking” (ford.: Az Emberek Meghekkelésének Művészete) című könyvében leírt definícióját fogadjuk el dolgozaunkban, amely a következő: „A Social Engineering a befolyásolás és 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.” [3] Mivel jelenlegi ismereteink szerint nincs általánosan elfogadott fordítása a magyar szaknyelvben (legtöbbször pszichológiai manipulációt tartalmazó támadásként szokták használni), ezért az angol Social Engineering (röviden SE) kifejezést fogom a továbbiakban használni. [23]

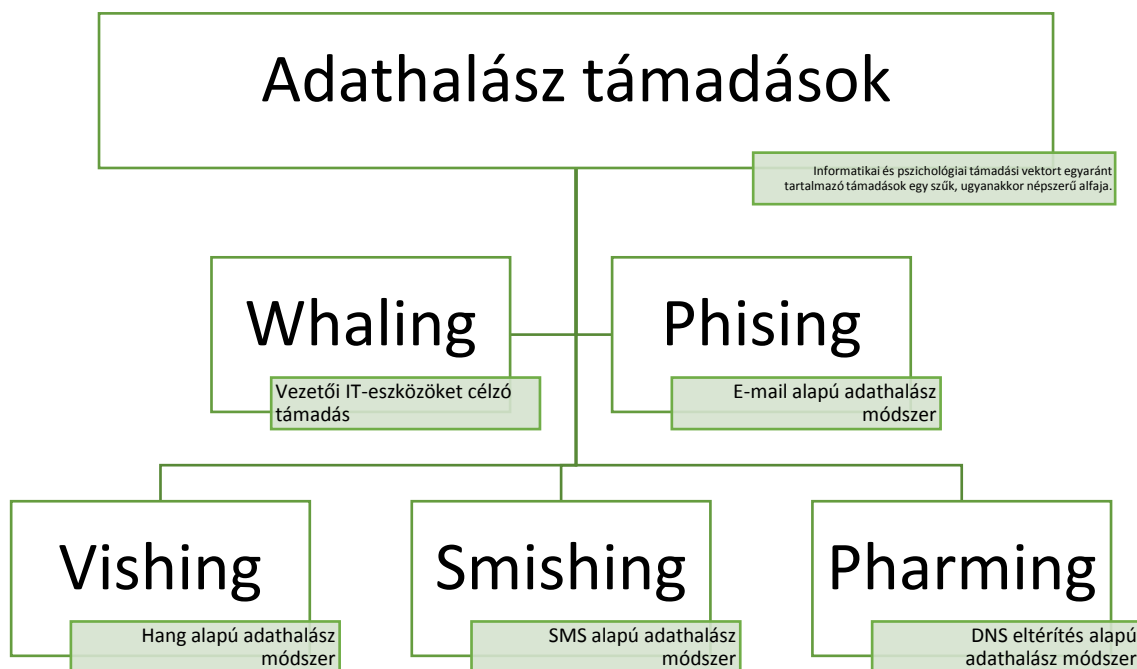
A Social Engineering támadások tárháza nagyon széles, emiatt a TDK dolgozatunkban nem kívánunk belemenni az összes támadási típus ismertetésébe. A támadásokat a hazai és nemzetközi szakirodalom széles körben részletezi. Nemzetközi szakirodalom kapcsán Christopher Hadnagy fent említett könyve, hazai szakirodalom kapcsán pedig Tóth Tamás folyóirat cikke [5] dolgozza fel a Social Engineering támadásokat, valamint is korábbi szerzői saját kutatás során [6] már részletezésre került. Amivel első körben foglalkozni szeretnénk most TDK dolgozatunk azok az adathalász támadások. Az adathalász támadások során a támadó valamilyen (jellemzően infokommunikációs csatornákon, de volt példa papír alapú adathalászatra is [7]) módon meggyőzi az áldozatot, hogy adja meg az adatait valamilyen célból olyan legendával, mely nem kelt gyanút az áldozatban. Példa lehet erre a manapság népszerű Netflix-es vagy GLS-es csalások emailben és SMS-ben egyaránt.[23]

¹ Christopher James Hadnagy egy amerikai író és információtechnológiai biztonsági tanácsadó. Az információbiztonság területén végzett társadalmi mérnöki munkájáról ismert.



1. ábra, szerző által készített, példa a Netflixes Smishingre

Az adathalász támadásokat Tóth Tamás módszere alapján klasszifikáltuk [5, pp 95-96] [23] a következő csoportokra:



2. ábra szerző által készített

Ezen belül is első sorban a Phishing-el, másodsorban pedig a Smishing-el foglalkoznánk a későbbiekben. Ezeket hasonlóan a Social Engineering-hez angolul fogjuk használni, mivel jelenleg nincs elterjedt és egységes magyar nyelvű kifejezés, ami egy szóban lefedné. [23]

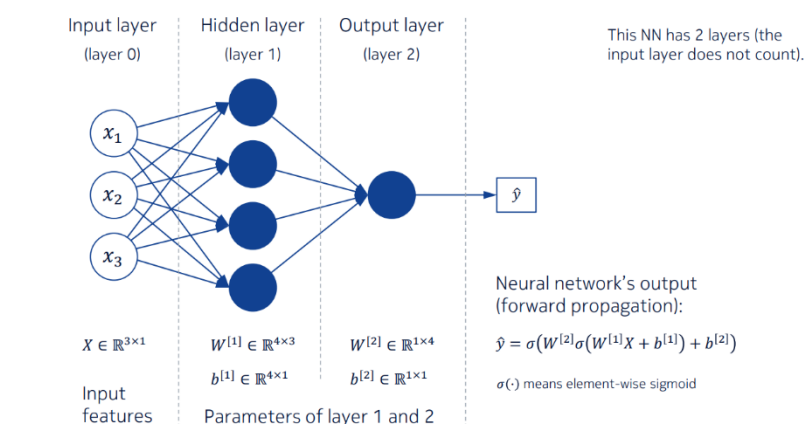
3.5. TECHNOLÓGIAI ALAPOK

A mély tanulás története az 1940-es évekig nyúlik vissza, amikor Warren McCulloch és Walter Pitts alkották meg az első mesterséges neuronhálózat elméletét. Azonban a módszer korai éveiben a számítási kapacitás korlátozott volt, és a kutatások lassan haladtak. Az igazi ugrást az 1980-as években tapasztaltuk, amikor a visszaterjesztéses neurális hálózatok (RNN) és a konvolúciós neurális hálózatok (CNN) megjelentek. [21]

Az igazi forradalmat azonban a 21. század elején hozta el a mély tanulásnak nevezett terület, különösen a mély neurális hálózatok terjedésével. A GPU-k és nagyobb adathalmazok elérhetősége lehetővé tette a mély neurális hálózatok hatékonyabb tanítását. 2012-ben például a ImageNet versenyen AlexNet nevű mély neurális hálózat nyert, jelentős teljesítményjavulást mutatva a klasszikus gépi tanulási módszerekhez képest [22]

A mai gépi tanulás alapját a neurális hálók adják. Ezek a hálók neuronokból állnak. Rendelkeznek bemeneti, kimeneti és úgynevezett rejtett rétegekkel (ez az összes réteg, ami a bemeneti és a kimeneti között van).

Neural network representation



13

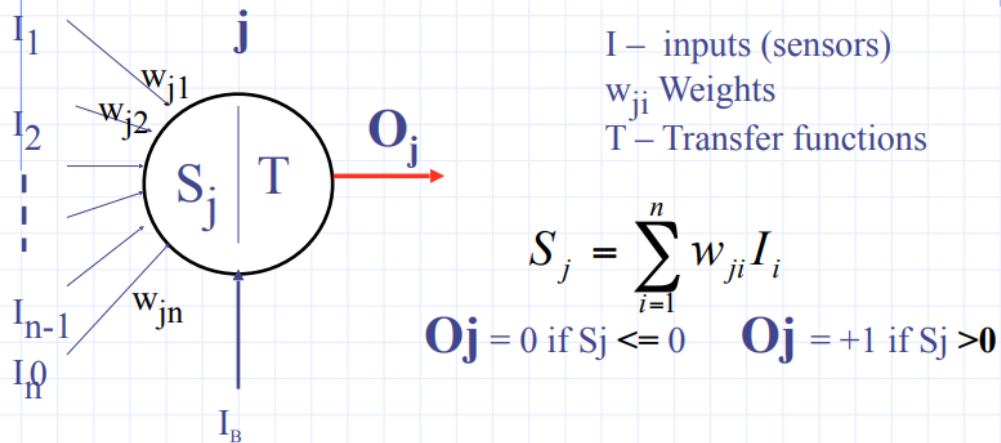
3. Képfelirat és forrás NOKIA [24]

A neuronok általános felépítése az lenti képen látható. Egy neuronnak általánosan vannak bemenetei, amihez súly érték van párosítva. Azért van erre szükség, hogy különböző súlyyal lehessen figyelembe venni a különböző bemeneti értékeket. Ezáltal könnyen módosíthatóvá válnak a bemeneti értékek.

McCulloch and Pitts Formal Neurons

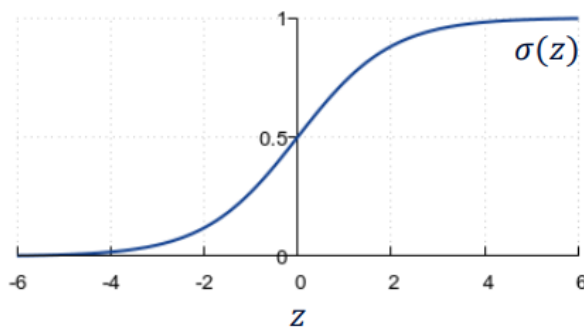
W. Mc Culloch és W. Pitts (1943)

„Firstly considered the brain as a computing device”



4. Képfelirat és forrás Óbudai Egyetem, dr. Kutor László professzor előadása, 2022 [25]

Az átviteli vagy aktivációs függvények azok, amik a végső műveleteket elvégzik a neuronokban. Ezek közül a 3 legnépszerűbb a Sigmoid átviteli függvény:

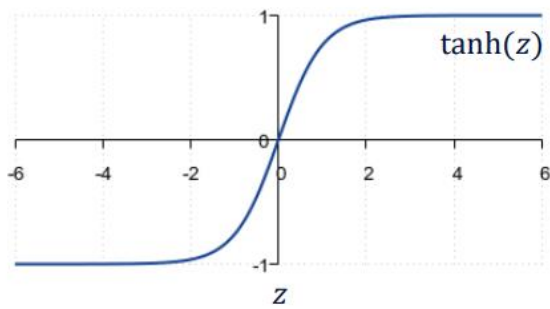


3. ábra Sigmoid átviteli függvény [24]

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

A sigmoidot leginkább a kimeneti rétegben alkalmazzák bináris klasszifikáció esetén.

A Tahn függvény szinte mindig jobb eredményt hoz a sigmoidnál és gyakran a rejtett rétegben használják:

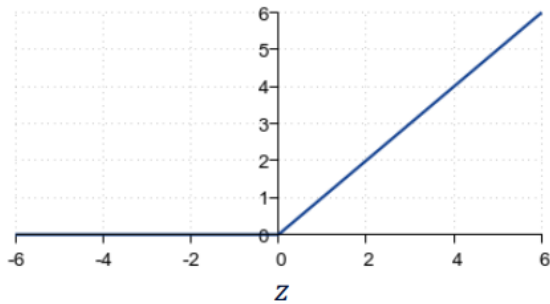


4. ábra Tahn átviteli függvény [24]

$$\sigma(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

[24]

Gyakori választás, főleg hatékonysági okokból kifolyólag a ReLU (rectified linear unit) átviteli függvény. Ez felgyorsítja a neurális hálót, mivel kevésbé számítás igényes, mint a korábbi kettő.



5. ábra ReLU átviteli függvény

$$\sigma(z) = \max(0, x)$$

[24]

4. A TÉMA AKTUALITÁSA

4.1. TÉMA AKTUALITÁSA ONLINE FORRÁSOK ALAPJÁN

A Social Engineering alapú támadások évről évre nagyobb számban jelennek meg és a TerraNova Security szerint a 2022-es évben az azonosított phishing támadások 96%-a [2] emailen keresztül érte a szervezeteket, ezért különösen időszerűnek tartottam ezen témával való foglalkozást. [23]

Továbbá megvizsgáltuk az egyik legnépszerűbb megoldást, amit phishing támadások ellen tudunk alkalmazni, ez a Cisco Talos Intelligence Group által működtetett PhishTank [26] ami korábban már ismert, többnyire nagy tömeg számára kiküldött adathalász oldalakat tud azonosítani, amely tovább igazolja azon hipotézisünket, hogy jelenleg a technológia nem képes kiszűrni a célzott támadásokat, pusztán a nagy tömegeket elérő támadás hullámokat.[23]

4.2. TÉMA AKTUALITÁSA SZAKÉRTŐI MEGKÉRDEZÉS ALAPJÁN

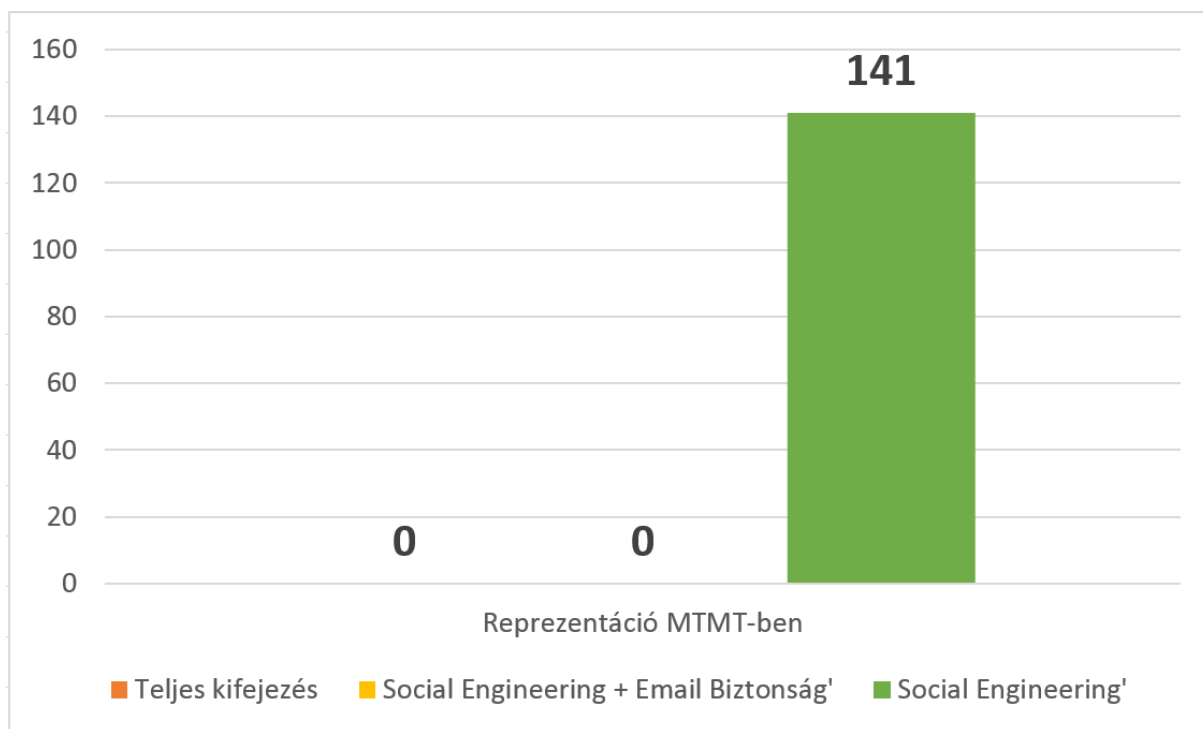
Továbbá személyesen is volt alkalmunk rövid interjú erejéig megkérdezni egy ISO 27001-es² auditort, egy információbiztonsági vezetőt, egy információbiztonsági tanácsadót, egy információbiztonsági szakértőt, egy telekommunikációs cég üzletágvezetőjét és egy szintén telekommunikációs cég vezető kutató mérnökét a témáról és mind a hat interjú alatt azt a visszajelzést kaptam, hogy az általuk működtetett email biztonsági megoldások meglehetősen primitív elven működnek és nem képesek feltétlen az ígért számokat hozni. Megerősítettek benne, hogy érdemes foglalkoznom a témával. Interjúnk során kiemelnénk, hogy a megkérdezettek között vegyesen voltak a versenyszférában és a közigazgatásban egyaránt dolgozók, ezzel növelve a kis mintánk reprezentációját.[23]

4.3. TUDOMÁNYREPREZENTÁCIÓ MTMT-BEN 2023 TAVASZ

A Magyar Tudományos Művek Tárának adatbázisában 2023 májusában a kulcsszavakra való keresést végeztünk. Az alábbi kulcsszavakra kerestem együttesen: "Social Engineering", "Email biztonság" és "mély tanulás". Sajnálatos módon nem sikerült találatokat találnunk. Az "Social Engineering" és "email" kulcsszavakra is eredménytelen volt a keresés. Azonban a "Social Engineering" kulcsszóra már jelentős mennyiségű találat érkezett, összesen 141 darab. Ezek között tudományos publikációk és konferencia előadások is szerepeltek szép számban. A "Email biztonság" és "mély tanulás" kulcsszavak keresése is nullás eredményt hozott az MTMT

² A Nemzetközi Szabványügyi Szervezet információ biztonsági menedzsment szabványa

rendszerében. Ezek alapján megállapítottam, hogy Magyarországon ezek a témák 2023 elején kevésbé kutatott területek. [23]



6. ábra TUDOMÁNYREPREZENTÁCIÓ MTMT-BEN 2023 TAVASZ | Szerző által szerkesztett

4.4. TUDOMÁNYREPREZENTÁCIÓ MTMT-BEN 2023 ŐSZ

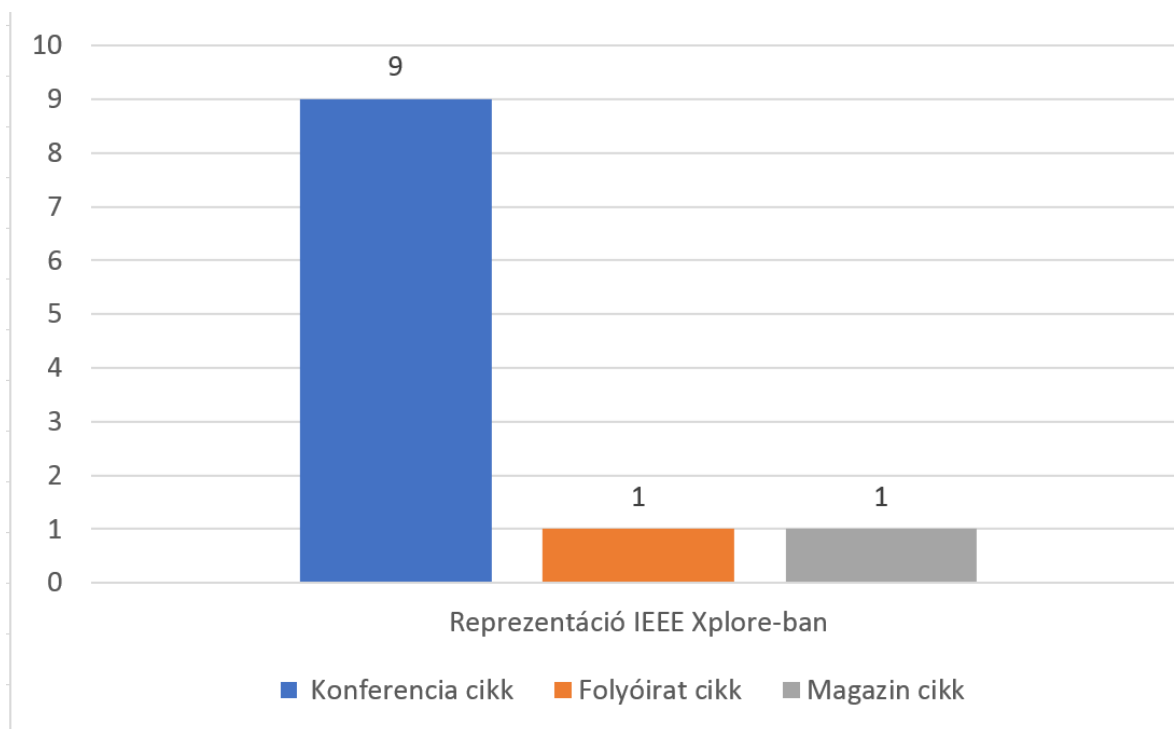
2023 novemberében újra elvégeztük a kulcsszavakra való keresést. Itt is ugyanúgy végeztük el a kulcsszavas keresést. Az alábbi kulcsszavakra kerestem együttesen: "Social Engineering", "Email biztonság" és "mély tanulás". Sajnálatos módon nem sikerült találatokat hoznia felhöz-nia a keresőmodulnak. Az "Social Engineering" és "email" kulcsszavakra is eredménytelen volt a keresés. Azonban a "Social Engineering" kulcsszóra már jelentős mennyiségű találat érkezett, 149 darab. Ezek között tudományos publikációk és konferencia előadások is szerepeltek szép számban. A "Email biztonság" és "mély tanulás" kulcsszavak is nullás eredményt hoztak az MTMT rendszerében. Ez alapján levontuk azt a konklúziót, hogy Magyarországon kis mérték-ben kutatták ezt a témát a korábbi trendekkel szemben.



7. ábra TUDOMÁNYREPREZENTÁCIÓ MTMT-BEN 2023 ŐSZ | Szerző által szerkesztett

4.5. TUDOMÁNYREPREZENTÁCIÓ IEEEEXPLORE-BAN 2023 TAVASZ

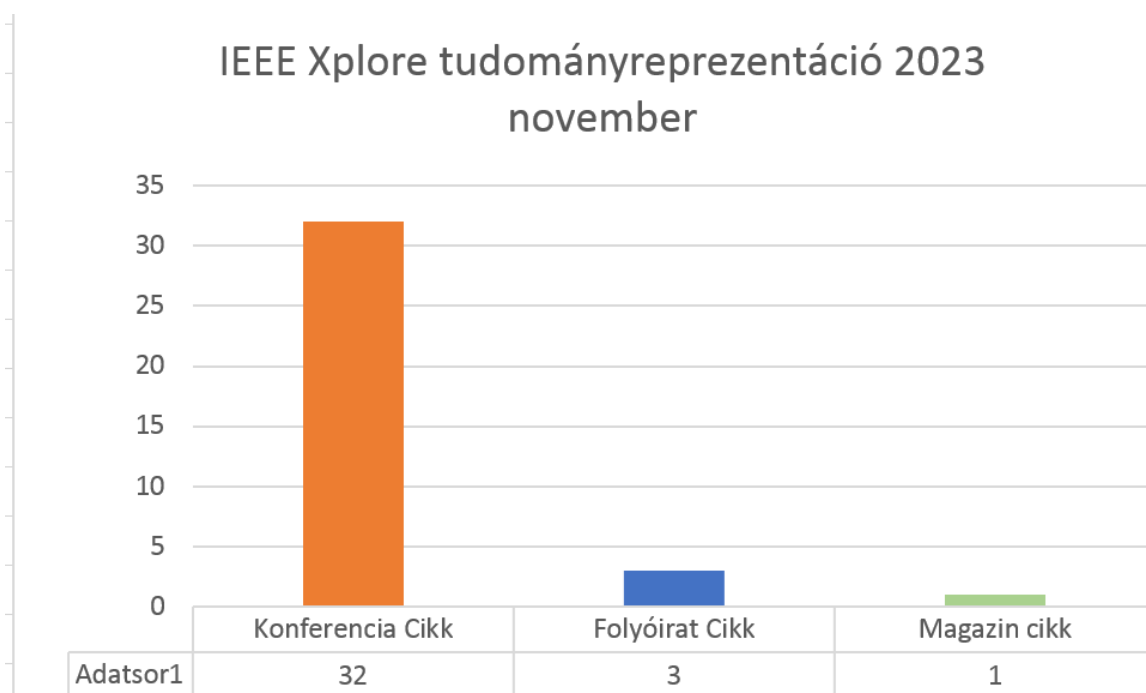
Az IEEE Xplore adatbázisban is megvizsgáltuk a téma reprezentációját. Mind három kifejezést tartalmazó keresést indítottunk a fejlett keresési módban (advanced search: ("All Metadata":Social Engineering) AND ("All Metadata":Email security) AND ("All Metadata":Deep Learning))melynek során 11 találatot kaptunk vissza az IEEE Xplore kereséséből, ebből 9 konferencia cikk, 1 folyóirat cikk és egy magazin cikk eloszlásban voltak. A fent említett találatok közül egy 2019-es, egy-egy 2020 és 2021-es, a többi pedig 2022-ben készült. Ezáltal arra a megállapításra jutottunk, hogy a tématerület nemzetközi kutatása sem jelentős és csak az elmúlt években kezdtek el vele foglalkozni. [23]



8. ábra TUDOMÁNYREPREZENTÁCIÓ IEEEEXPLORE-BAN 2023 TAVASZ /Szerző által szerkesztett

4.6. TUDOMÁNYREPREZENTÁCIÓ IEEEEXPLORE-BAN 2023 ŐSZ

2023 novemberében szintén megvizsgáltuk a téma reprezentációját az IEEE Xplore adatbázisban. Az előző kereséssel equivalensen mind a három kifejezést tartalmazó keresést indítottunk a fejlett keresési módban (advanced search: ("All Metadata":Social Engineering) AND ("All Metadata":Email security) AND ("All Metadata":Deep Learning)) melynek során 36 találatot kaptunk. Ez a tavaszi 11-nél 25-el több. Ebből arra a következtetésre jutottunk, hogy a téma elkezdett a nemzetközi mérnöki adatbázisban felkapott lenni. Releváns vele most foglalkozni és jelenleg aktívan kutatott terület.

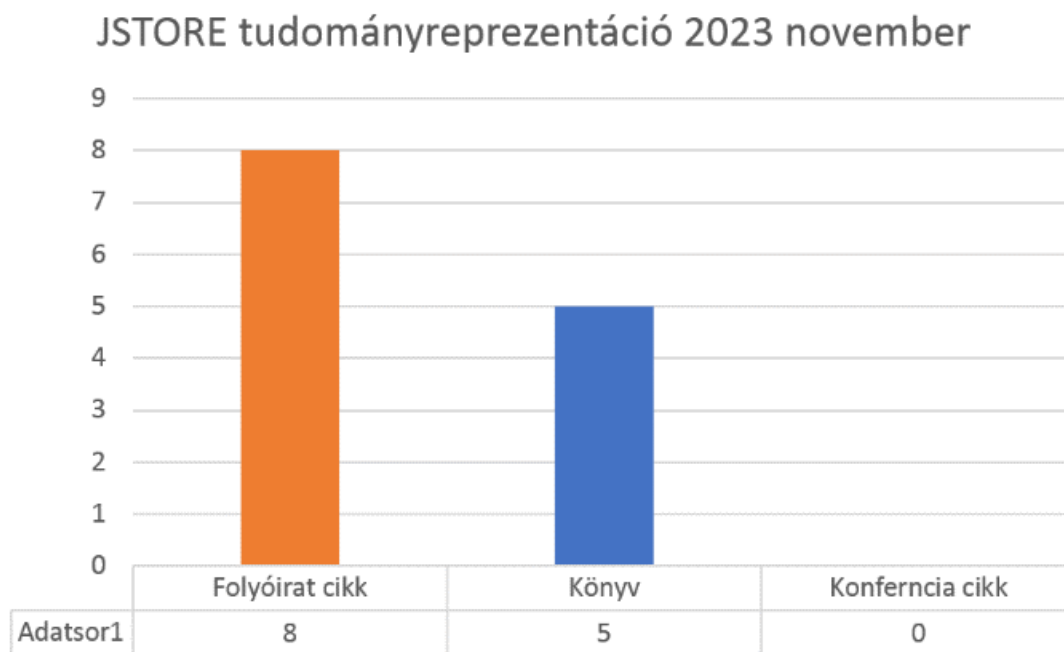


9. ábra TUDOMÁNYREPREZENTÁCIÓ IEEEEXPLORE-BAN 2023 ŐSZ | Szerző által szerkesztett

4.7. TUDOMÁNYREPREZENTÁCIÓ JSTORE-BAN 2023 ŐSZ

2023 novemberben tudományrepresentációt megvizsgáltuk a Journal Storage (JSTORE)³ adatbázisban is. Itt is a fejlett keresési módban az előzőekben leírt 3 kifejezésre rákeresve a grafikonon szereplő adatokat kaptuk

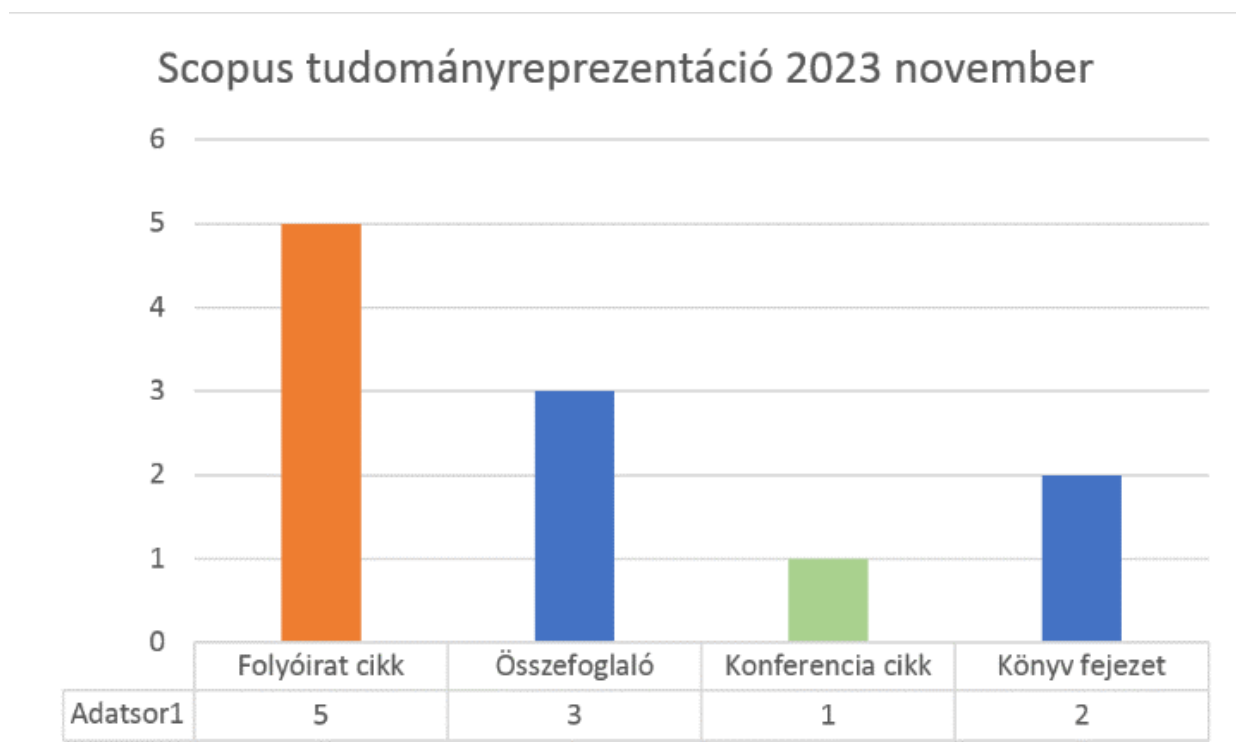
³ Ugyan az Óbudai Egyetem nem rendelkezik még JSTORE előfizetéssel, de a Nemzeti Közszolgálati Egyetem könyvtára a segítségünkre volt kutatásunk során, amit ezúton is köszönünk nekik.



10. ábra TUDOMÁNYREPREZENTÁCIÓ JSTORE-BAN 2023 ŐSZ | Szerző által szerkesztett

4.8. TUDOMÁNYREPREZENTÁCIÓ SCOPUS-BAN 2023 ŐSZ

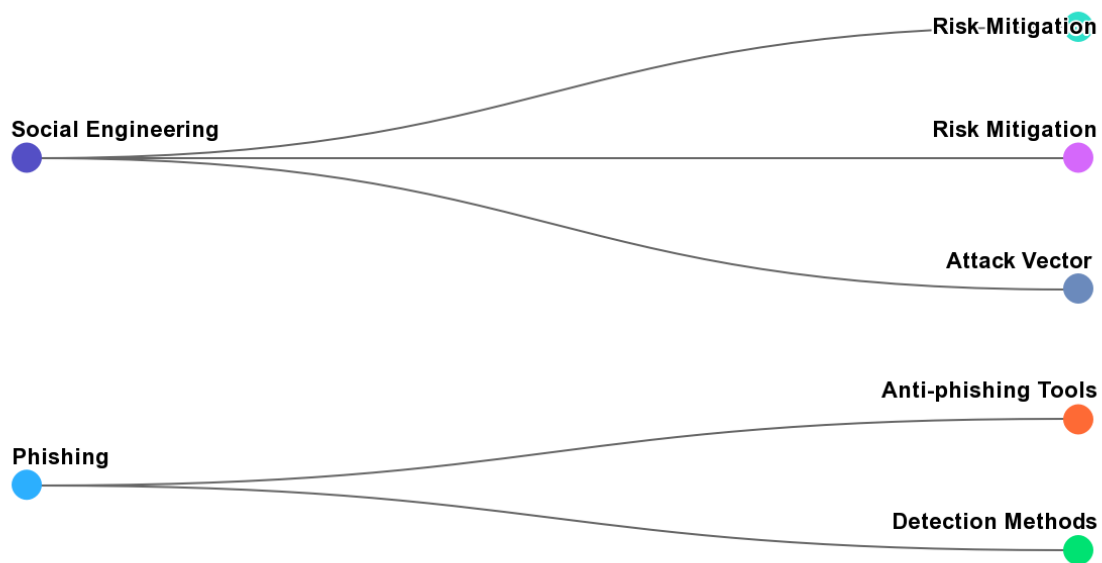
2023 novemberében megvizsgáltuk a téma tudományrepresentációját az Elsevier Scopus adatbázisában is. Itt a fejlett keresési módban az alábbi lekérdezést írtuk: „ALL ("Social Engineering") AND ALL ("Email security") AND ALL (deep AND learning) AND (LIMIT-TO (SUBJAREA , "COMP"))., Itt is hasonlóan a JSTORE-os lekérdezéshez a „Computer Science” az az a számítás vagy számítógépes tudomány területre szűrtünk, mivel szűrés nélkül sok fals találatot kaptunk. A lekérdezés 5 folyóirat cikket, 3 tudományos összefoglalót/áttekintést, 1 konferencia cikket és 2 könyv fejezetet.



11. ábra TUDOMÁNYREPREZENTÁCIÓ SCOPUS-BAN 2023 ŐSZ

4.9. SCOPUS AI – TÉMA FELDERÍTÉSE

A Scopus béta verziójú mesterséges intelligencia téma felderítőjével is elkészítettünk egy grafikont, amivel gráfok segítségével láthatóvá tettük, hogy hogyan kapcsolódnak a cikkek téma-területei.



12. ábra SCOPUS AI – TÉMA FELDERÍTÉSE | Szerző által szerkesztett

5. PROBLÉMÁK MEGFOGALMAZÁSA

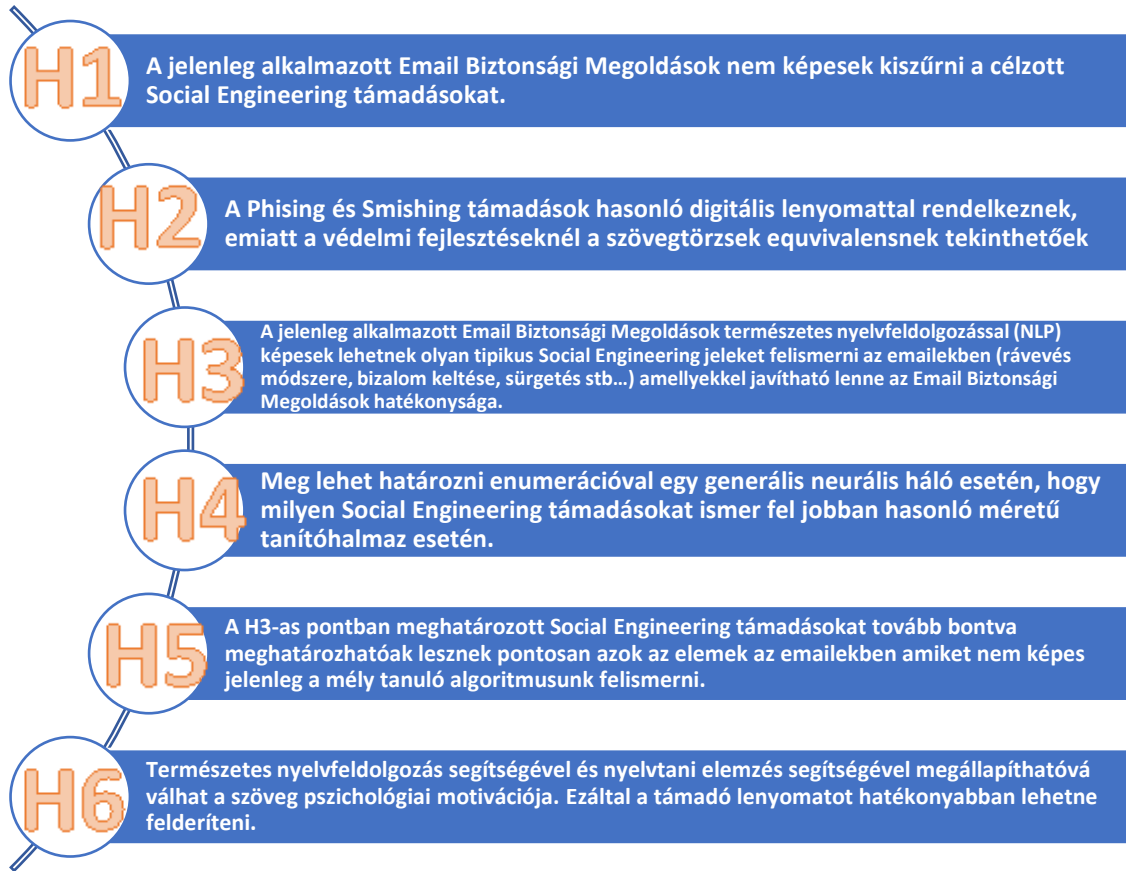
5.1. ÉSZREVÉTELEK

Számos műszaki és informatikai megoldással találkoztunk, amelyek nem voltak képesek megfelelő hatékonysággal, vagy egyáltalán nem tudták detektálni és kezelni a társadalmi mérnöki támadásokat. Egy támadó által küldött kártékony emailt számos ponton meg lehetne akadályozni, attól függően, milyen típusú támadásról van szó. Példaképpen kipróbáltuk, hogy egy átlagos védelemmel rendelkező üzleti laptopra (reaktív vírusirtó, tűzfal, email biztonsági megoldás a levelező szerver előtt) küldünk egy kártékony emailt, és meglepve tapasztaltuk, hogy semmilyen riasztás nem érkezett. [23]

Továbbá komoly problémának látjuk, hogy a támadók egyre gyakrabban alkalmaznak társadalmi mérnöki támadásokat. A konvencionális támadásokra már sikerült hatékony műszaki védelmet kifejlesztenünk, azonban a humán tényező, amely a biztonsági lánc leggyengébb láncszeme, ilyen ütemben nem fejlődött. Korábbi kutatásaink és jelenlegi interjúink azt mutatják, hogy nagyon nehéz a munkavállalókat felkészíteni a professzionális szinten kidolgozott támadások ellen. A biztonságtudatosító képzések hatékonysága gyakran hasonlít egy kevésbé hatékony tűz- és munkavédelmi előadáséhoz. Ezen a helyzeten különböző módszerekkel, például gamifikációval lehet javítani [8] [23].

Kiemelnénk Oroszi Eszter Diána biztonságtudatosítási szabadulószoftver megoldását, amely sokkal hatékonyabbá teheti ezeket a képzéseket. Mivel azonban az emberek képzése nehéz és költséges, a szervezetek vezetői általában inkább olyan műszaki megoldásokra vágynak, amelyeket egyszer megvásárolhatnak, és utána minimális képzéssel vagy anélkül is működőképesek maradnak. [23]

5.2. HIPOTÉZISEK



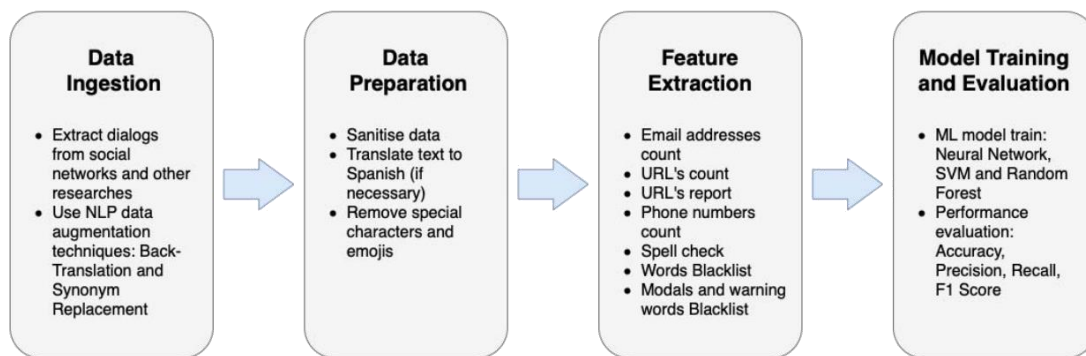
13. ábra Hipotézisek [23] | Szerző által szerkesztett

6. SZEKUNDER KUTATÁS

6.1. IRODALOMKUTATÁS EREDMÉNYEI

6.1.1. Technológiai 11.05

Juan C. Lopez és Jorge E. Camargo [14] modellje NLP-t, gépi tanulást, vektoros gép segítést és véletlenszerű erdő algoritmust alkalmaz. Ezzel állításuk szerint 84%-os pontosságot tudtak elérni. A rendszerük grafikusan a következő képpen néz ki [23]:



14. ábra [14, p 178, fig1]

Ennél a modellnél nagyon előre mutatónak érzem, hogy NLP-vel már bizonyos Social Engineering támadási formákat már képes felismerni, mint pl.: ráhatás, túlterhelés és a tekintély bizonyos fajtáit. [14, p 179] Ezen felül a konvencionális eszköztárból is használtak URL vizsgálatot, szó fekete listát és telefonszám vizsgálatot. [23]

- Bronjon Gogoi és Tasiruddin Ahmed [15, p 2] modellje transfer learning-et alkalmaz annak érdekében, hogy kevesebb adatból is képes legyen pontosabb döntéseket hozni. A modell alapjául a Google által fejlesztett BERT (Bidirectional Encoder Representation from Transformers) és DISTILBERT modelleket használták, mivel ezekre nagy előre tanított corpusok voltak elérhetőek. Használtak továbbá NLP-t, amit NSP-vel (Next Sentence Prediction – Következő Mondat Jóslása) és NER (Named Entity Recognition - Nevezett Entitás Felismeréssel). A modellük grafikus szemléltetése [35]

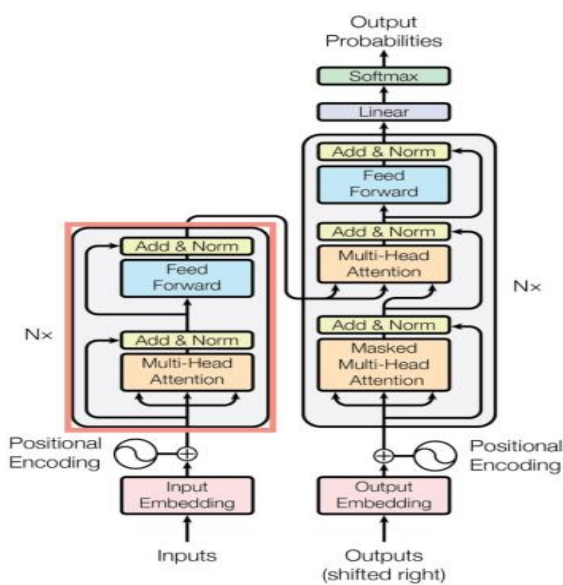


Fig. 3. Transformer Architecture

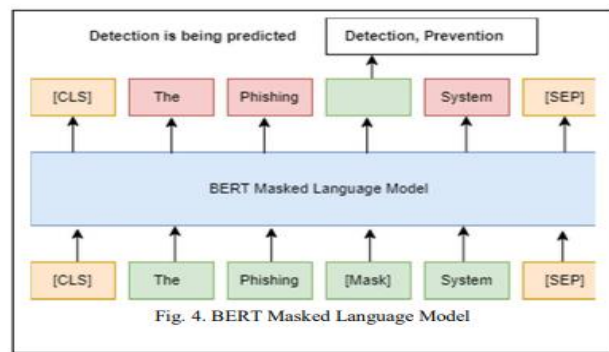


Fig. 4. BERT Masked Language Model

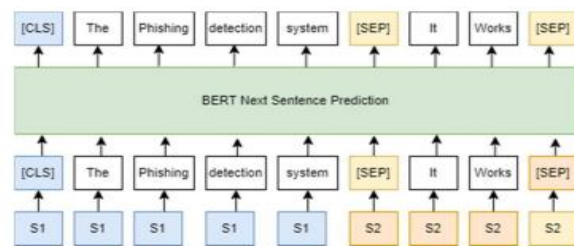


Fig. 5. BERT Next Sentence Prediction Model

15. ábra [35]

- Egy Indiai kutatás azt vizsgálta hogyan lehet az URL-eket és az emailben lévő QR kódokat gépi és mély tanulással érzékelni. Kutatásuk hatékonysága az URL és QR kód vizsgálat során átlagosan 90% felett volt [15]
- Szintén Indiai kutatók egy Konvolúciós neurális hálót tanítottak fel és véletlenszerű erdő algoritmust használtak a hatékonyság növelése érdekében [16]
- Török kutatók SMS és Email spameket (bináris klasszifikációt használva csak 2 részre osztották a halmazt: SPAM és Nem SPAM) kiszűrő modellt készítettek amivel magas hatékonyságot értek el, de Ők is csak 5000 alatti adathalmazt használtak. [17]
- Pakisztáni kutatók szintén megállapították, hogy az eddigi megoldások nagy része csak az URL-t vizsgálja és nem magát a törzs szöveget. [18]
- Kínai kutatók egy LSTM (hosszú rövid távú memória) és extrém gradiens támogató fa alapú megoldást dolgoztak ki az adathalász támadások ellen. 500.000-es corpusszal dolgoztak amit a CMU-CERT-től kaptak. Jelenlegi prototípusuk megfelelő alapként szolgál a további kutatásaikhoz, amit folytatni fognak. [19]

6.1.2. Nyelvi

Szekunder kutatásunk széles betekintést engedett az adathalászat, a spam és az internetes megtevesztés témakörébe. Első sorban kezdjük rövid történeti áttekintéssel.

Az SMTP fejlesztésével az 1980-as évek elején indult el a spam korszaka, amikor ezt kezdték alkalmazni az e-mailek küldésére. A spamküldők kihasználták az eredeti SMTP tervezését, amely nem követelte meg az e-mail küldők azonosítását. A közép-1990-es években megjelent az automatizált tömeges e-mail küldés technológiája, amit a Floodgate nevű első kereskedelmi "spamware" hozott el. Ezt követően a szolgáltatók és spammerek közötti harc alakította a spam piac dinamikáját. Az adminisztrátorok kezdetben kulcsszavak alapján szűrték a spam üzeneteket, de a spammerek gyorsan alkalmazkodtak, például helyesírási hibákat vetettek be. (Deborah Schafferh SchafferThe Language of Scam Spams) Az anti-spam fejlődése az azonosítási protokollok, gépi tanulás, tömeges adatgyűjtés és IP feketelistázás bevonásával zajlott. A spammerek kreatív taktikákkal próbálták megkerülni a szűrőket, de a feketelistázás vált a leghatékonyabb fegyverré. A nagyobb szolgáltatók, mint a Yahoo! Mail és a Gmail, kifejlesztették dedikált anti-spam csapataikat, míg a kisebb szolgáltatók nehézkesen tudtak versenyezni a magas költségek és a koncentráció növekedése miatt. [28]

Jensen tanulmányában pedig meg is állapítja, hogy az adathalászat és a spam elleni harc legnagyobb mozgatórugója a gazdasági veszteség [29] amit Scherbina és Shulusche is megerősít a folytatott tanulmányaik alapján, éves 3,3%-ra teszik ezen kárt az USA-ban.[30] Mindemellett a hadsereg is nehézségekkel néz szembe a kiberbiztonsági szektorban ez alatt az idő alatt. [31]

Elindulnak tehát a kutatások a spam kiszűrésére, új IT eszközöket fejlesztenek különböző módszerekkel, többek között Kernel-t és gépi tanulást használva az adathalász üzenetek és weboldalak hatékonyabb észleléséhez. [32]

Kísérletek születnek a pszichológiai modellek kapcsán is, mind a belső [33] mind a külső fenyegetések kiszűrésére (Dark Data) és megszületik az online átverés-detektálási megközelítések, melyek hangsúlyozzák az információbiztonsági jelek szerepét a phishing észlelésében, és figyelembe veszi az egyéni hajlamot a jelzés értelmezésében és a hitelesség kérdéskörére teszik a hangsúlyt az adathalászat problémájának megoldására. [34]

7. PRIMER KUTATÁS

7.1. SZAKÉRTŐI MEGKÉRDEZÉSEK

A készített interjúk első része inkább a témám irányát határozta meg, a második része pedig olyan kihívásokat vetett fel, amik jelenleg még nem megoldottak. Az egyik leghasznosabb interjút egy telekommunikációs - szolgáltató cég üzletágvezetőjével sikerült készítenünk, ahol betekintést kaphattunk a telefonszolgáltatások biztonsági architektúrájába. Ezen interjú alkalmával rákérdeztünk arra is, hogy az 1. ábránkban említett pl.: Netflixes csalások ellen milyen védekezési mechanizmusokat használnak és mi az oka, hogy ennyire elterjedtek az SMS-es adathalás támadások. Azt a meglepő választ kaptuk, hogy két védekezési mechanizmust használ a megkérdezett szakértő szervezete (ugyanakkor kiemelte, hogy nem csak ők, hanem a többi Magyarországi szolgáltató is). Az első egy olyan empirikus predikció, amely gyanúsan sok kiküldött SMS esetén beriaszt és jelzi a szolgáltatónak, hogy pl.: 500.000 SMS-t küldtek ki ugyanazzal a szöveggel. A másik védelmi mechanizmus SMS-ek szűrésére az a jelentés alapú telefonszám kitiltás. Ez esetben, amikor kellő számú bejelentés érkezik a szolgáltatóhoz és kivizsgálják a gyanús eseteket akkor ideiglenesen letilthatják a telefonszám SMS küldési lehetőségét a hálózatról. Ezen túl viszont mást nem alkalmaznak. Ezzel a szolgáltatások biztonságát érintő megosztott felelősségi modell szerint 4 több felelősség helyeznek a felhasználóra mind a lakossági ügyfelek, mind a vállalatok esetén. Amennyiben viszont a vállalatoknál nincs megfelelő hatékonyságú biztonságtudatosító képzés ezeket a támadásokat nagy eséllyel nem fogják tudni kivédeni. [23]

A többi interjú tartalma és eredménye megegyezett. Interjú alanyainkkal email biztonságról beszélgettünk és minden interjún azt az eredményt kaptuk vissza, hogy a phishing emaileket sehogyan nem tudják megfelelően szűrni. Egy alkalommal merült fel, hogy megoldás lehet, ha az email szabályzatban letiltjuk a szervezeten kívüli emailek fogadását ezzel tovább erősítve a Zero Trust⁵ megközelítést, de ez csak kevés helyen vitelezhető ki, mivel a vállalatoknak nem a belső levelezésük termel majd bevételt. [23]

⁴ A megosztott felelősségi modell (angolul: Shared Responsibility Modell) a felhő biztonságban elterjedt biztonsági kifejezés, de véleményünk szerint más kiberbiztonsági területeken is remekül használható különböző kockázati tényezőinek felírására a sok tényezős biztonsági kihívások elemzésére. [9]

⁵ Bizalmatlanság elve. Egy olyan kiberbiztonsági jó gyakorlat a jogosultság kezelésnél, amivel csak az kap jogosultságot az adott folyamathoz vagy erőforráshoz, akinek feltétlenül szüksége van rá. Ezzel megelőzhetővé válik, hogy a fertőzés a szervezeten belül elterjedjen, illetve a visszaélések is visszaszoríthatóak vele.

7.2. LAKOSSÁGI MEGKÉRDEZÉS (KÉRDŐÍVVEL)

7.2.1. Módszertan

A készített interjúk két fő részre oszthatók. Az első rész során a témakörrel kapcsolatos irányvonalak és kihívások kerültek felszínre, amelyek jelenleg még nincsenek teljesen megoldva. Ebben a vonatkozásban kiemelkedő interjúnk volt egy telekommunikációs szolgáltató cég üzletágvezetőjével, akit Kollár Csaba 2018-as [20] cikke alapján szakértőnek lehet tekinteni, hiszen ezt foglalkozása és tapasztalata igazolja. Az interjú során az alany részletesen kifejtette a telefonszolgáltatások biztonsági architektúráját és azzal kapcsolatos védekezési mechanizmusokat. A szervezete tehát két fő védelmi mechanizmust alkalmaz. Az egyik az empirikus predikció, ami gyanúsan nagy mennyiségű SMS esetén riadót vált ki, figyelmeztetve a szolgáltatót, ha például ugyanazzal a szöveggel nagy mennyiségű SMS-t küldenek. A másik védelmi mechanizmus a jelentés alapú telefonszám kitiltás, ami akkor aktiválódik, ha a felhasználók jelentik a gyanús tevékenységet, és a szolgáltató vizsgálja a jelentéseket. Ebben az esetben ideiglenesen letilthatják a telefonszám SMS küldési lehetőségét a hálózaton. Tehát több felelősséget helyeznek a felhasználóra, mind a lakossági ügyfelek, mind a vállalatok esetén. Amennyiben viszont a felhasználó tájékozatlan, vagy nincs kiképezve ilyen üzenetek szűrésére, ezt meglehetősen bizonytalan kimenetelű módszernek találom.

Más interjúk tartalmilag hasonló eredményeket hoztak. Az emailbiztonságról folytatott beszélgetések során minden interjúalany azt emelte ki, hogy a phishing emaileket még mindig nehéz megfelelően szűrni. Egyik esetben felvetődött az ötlet, hogy az email szabályzatban a szervezeten kívüli emailek fogadását le kell tiltani, ami tovább erősítheti a Zero Trust (ford.: Nulla Bizalom) megközelítést. Azonban ez a megoldás csak néhány helyen alkalmazható, mivel sok vállalat számára az olyan emailek is fontosak lehetnek, amelyek a vállalaton kívülről érkeznek, és az ilyen tiltás az üzleti tevékenységet is negatívan befolyásolhatja, mivel nem minden esetben a belső levelezés generálja a bevételeket.

7.2.2. Eredmények

Az adatahalászat kérdőívet 140 válaszadó töltötte ki, akik különböző neműek és korcsoportokba tartoznak. A válaszok szerint a résztvevők többsége férfi (56%), de nők (39%) és nem-bináris (2%) személyek is kitöltötték a kérdőívet. További 2% nem kívánt válaszolni és 1% pedig az egyéb kategóriába sorolta magát.

A kérdőív válaszadói között széleskörű korcsoport-eloszlás figyelhető meg. A legnagyobb arányban az 18-25 év közötti korcsoport szerepelt, akik a válaszadók 73,5%-át tették ki. Ezt követték a 26-35 év közöttiek 10%-kal, majd a kis- és fiatalkorúak (13-17 év) 9,2%-kal, az 36-45 év közöttiek 4,3%-kal, az 46-55 év közöttiek 2,1%-kal, és végül az 65 év felettiek 0,7%-kal. A válaszadók között tehát széles körű korcsoportok képviseltették magukat, amelyek között az 18-25 év közöttiek domináltak, amivel teljesítettük azon célkitűzésünket, hogy a kutatásunk elsődlegesen a saját kortársainkra fókuszáljuk.

Az iskolázottság terén a fiatal felnőttek túlnyomó többsége miatt hasonlóan mérsékelt a szórás. A válaszadók között az általános iskolai végzettséggel rendelkezők 18,6%-ban képviseltették magukat. A leggyakoribb végzettség az érettségi volt, amelyet a válaszadók 31,4%-a szerzett meg. A középfokú szakirányú végzettséggel rendelkezők 18,6%-ban szerepeltek. A BA/BSc vagy főiskolai végzettséggel rendelkezők 20%-ot tettek ki. A felsőoktatási szakképzést 2,9%-an, az MA/MSc vagy egyetemi végzettséget 5% szerzett. A PhD-t szerzettek 1,4%-ot képviseltek a válaszokban, míg az egyéb végzettséggel rendelkezők aránya 2,1% volt. Így az iskolázottság terén is sokféle végzettséggel rendelkező válaszadók találhatók a mintában, de az érettségivel és a BA/BSc végzettséggel rendelkezők egyértelműen dominálnak.

Az adathalászat jelensége iránti tudatosság és az áldozattá válás mértéke is fontos részét képezi a kutatásnak. A válaszadók 132-en, azaz 94%-a már hallott az adathalászatról, míg csak 8-an, azaz 6%-a nyilatkozott úgy, hogy még soha nem találkozott ezzel a kibertámadási formával. Az áldozattá válás szempontjából a válaszok szintén beszédeseek. A válaszadók 107-en, azaz 76%-a mondta, hogy már próbálkoztak az adatai kicsalásával, míg 33-an, azaz 24%-a nyilatkozott úgy, hogy még nem volt áldozata ilyen támadásnak. Azok között, akik már áldozattá váltak, 20-an, azaz 14% mondta, hogy legalább egyszer esett már adathalászat áldozatául, míg a többiek szerencsésebbek voltak, és nem sétáltak bele ilyen támadásokba. A válaszok alapján tehát elmondható, hogy a megkérdezettek nagy része hallott már az adathalászat jelenségéről, tehát tudatosabbak a jelenséggel kapcsolatban, amivel gyakran találkoznak is és kevésbé esnek általa csapdába.

Ugyanakkor azok közül, akik válaszoltak a kérdésre, hogy „Tudnak-e a környezetükben olyan személyről, aki esett adathalászat áldozatául?”, a válaszok alapján a következő arányok állnak fenn: az igen válaszadók aránya 36%, míg a nem válaszadók 64%-ot képviselnek. A válaszadók között, akik igen válaszoltak, különböző mértékű áldozattá válások is előfordultak. Ebből adódóan, az 1 fő által tapasztalt esetek aránya 1,4%, a 2-5 fő közötti eseteké 22,1%, 5-10 fő közötti eseteké 3,6%, 10-25 fő közötti eseteké 2,1%, és több mint 100 fő áldozatává válás is előfordult,

amelyet a válaszadók 1,4%-a jelentett. Fontos kiemelni, hogy a 25-50 fő közötti esetről és a 50-100 fő közötti esetről szóló válaszok aránya mindkét esetben 0% volt ebben a mintában. A válaszok összességében bemutatják a környezetükben esett phishing áldozatok változatos számát és gyakoriságát az adatokban.

A "Többször is megtörtént-e a támadás?" kérdésre adott válaszokat elemezve megállapítható, hogy a kitöltők különböző mértékben tapasztalták az adathalász támadásokat. A válaszok eloszlása a következőképpen alakult: 9,5% azok, akik szerint elég gyakran érték őket ilyen támadások, 38,1% pedig azt mondta, hogy néha megesik velük. Emellett 23,8% azt állította, hogy csupán egyszer érte már ilyen támadás, míg 22,2% szerint bár még nem tapasztalták, megeshetett volna velük. A válaszadók 6,3%-a úgy véli, hogy nem eshetnek áldozatul ilyen támadásoknak, és 0,0% nem kívánta megosztani adatait ezzel kapcsolatban.

A keletkezett kár értékét tekintve az elenyészőtől, a pár ezer forintos és néhány százezer, majd pár milliós károkon keresztül egészen a több száz milliós károkig számoltak be minden féléről a válaszadók szabad, szöveges válaszaikban. Volt olyan is, aki leginkább a kár helyrehozatalára szánt időt fájlalta különösen utólag.

Az adatok alapján a válaszadók közül 88,6% hallott az Email Phishing-ről (Email Adathalászat), míg a Spear Phishing (Lándzsás adathalászat+) típusról 45,7% tájékozott. A Pharming (Pharming) jelenségéről 42,8% hallott már, míg a Vishing (Hang-phishing) típusról 57,1% számolt be. A Smishing (SMS-phishing) ismertsége 72,1%, a Clone Phishing (Klónozó-phishing) pedig 50%. A Whaling (Nagyhalászat) jelenségről 20% hallott, míg a Social Media Phishing (Társadalmi médiaphishing) típusra 69,3% emlékszik. Az Üzleti e-mail meghekkelés (Business Email Compromise - BEC) típust 35% ismerte fel, míg a Kiberbűnözési támadások (Ransomware Attacks) típusáról 47,1% hallott.

Az összegzés során azt látjuk, hogy az adathalászat széles körű probléma, és sokféle típusú támadás érintheti az embereket. Az emberek különböző módon reagálnak az ilyen támadásokra, és vannak, akik már többször is áldozattá váltak, míg mások szerencsére elkerülték ezeket a veszélyeket, de többen környezetükben nem voltak ennyire szerencsések. Az adathalászattal kapcsolatos ismeretek terén még mindig van fejlődési lehetőség, és az embereknek fontosnak kell tartaniuk az online biztonságukat.

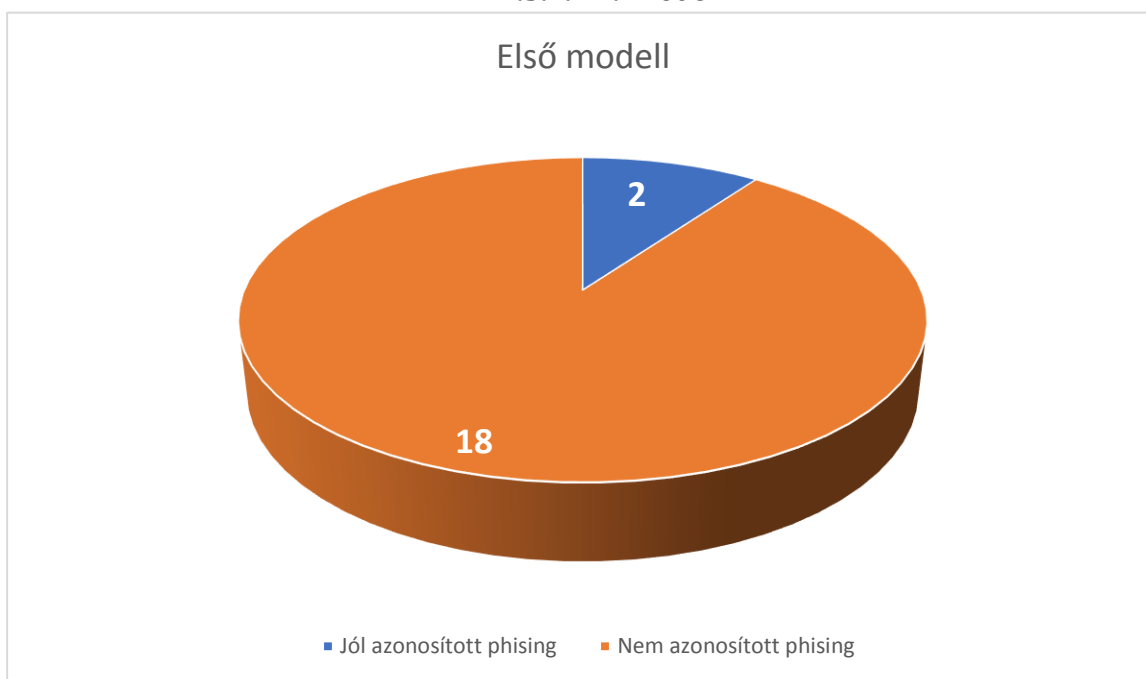
7.3. KORÁBBI MÉRÉSEK BEMUTATÁSA

Sokszor futottam bele primer kutatásom során abba, hogy nagyon magas hatékonyságúnak ígérkező klasszifikáló modelleket találtunk a közösség által fejlesztett Kaggle nevű oldalon. [10]

Itt sokszor 90+ % hatékonyságú modelleket láttunk, viszont ez gyanúsnak tűnt, mivel amikor korábban készítettünk egy interjút egy telekommunikációs cég kutató mérnökével, tőle azt az információt kaptuk, hogy annak ellenére, hogy a piacvezető beszállítók ilyen technológiában 30-40%-nál nagyobb hatékonyságot nem tudtak elérni. Ebből arra következtethetünk, hogy ahol nagyon magas hatékonyság látszik, ott a modellt túltaníthatták, ennek következményében az eredeti adat halmazával jó eredményeket tud elérni, viszont valós támadások esetén már sokkal kevésbé. [23]

Ennek igazolására 3 modellnél [11][12][13] megmértük mi történne, ha phishing emaileket írnánk, amiket a papíron jól teljesítő modelleknek kellene klasszifikálnia. Mind három modell esetében 20-20 emaillel végeztünk méréseket. Fontos azonban kiemelni, hogy a használt modellek alapvetően binárisan foglalkoztak az email biztonsággal. Vélhetően az egyszerűség és vagy a hatékonyság (futási időre értendő, nem pontosságra) kedvéért, de nem tesznek különbséget a SPAM, phishing vagy egyéb más kártékony emailek között. A SPAM kategóriába kerül minden kártékony email. Ezt az eredmények értékelésénél fontos figyelembe venni. Az eredmények grafikus formában itt láthatóak: [23]

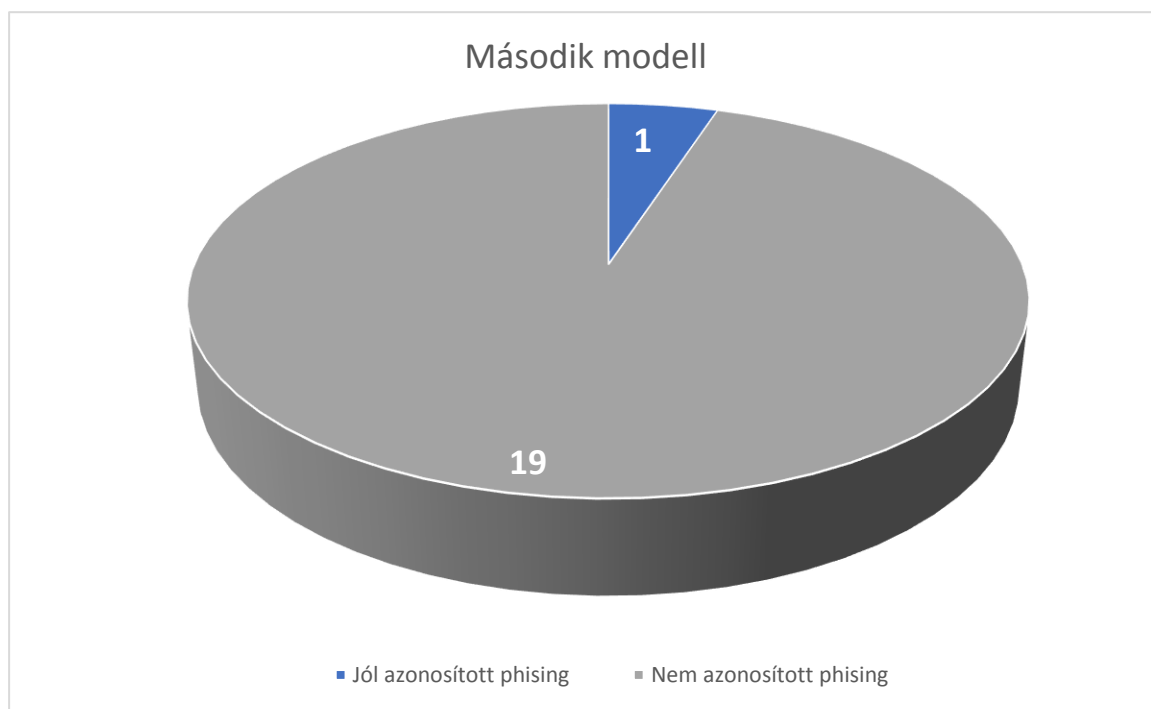
7.3.1. 1. Modell



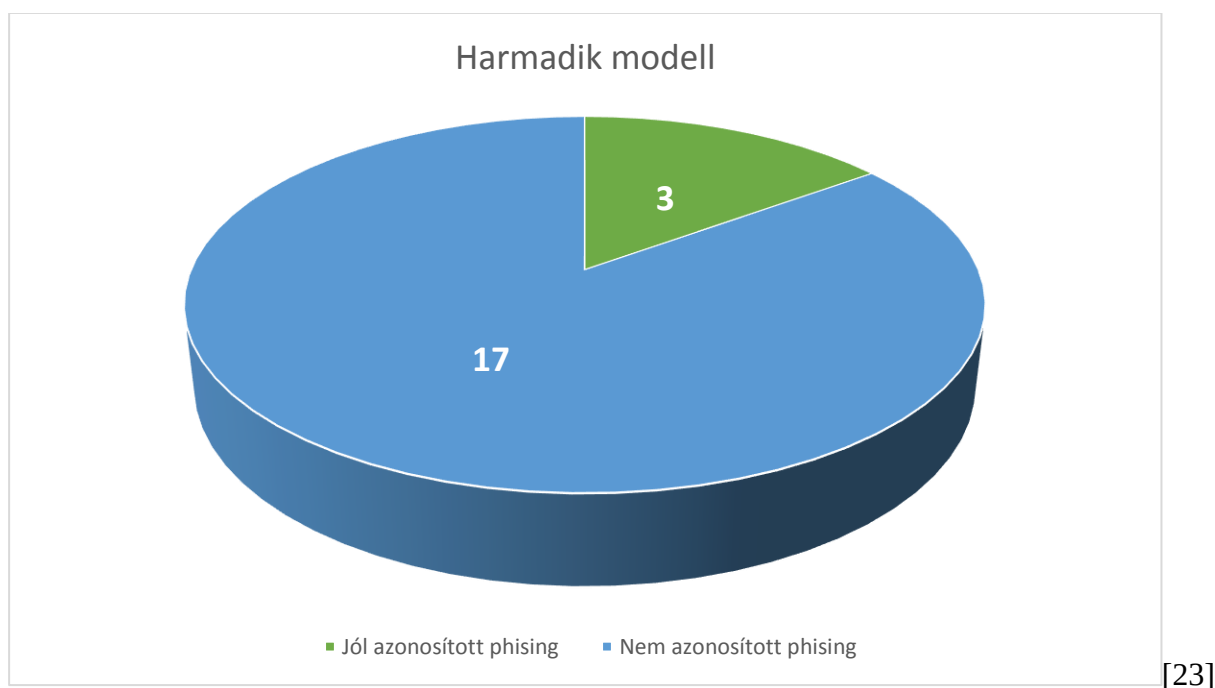
[23]

7.3.2. 2. Modell

[23]



7.3.3. 3. Modell



7.3.4. További modellek sikertelensége

Bizonyos modellek esetén pedig a még emberek számára is feltűnően egyértelmű támadások azok amiket nem érzékelnek. Ezeknek az oka az lehet, hogy nagyon kis méretű és főleg egy

típusú adathalmazból túltanították a modellt, ezáltal 1 nagyon kis szegmensset nagyon jól lefed, de a támadások nagy többségét nem fogja érzékelni.

Oracle sphere test!

```
[ ] prediction_on_test_data = model.predict(X_test_features)
    accuracy_on_test_data = accuracy_score(Y_test, prediction_on_test_data)
```

```
[ ] print('Accuracy on test data: ', accuracy_on_test_data)
```

Accuracy on test data: 0.9659192825112107

Predictive System INPUT

```
[ ] input_mail = ["Won 5 free entry for the basketball match click on the link to achive: http://totallyscam.com/scam/phising/spam/thisistotallynotavirus.exe"]
# Text --> Vector
input_data_features = feature_extraction.transform(input_mail)

#PRED
prediction = model.predict(input_data_features)
print(prediction)

if(prediction[0] ==1):
    print('Ham Mail')
else:
    print('Spam mail')

[1]
Ham Mail
```

16. ábra Szerző által készített kód részlet

7.4. EMAIL EXPORT-IMPORT ELŐRE PONTOZÁS

A szakirodalom, valamint az iparban használatos megoldásokban [19] is bevett rendszer az, hogy előre pontozzák a kockázatot különböző tényezők alapján.

Ezt mi is megtettük. Ezért a Microsoft Office365 Outlook levelező rendszeréhez készítettünk egy Microsoft Power Automate folyamatot, ami az új beérkező emaileket mielőtt a felhasználó olvasná kiszedi a beérkezett email-ek közül és elvégez rajta olyan ellenőrzéseket, amivel a nagyobb kockázati besorolásokat meg lehet tenni, majd átadja az emailt egy mély tanuló modellnek. Ezután visszakapja a Power Automate folyamat a feldolgozott emailt és a kockázati százalék alapján a rendszerüzemeltetés beállításai alapján dönt arról, hogy az email átengedhető-e vagy eldobandó.

A Power Automate folyamatunk először is megvizsgálja a küldő felet. Amennyiben saját tartományon belüli a küldő, úgy nem emeli a kockázati pontszámot, mivel a saját tartományon belülről érkező támadások nagyrészt nem adathalász alapúak. Megvizsgálja, hogy a tárgy és a szövegtörzs tartalmaz-e gyakran alkalmazott adathalász mondatot. Valamint megvizsgálja, hogy tartalmaz-e kártékonynak tűnő csatolmányt.

7.5. ÚJ TÁMADÁSI OSZTÁLYOK LÉTREHOZÁSA

Korábbi kutatásaink során azt a konklúziót vonhattuk le, hogy jelenleg a szakirodalom nem rendelkezik elég mély osztályozási rendszerrel a támadások klasszifikációja kapcsán. Ezért világossá vált, hogy amennyiben az adathalász támadásokat mély tanulás során jobban detektálhatóvá akarjuk tenni szükségünk van több osztályra, ami további rétegeket jelenthet a neurális hálónk számára. Az új osztályok elkészítését elkezdtük, de mivel nagyon sok nyelvi elem is befolyásolja az új osztályok kialakítását ezért a Magyar Kiberbiztonsági Klaszter-től igényeltünk email adatbázisokat kutatási célra. Ennek mélyrehatóelemzése után előreláthatólag 2024 tavaszán kerülhetnek bemutatásra az új osztályok.

7.6. MÉRÉSEK SZABADON HASZNÁLHATÓ NAGY NYELVI MODELLEKKEL

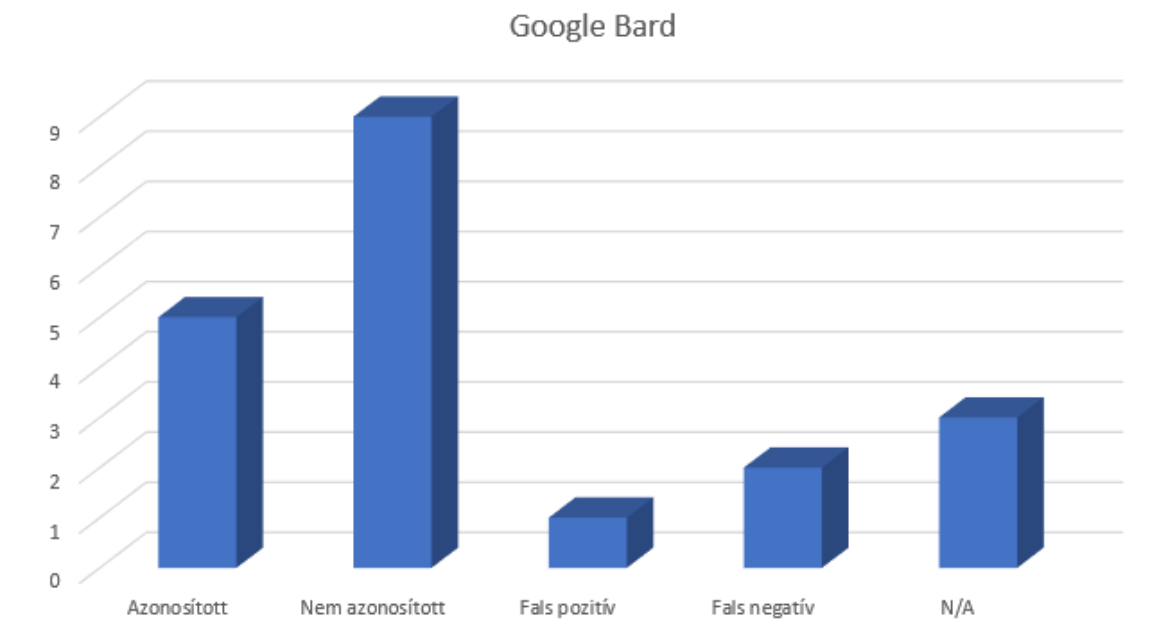
Mivel manapság nagyon felkapottak lettek az NLP alapú nagy nyelvi modellek ezért 2 ilyen modell esetén megvizsgáltuk az ilyen nagy nyelvi modellek alkalmazhatóságát.

7.6.1. Bard

A Google által fejlesztett Bard hasonlóan a ChatGPT-hez egy nagy nyelvi modell, amivel egy chat felületen képesen vagyunk interaktálni. Kutatásunk során megvizsgáltuk, hogy egy ilyen nagy nyelvi modell, amely természetes nyelvfeldolgozáson alapul képes-e segítségünkre lenni abban, hogy kiszűrjük az adathalász támadásokat.

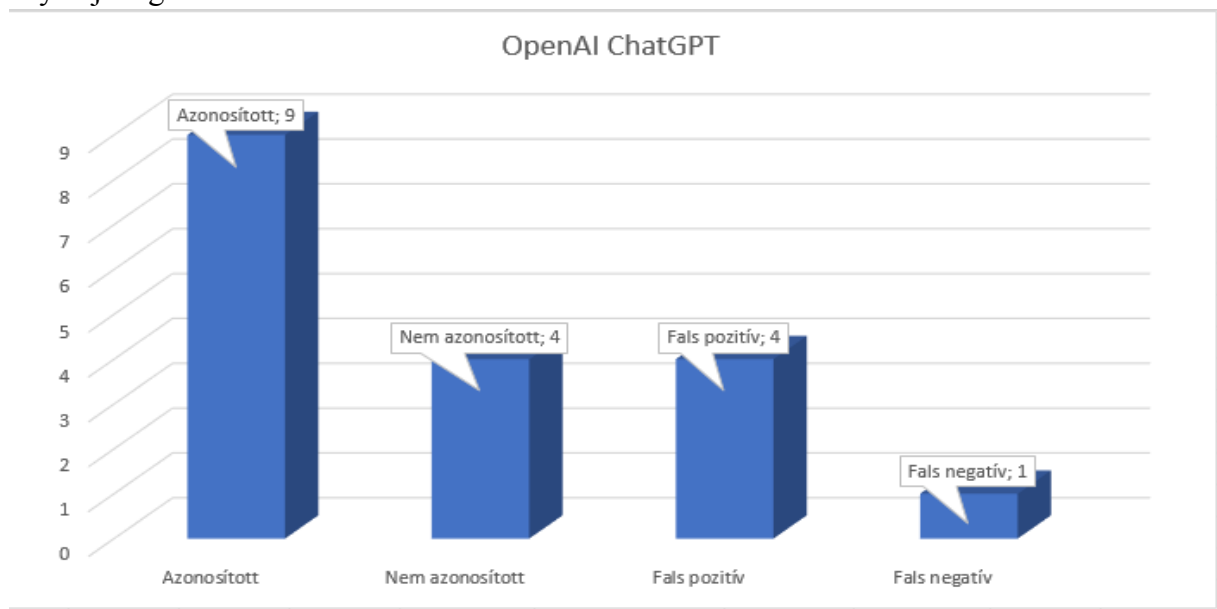
Hasonlóan a korábbi 3 modell vizsgálatához ezt a modellt is alávetettük olyan mérésnek, aminek során 20 támadó email kapott, amiről el kellett döntenie, hogy szerinte kártékony-e.

Az adatok a grafikonon láthatóak. Amit külön kiemelnék, hogy egy támadásnál leírta, hogy nem tudja megítélni. Amit összevetve a ChatGPT-vel pozitív lépésnek tartok, mivel a ChatGPT sosem ad „nem tudom” választ, hanem mindig megmagyarázza szerinte miért van igaza. Ez értelemszerűen sok fals eredményhez vezet.



7.6.2. ChatGPT

Chat GPT-vel is elvégeztettük a méréseket. Ahol elég magas azonosítási százalékot tapasztalhattunk, viszont ki kell emelni, hogy nagyon magas a fals pozitív aránya (20%). Ez valós helyzetben azt jelentené, hogy amennyiben valós email biztonságához használnánk a ChatGPT-t úgy 100 valódi emailból 20 eldobásra kerülne. Ezért jelenlegi formájában elvethető a ChatGPT ilyen jellegű alkalmazása.



7.7. SAJÁT MÓDSZER KIDOLGOZÁSA

Kutatásunk során elkezdtünk egy saját módszert kidolgozni.

Az alábbi folyamat ábrán ennek a béta verziója található.



17. ábra Szerző által szerkesztett

8. KONKLÚZIÓ - TÉZISEK

Mivel A kutatásunk célja az volt, hogy megvizsgáljuk és fejlesszük az email alapú Social Engineering támadások elleni védekezési lehetőségeket, különös tekintettel a mély tanulás és a természetes nyelvfeldolgozás (NLP) alkalmazására. Az információs technológia rohamos fejlődése közepette a biztonság megőrzése nem tartott lépést a támadók egyre kifinomultabb módszereivel. A hagyományos védelmi megoldások sokszor nem bizonyultak elegendőnek, különösen a pszichológiai manipulációval kombinált támadások ellen. Kutatásaink és szakirodalmi áttekintésünk során számos fontos megállapítást tettünk.

1. Eredmények összefoglalása

1.1 Tudományos és technológiai alapok:

- A mély tanulás és a természetes nyelvfeldolgozás jelentős fejlődésen ment keresztül az utóbbi években, különösen a nagy nyelvi modellek (LLM) megjelenésével.
- A gépi tanulás technológiák, mint a ReLU, Tanh és Sigmoid átviteli függvények, valamint a neurális hálózatok különböző típusai (RNN, CNN) alapvető szerepet játszanak a modern biztonsági megoldások fejlesztésében.

1.2 Social Engineering támadások:

- A social engineering, különösen az adathalász támadások (phishing, smishing), egyre nagyobb kihívást jelentenek a vállalatok és szervezetek számára.
- A támadások klasszifikálása és tipizálása, valamint a támadási módszerek megértése kulcsfontosságú a védekezés hatékonyságának növelésében.

1.3 Kutatási eredmények:

- Korábbi kutatásaink és szakirodalmi áttekintéseink során arra a következtetésre jutottunk, hogy a technológiai megoldások önmagukban nem elegendőek a social engineering támadások elleni védekezésben.
- A humán alapú biztonságtudatossági képzések szerepe kiemelkedő, de ezek hatékonysága is korlátozott lehet.

1.4 Aktualitás és tudományos reprezentáció:

- Az MTMT és a nemzetközi adatbázisok (IEEE Xplore, Scopus, JSTORE) elemzése alapján megállapítottuk, hogy a social engineering és az email biztonság témakörei hazánkban és nemzetközileg is egyre nagyobb figyelmet kapnak.

- Az elmúlt években növekedett a kapcsolódó publikációk száma, ami jelzi a téma kutatásának növekvő jelentőségét.

1.5 Javaslatok a jövőbeli kutatásokhoz

- Integrált megközelítés: A social engineering támadások elleni védekezéshez integrálni kell a technológiai és humán megközelítéseket. A gépi tanulás és NLP technikák alkalmazása mellett fontos a biztonságtudatossági képzések fejlesztése és hatékonyságuk folyamatos mérése.
- Adathalász támadások elleni védelem: További kutatások szükségesek az adathalász támadások detektálására és megelőzésére szolgáló technológiai megoldások fejlesztésében, különös tekintettel a célzott támadások elleni védelemre.
- Interdiszciplináris együttműködés: Az információbiztonság és a társadalompszichológia közötti szorosabb együttműködés szükséges a támadási módszerek jobb megértéséhez és az ellenük való hatékony védekezés kialakításához.

2. Zárszó

A kutatásunk során megállapítottuk, hogy a social engineering támadások elleni védekezésben jelentős potenciál rejlik a gépi tanulási és NLP technológiák alkalmazásában. Azonban ezek a technológiák önmagukban nem elegendőek; szükség van a humán tényezők figyelembevételére és a biztonságtudatosság növelésére is. A jövőbeli kutatásokban az integrált megközelítés alkalmazása és az interdiszciplináris együttműködés további lehetőségeket nyithat a kiberbiztonság terén.

A kutatás során bizonyítottuk a H1-es hipotézisünket, amit az interjúk eredményei és a saját méréseink is alátámasztottak. A H3-as hipotézisünk a szakirodalomban bizonyítást nyert, ugyanakkor a jövőben szeretnénk elvégezni a saját méréseinket is. A H2, H4, H5 és H6-os hipotézisek jelenleg még nem eldönthetőek a rendelkezésre álló információkból.

Összességében a kutatás eredményei rávilágítanak arra, hogy a social engineering támadások elleni védekezéshez egy komplex, több szintű megközelítés szükséges, amely magában foglalja a technológiai innovációt és az emberek tudatosságának növelését egyaránt. Az ilyen jellegű kutatások és fejlesztések alapvető fontosságúak ahhoz, hogy a jövőben hatékonyan tudjunk szembenézni a kiberbiztonsági kihívásokkal.

9. MUNKAMEGOSZTÁS

A kutatás különböző részeinek elkészítésében Krasznyánszki Brúnó és Csizmadia Réka közösen vettek részt, az alábbi munkamegosztás szerint:

- **Krasznyánszki Brúnó:**
 - A műszaki alapok kidolgozása
 - A nyelvi rendszerek alapjainak és tesztjeinek elkészítése
 - Interjúk lebonyolítása
 - Az emailbázis összeállítása és elemzése
 - Mérések tervezése és elvégzése
- **Csizmadia Réka:**
 - Szekunder kutatás a szakirodalmi alapokból
 - Kérdőív elkészítése és annak összegzése
 - A nyelvi alapok kidolgozása
 - A kutatás nemzetközi kontextusba helyezése és a szakirodalom feldolgozása
- Mindkét kutató jelentős mértékben hozzájárult a projekt sikeréhez, különböző szaktudásukat és tapasztalataikat ötvözve egy komplex, több szempontú elemzés létrehozásában. Az eredmények alapján további együttműködésük révén még átfogóbb és mélyebb kutatások elvégzése válhat lehetővé a jövőben.

10. IRODALJEGYZÉK

- [1] ACM CSS Téma klasszifikáció az OTDT követelmények szerint az érintett témából generálva: <https://dl.acm.org/ccs> (generálás ideje: 2023. 05. 01 22:43)
- [2] <https://terranovasecurity.com/cyber-security-statistics/> (hozzáférés: 2023. 05. 02 14:44)
- [3] Christopher Hadnagy „The Art of Human Hacking” ISBN: 978-1-118-02801-8 WileyPublishing, Inc. (2011) p.23 szerző által fordított
- [4] Kevin D. Mitnick „A legendás hacker - A megtévesztés művészete” 9789632065557
- [5] TÓTH TAMÁS „AZ EGYES SOCIAL ENGINEERING MÓDSZEREK ELHATÁROLÁSA ÉS RENDSZEREZÉSE”, Katonai Nemzetbiztonsági Szolgálat Szakmai Szemle XVIII. évfolyam 1. szám 2020. március HU ISSN 1785-1181
- [6] Krasnyánszki Brúnó „Hogyan változtatta meg a XXI-ik századi kibervédelmet a Social Engineering?”, Évf. 4 szám 3 (2022): Biztonságtudományi Szemle Információbiztonság rovat
- [7] Dub Máté, Szakdolgozat, 'A kiberbiztonság humán aspektusának vizsgálata social engineering technikával', Nemzeti Közszerkezleti Egyetem Államtudományi és Nemzetközi Tanulmányok Kar Közszerkezési és Infotechnológiai Tanszék Közigazgatás-szerkező szak Nemzetközi közigazgatási szakirány pp, 35-50
- [8] Oroszi Eszter Diána, Biztonságtudatossági szabaduló- szoba, mint új programelem az információbiztonsági képzésekben Dunakavics A Dunaújvárosi Egyetem online folyóirata 2020. VIII. évfolyam III. szám pp 51-62
- [9] Shared Responsibility Modellről lásd a Felhő Biztonsági Szövetség megfogalmazását: <https://cloudsecurityalliance.org/blog/terms/shared-responsibility-model/> (utolsó hozzáférés 2023. 05. 03 12:24:38)
- [10] <https://www.kaggle.com/> (utolsó hozzáférés 2023. 05 .03 12:25:21)
- [11] <https://www.kaggle.com/code/mfaisalqureshi/email-spam-detection-98-accuracy> (utolsó hozzáférés 2023. 05 .03 14:13:32)
- [12] <https://www.kaggle.com/code/harshsinha1234/email-spam-classification-nlp>(utolsó hozzáférés 2023. 05 .03 14:14:43)
- [13]<https://www.kaggle.com/code/balaka18/email-spam-classification> (utolsó hozzáférés 2023. 05 .03 14:15:53)
- [14] Juan C. Lopez, Jorge E. Camargo, 'Social Engineering Detection Using Natural Language Processing and Machine Learning' 2022 5th International Conference on Information and

Computer Technologies (ICICT), DOI 10.1109/ICICT55905.2022.00038

[15] Bronjon Gogoi, Tasiruddin Ahmed, 'Phishing and Fraudulent Email Detection through Transfer Learning using pretrained transformer models', 2022 IEEE 19th India Council International Conference (INDICON) | 978-1-6654-7350-7/22/\$31.00 ©2022 IEEE | DOI: 10.1109/INDICON56171.2022.1004009

[15] S. S. S. P, S. Velpula, R. Parise, N. K. Vamsi and S. K. Chaitanya, "Phishing Attack Detection Using Convolutional Neural Networks," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1381-1385, doi: 10.1109/ICACCS57279.2023.10113077.

[16] A. A. Abdullahi and M. Kaya, "A Deep Learning Based Method to Detect Email and SMS Spams," 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 2021, pp. 430-435, doi: 10.1109/DASA53625.2021.9681921.

[17] A. Karthikeya, Y. B. Sai, S. Hariharan, A. C. Rao, D. Jignash and A. B. Prasad, "Prevention of Cyber Attacks Using Deep Learning," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023, pp. 1332-1336, doi: 10.1109/ICACCS57279.2023.10112794.

[18] S. Tareen, S. U. Bazai, S. Ullah, R. Ullah, S. Marjan and M. I. Ghafoor, "Phishing and Intrusion Attacks: An Overview of Classification Mechanisms," 2022 3rd International Informatics and Software Engineering Conference (IISEC), Ankara, Turkey, 2022, pp. 1-5, doi: 10.1109/IISEC56263.2022.9998205.

[19] D. He et al., "An Effective Double-Layer Detection System Against Social Engineering Attacks," in IEEE Network, vol. 36, no. 6, pp. 92-98, November/December 2022, doi: 10.1109/MNET.105.2100425.

[20] Kollár Csaba A SZAKÉRTŐVÉ VÁLÁS, ILLETVE A SZAKÉRTŐK KIVÁLASZTÁSÁNAK ÉS MEGKÉRDEZÉSÉNEK MÓDSZERTANI KIHÍVÁSAI VEZETÉSTUDOMÁNY / BUDAPEST MANAGEMENT REVIEW XLIX. ÉVF. 2018. 02. SZÁM/ ISSN 0133-0179 DOI: 10.14267/VEZTUD.2018.02.07

[21] [<https://www.historyofinformation.com/detail.php?entryid=782>]

[22] . [<https://www.pinecone.io/learn/series/image-search/imagenet/>]

[23] Krasnyánszki Brúnó 'Mesterséges intelligencia alapú email biztonsági megoldások alacsony hatékonyságának gyökér okainak vizsgálata Social Engineering támadások során ' Óbudai Egyetem 57. Intézményi Tudományos Diákköri Konferencia TDK dolgozat

[24] NOKIA Internal Introduction to AI Presentations/2/24 (nem nyilvános előadás)

- [25] Dr. Kutor László Intelligent Systems Presentation 2/12 (nem nyilvános előadás)
- [26] <https://phishtank.org/>
- [27][J. M. Rao és D. H. Reiley, „The Economics of Spam”, *Journal of Economic Perspectives*, köt. 26, sz. 3, o. 87–110, aug. 2012, doi: 10.1257/jep.26.3.87.]
- [28]<https://workspace.google.com/blog/identity-and-security/an-overview-of-gmails-spam-filters>
- [29] K. Jensen, „A Matter of Concern: Kenneth Burke, Phishing, and the Rhetoric of National Insecurity”, *Rhetoric Review*, köt. 30, sz. 2, o. 170–190, márc. 2011, doi: 10.1080/07350198.2011.552378.],
- [30] A. Scherbina és B. Schlusche, „The Effect of Malicious Cyber Activity on the US Corporate Sector”
- [31] „Government, Military Face Severe Shortage Of Cybersecurity Experts”.
- [32] A. Abbasi, F. “Mariam” Zahedi, D. Zeng, Y. Chen, H. Chen, és J. F. Nunamaker, „Enhancing Predictive Analytics for Anti-Phishing by Exploiting Website Genre Information”, *Journal of Management Information Systems*, köt. 31, sz. 4, o. 109–157, jan. 2015, doi: 10.1080/07421222.2014.1001260.
- [33] Greitzer, Kangas, Noonan, Brown, és Ferryman, „Psychosocial Modeling of Insider Threat Risk Based on Behavioral and Word Use Analysis”, *e-Service Journal*, köt. 9, sz. 1, o. 106, 2013, doi: 10.2979/eservicej.9.1.106.
- [34] University of Malta, I. Reid, L. Gozna, University of Leicester, J. Boon, „From Tactical to Strategic Deception Detection: Application of Psychological Synthesis”, *JSS*, köt. 10, sz. 1, o. 81–101, márc. 2017, doi: 10.5038/1944-0472.10.1.1528.
- [35] : <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10040097>

