

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

Nagy nyelvi modellel támogatott nyílt forrású információgyűjtés a kibertérben

Jelige: AI információvadászat

Pályamunka lezárásának időpontja: 2024. április

2025.

A KULTURÁLIS ÉS INNOVÁCIÓS MINISZTERIUM ÚNKP-23-2-I-NKE-30 KÓDSZÁMÚ ÚJ NEMZETI KIVÁLÓSÁG PROGRAMJÁNAK A NEMZETI KUTATÁSI, FEJLESZTÉSI ÉS INNOVÁCIÓS ALAPBÓL FINANSZÍROZOTT SZAKMAI TÁMOGATÁSÁVAL KÉSZÜLT.



TARTALOMJEGYZÉK

1. BEVEZETÉS	5
1.1. A vizsgált téma bemutatása	6
1.2. A kutatás célja.....	6
1.3. Hipotézisek	7
1.4. Módszertan	8
2. A NYÍLT FORRÁSÚ INFORMÁCIÓSZERZÉS (OSINT).....	9
2.1. Fogalmi meghatározások	9
2.1.1. <i>Az információ és a nyílt információ</i>	9
2.1.2. <i>A hírszerzés formái</i>	10
2.1.3. <i>A nyílt forrású információgyűjtés (OSINT)</i>	11
2.2. Az OSINT főbb jellemzői.....	13
2.2.1. <i>Források és módszerek</i>	13
2.2.2. <i>Az internetes nyílt forrású információgyűjtés</i>	15
2.2.3. <i>Az aktív és a passzív felderítés</i>	16
2.2.4. <i>Az „internetes OSINT eszközök”</i>	16
2.3. Az OSINT megjelenése a Cyber Kill Chain szakaszaiban	17
2.3.1. <i>Felderítés (reconnaissance)</i>	20
2.3.2. <i>Fegyverkezés (weaponization)</i>	21
2.3.3. <i>Szállítás (delivery)</i>	22
2.3.4. <i>Kihasztnálás (exploitation)</i>	23
2.3.5. <i>Telepítés (installation)</i>	23
2.3.6. <i>Futtatás és vezérlés (command & control)</i>	24
2.3.7. <i>Célok elérése (actions on objectives)</i>	25
2.4. Védelmi megoldások OSINT-ellen	26
2.5. A kibertéri nyílt forrású információszerzés SWOT elemzése	27
2.6. A mesterséges intelligencia megjelenése az OSINT-ben	30
3. A NAGY NYELVI MODELL (LLM)	31
3.1. Mesterséges intelligencia (MI)	31
3.2. A gépi tanulás módszerei (ML)	33
3.3. Természetes nyelvfeldolgozás (NLP).....	36
3.4. A nagy nyelvi modellek felépítése, működése és képességei.....	37
3.4.1. <i>A transzformátor architektúra</i>	39
3.4.2. <i>Fejlesztési és üzemeltetési szükségletek</i>	40
3.4.3. <i>Képzési adatkészletek általános típusai</i>	41

3.4.4.	<i>Nyílt adatkészletek gyűjtési eszközei</i>	43
3.4.5.	<i>A prompt engineering</i>	44
3.4.6.	<i>A tokenizáció</i>	45
3.4.7.	<i>Képességek és a főbb alkalmazási területek</i>	45
3.4.8.	<i>A multimodal képességek megjelenése</i>	46
3.5.	Nagy nyelvi modellek napjainkban	47
3.5.1.	<i>Hostolt nagy nyelvi modellek</i>	49
3.5.2.	<i>Nyílt forráskódú nagy nyelvi modellek</i>	51
3.5.3.	<i>A dark webbel kapcsolatos nagy nyelvi modellek</i>	52
3.5.4.	<i>Valós idejű információgyűjtési képesség</i>	54
3.6.	A nagy nyelvi modellek kiberbiztonsági sebezhetőségei	55
3.7.	Nagy nyelvi modellek SWOT analízise	58
4.	NAGY NYELVI MODELLEK OSINT ESZKÖZÖKBEN	62
4.1.	LLM OSINT PoC	62
4.2.	InfoHound és Ollama	65
4.3.	NexusXplore OSINT Research Assistant (ORA)	69
4.4.	Autonóm LLM hack agent (ALLMHA)	69
5.	ÖSSZEFOGLALÁS	72
5.1.	Összegzés és konklúzió	72
5.2.	Az OSINT-LLM keretrendszerek azonosított típusai és jellemzői	74
5.2.1.	<i>Szolgáltatás alapú</i>	75
5.2.2.	<i>Helyben integrált</i>	77
5.3.	Az OSINT-LLM megjelenése a Cyber Kill Chain szakaszaiban	78
5.4.	Hipotézisek bizonyítása	80
5.5.	Eredmények	82
5.6.	Az LLM további várható megjelenése az OSINT tevékenységben	82
5.7.	OSINT-LLM tevékenység elhárítását támogató javaslatok	84
5.7.1.	<i>Kollektív elhárítás</i>	84
5.7.2.	<i>Defenzív elhárítás</i>	85
5.7.3.	<i>Offenzív elhárítás</i>	86
6.	A KUTATÁS ALKALMAZÁSÁNAK TERÜLETEI ÉS TOVÁBBI IRÁNYAI	86
7.	RÖVIDÍTÉSEK JEGYZÉKE	88
8.	IRODALOMJEGYZÉK	89

1. BEVEZETÉS

A digitalizáció korában élünk, ami folyamatosan átformálja társadalmunk területeit. Az életünk jelentős eseményei egyre inkább átkerülnek a fizikai világból a kibertérbe, továbbá szoros kapcsolatot alakítanak ki azzal. E párhuzamosan kialakuló folyamat során a kibertérben végrehajtott tevékenységek már **számottevő hatást gyakorolnak a fizikai világra**. Ennek eredményeként a modern társadalomban a kiberbiztonság és az információbiztonság **kulcsfontosságú területekké váltak**. A kibertéri adatmennyiség rohamos növekedése, amelyet a „big data” jelenségként azonosíthatunk, forradalmasította az adattárolás, az adatelemzés és az adatvizualizáció technológiáit. Ezek az adathalmazok alkotják a **gépi tanulás és egyéb mesterséges intelligencia alapú technológiák** fundamentumát is.

A mesterséges intelligencia, különösen a **2022-ben bemutatott ChatGPT** markáns változásokat indított el társadalmunk állami és gazdasági aspektusaiban. Képességei és népszerűsége **alapjaiban rengette meg** a technológiai vállalatok **mesterséges intelligenciával kapcsolatos politikáját**. A technológia eredményeinek elismerésével egyidejűleg azonban **adatvédelmi és biztonsági kérdések is felmerültek**. A ChatGPT 2023 nyarán végrehajtotta **Microsoft Bing integrációját**, amely lehetővé tette a mesterséges intelligencia számára, hogy közvetlenül az internetről hajtson végre információgyűjtést, amely még **tovább fokozta ezeket az aggályokat**. Ezzel a fejlesztéssel tulajdonképpen **egy hírszerzési eszközt** adott az egyszerű felhasználók kezébe.

A digitalizációval párhuzamosan a hírszerzési formák között egyre népszerűbb a **nyílt forrású információgyűjtés** (Open Source Intelligence – OSINT), amely részletes és aktuális információk birtokába juttatja művelőit. A mesterséges intelligencia, jelen esetünkben a nagy nyelvi modellek képessége valószínűleg **transzformációs hatást váltott ki** a nyílt forrású információgyűjtésre. Meglátásom szerint a jövőben jelentkező **változások még érdekibbek lesznek** ezen a területen, és az ezzel szembeni **védelmi nézeteket is forradalmasítani fogja**. E tudományos dolgozat célja, hogy a két tématerület, vagyis a nyílt forrású információgyűjtés és a nagy nyelvi modellek ismertetésével, valamint azok **kölcsönhatásának** vizsgálatával átláthatóságot és racionalitást biztosítva támassza alá ezt a hipotézist.

1.1. A vizsgált téma bemutatása

A természetes nyelvfeldolgozás (nature language process, a továbbiakban: NLP) legeredményesebb technológiai módszere, **a nagy nyelvi modellek alkalmazása** nem egy újkeletű technológia, már **régóta az emberi tudás része**. Azonban egy megjelenő új architektúra, **a figyelemközpontú transzformátor hatására óriási fejlődésnek indult** a terület. A 2022-ben bemutatott transzformer alapú ChatGPT sikereinek köszönhetően pedig **jelentős forrásokat vonzott be a fejlesztések ezen ága**, így napjainkban és a jövőben is számos olyan nagy nyelvi modell megjelenésre számíthatunk, **amelyek eredményesebbek lesznek** a jelenleg ismert megközelítésektől. A nagy nyelvi modellek **rugalmassága** adja meg a lehetőséget arra, hogy egyre több és több területen (pl.: orvosi diagnosztika, számítógépes programozás stb.) jöhessen létre olyan **nagy nyelvi modell specifikáció**, amely újra formálja az eddigi folytatott gyakorlatokat és eljárásokat. E területek alól **a kiberbiztonság sem kivétel**.

A nyílt forrású információgyűjtés a **kiberbiztonság megteremtésének fontos részét képezi**, hiszen a kibertámadások előkészületeinek fontos aspektusa, így **védelmi szempontból** is ajánlatos vizsgálni legújabb képességeit. A dolgozat címében jelölt téma jól körülhatárolt, a nagy nyelvi modellek és a nyílt forrású információgyűjtés **„aktuális tudásmetszete”** jól megfogalmazható **tudományos problémaként**. Megítélésem szerint e két téma közül bármelyik önálló vizsgálata is aktuális eredményeket hozna felszínre, így hát kifejezetten nehéz helyzetben vagyok akkor, amikor egy tudományos dolgozat korlátjai közé kell szorítanom a két tudáshalmaz unióját.

1.2. A kutatás célja

A kutatás alapvető célja egy olyan két témát átfogó és harmonizáló információbázis összeállítása, amely a későbbiek során tudományos publikációkká formálva **gazdagítja a hazai szakirodalmi művek tárá**t és **tudományos értékével segíti a téma művelőit a következő időkben**. Speciális célja azonban többértű.

Egyrészt megpróbálja azonosítani a nagy nyelvi modellek helyzetét és értelmezését a **kibertéri, főként támadó szándékú OSINT tevékenységekben**, valamint ezzel kapcsolatban olyan megállapításokat tenni, amelyek **hozzájárulhatnak a hazai kiberbiztonság hatékonyságának növeléséhez**. Másrészt a témákban való részletes elmélyülés, valamint a legújabb tudományos

eredmények figyelembevétele alapján olyan megállapítások megtétele, amelyek **a jövőben megjelenő vektorokat és változásokat igyekeznek azonosítani**, ezzel úgy szint hozzájárulva a proaktív kiberbiztonsági politika megteremtéséhez.

1.3. Hipotézisek

A dolgozat négy hipotézist állít és kíván vizsgálni:

H1: A nagy nyelvi modellek OSINT eszközként értelmezhetőek.

Néhány nagy nyelvi modell (pl.: ChatGPT, Gemini) a szolgáltatók által megadott információk alapján képes az interneten történő valós idejű információkeresésre. Ez megteremti a lehetőséget arra, hogy az egy önálló, nyílt információgyűjtő eszközként értelmezzük azt. Ennek bizonyítására törekszek.

H2: A nagy nyelvi modellek bővíthetik a kibertéri OSINT tevékenységek információfeldolgozási határait.

Ebben a hipotézisben az LLM információfeldolgozási hatásait vizsgálom az OSINT alkalmazás tükrében. Arra keresem a választ, hogy képes-e az LLM feloldani az OSINT-ban rejlő korlátozottságokat.

H3: A nagy nyelvi modellek hatékonyabbá teszik a kibertámadásokkal kapcsolatos OSINT tevékenységet.

Ez a hipotézis a nagy nyelvi modellek és az OSINT kibertámadásban való együttes alkalmazásának megjelenését vizsgálja, azonosítani kívánja azokat a speciális jellemzőket, amelyek ebben a használati módban érvényesülnek.

H4: Léteznek olyan elhárítási technikák, amelyek csökkenthetik a nagy nyelvi modellekkel végzett nyílt forrású információgyűjtés kiberbiztonsági kockázatait.

A nagy nyelvi modellek olyan forradalmi technológiák, amelyek jelentős változásokat hozhatnak az OSINT és a kiberbiztonság világában. E változások negatív kiberbiztonsági következményeire válaszként a kockázatsökkentő megoldásokat kívánok azonosítani.

1.4. Módszertan

A címben szereplő kettősség egy **újszerű megközelítés**, és nemzetközi tudományos berkekben is **kevésbé kutatottnak mondható**. Az OSINT és a nagy nyelvi modellek önálló területként, tudományos szempontból azonban népszerű területek. A témaegyüttes klasszikus és hatékonynak mondható **kvantitatív eszközökkel történő megközelítése** számomra nehezen értelmezhető. Ugyanakkor újszerűsége és innovatív jellegének köszönhetően a **kvalitatív módszertanok közül a dokumentumelemzés** módszertana alkalmas a témával kapcsolatos információk kinyerésére. Elgondolásom alapján erre a módszertanra helyezve a kutatás legnagyobb hangsúlyát – **a lehető legszélesebb horizonton vizsgálódva** – értékes eredményeket várhatunk.

A témában íródott szakirodalmak felkutatására irányuló keresések a **Publish or Perish 8** akadémiai szoftverprogrammal történtek olyan **tudományos adatbázisokban**, mint a **Google Scholar**, **Scopus**, **Web of Science** archívumai, valamint a nem publikált, de a legújabb tudományos eredményeknek otthont adó **arXIV** nyílt forrású archívumban. A legújabb gyakorlatban megjelenő nyilvánosan közzétett fejlesztések megismeréséhez a **GitHub fejlesztői platformon tárolt projektek kerületek vizsgálat alá**. Leginkább a nagy nyelvi modell integrációt tartalmazó nyílt forrású információgyűjtési eszközök, valamint ezek jellemzői és kapcsolatai kerülnek elemzésre.

A dolgozat első felében bemutatásra kerül a **nyílt forrású információgyűjtés tevékenysége**, valamint a kibertámadások során alkalmazható OSINT eszközök köre. Hasonlóan részletezésre kerül a mesterséges intelligencia, a természetes nyelvi feldolgozás, illetve a nagy nyelvi modellek szerkezete, képességei. Utóbbi esetében a **legújabb modellekből** és további **specifikumokból példálózunk** annak érdekében, hogy átfogó képet kapjunk az alkalmazási jellemzőkről. **SWOT analízis** alá lesz vetve mindkét tématerület, melyeknek eredményei felhasználásra kerülnek a későbbiekben.

A dolgozat eredményei között a **szintézis módszertanát** alkalmazva a feltárt és összeállított információbázis elemzése, ezzel összefüggő megállapítások megtétele, illetve a **jövőben várható irányok** megfogalmazása történik meg. A dolgozat megkísérli azon **stratégiai elemek** azonosítását, amelyek integrálhatók a különböző szervezetek **védelmi politikájába** annak érdekében, hogy a nagy nyelvi modellekkel végzett nyílt információgyűjtés okozta kockázatok kezelésére intézkedhessenek.

2. A NYÍLT FORRÁSÚ INFORMÁCIÓSZERZÉS (OSINT)

A kutatások során fontos szerepet töltenek be a fogalmi meghatározások, hiszen ezek biztosítják az alapokat, amelyre a kutatás épül és ezek alkalmazásával biztosíthatjuk a kommunikációs egyértelműséget. A következőkben a témaegyüttes első szereplőjének, a nyílt információgyűjtésnek alapvetéseit, formáit, jellemzőit mutatjuk be. Továbbá, a hipotézisek vizsgálatának okán a kibertámadásban betöltött szerepére is szükséges kitérünk, valamint a két terület átvezetése érdekében bemutatásra kerül a mesterséges intelligencia kapcsolata is.

2.1. Fogalmi meghatározások

A görcső alá vett koncepció egyik felének (a nyílt forrású információszerzésnek) a definiálása egyszerű feladatnak tűnik, ugyanakkor **mélyebb részleteiben nem egyértelmű**. Egyszerű abban a tekintetben, hogy a korábbiakban tudományosan megfogalmazott gondolatok vagy jogszabályi értelmezések felhasználása **nem jelent különösebb kihívást**, ám nem egyértelmű, mert a fogalmakban megjelölt tevékenységek, módszerek és eszközök a gyakorlati megvalósulás során **helyenként elvesztik élesen megfogalmazott határaikat**, amelyekkel így összefüggésben **konfliktusok is megjelennek**.

Az alábbi fejezet rész nem kívánja kritikusan és részletekbe menően vizsgálni a nyílt forrású információszerzést, de egy rövid és praktikus bemutatást kíván adni annak – a kutatás szempontjából releváns – érintettségeiről. A téma tulajdonképpen az információ értéke körül forog, amellyel szemben annak védelmére és megszerzésére irányuló célok is megfogalmazódnak a mindennapokban. A dolgozat törzsanyaga főként utóbbival foglalkozik, de **eredményei defenzív célzatúak**. Ezzel a szakirodalomban nincs egyedül, hiszen tudományos szempontból az információszerzés önmagában is, de **védelmi irányultsággal való feldolgozása** egyértelműen népszerű téma.

2.1.1. Az információ és a nyílt információ

Az **információ** fogalmával jogszabályi szinten foglalkozik az állami és önkormányzati szervek elektronikus információ- biztonságaról szóló 2013. évi L. törvény, miszerint: *„bizonyos tényekről, tárgyokról vagy jelenségekről hozzáférhető formában megadott tapasztalat, megfigyelés vagy*

ismeret, amely valakinek a tudását, ismeretkészletét, annak rendezettségét átalakítja, megváltoztatja, illetve befolyásolja, továbbá bizonytalanságát csökkenti vagy megszünteti.”¹

A **nyílt** jelző alatt azokat **az elérhető forrásokat** kell érteni, amelyekhez a hozzáférés legálisan, törvénybe nem ütköző eszközökkel történik, jellemzően ilyen az előfigyelés, a megfigyelés, a személyes megkeresése, megvásárlás, és az internetes kereséssel összeállított információk köre. A nyílt információforrások fajtái sokrétűek, de napjainkban **a világháló nyílt információk legnagyobb tárháza**, amelynek további részleteire még kitérünk.²

2.1.2. A hírszerzés formái

A múltba visszatekintve az információ megszerzése, annak hagyományos jellegét tekintve leginkább a különféle hírszerző szervek feladata volt, amely során **nyílt és titkos eszközöket egyaránt igénybe vettek**. Az utóbbi évtizedekben a hírszerzés forráskörnyezetei jelentősen átalakultak, és ez egyértelmű kihatással járt a hírszerzési tevékenységek megjelenésére is. A tevékenységnek számos célja lehet, a közeli múltban például a terrorizmus, a nemzetközi bűnelkövetés, gazdasági és ipari hírszerzés, a migráció volt az. Manapság azonban a technológiai fejlődésnek köszönhetően **robbanásszerűen megnőtt a kibertérbe helyezett információs források** mennyisége.³ Ez a dolgozat témáját tekintve erre fókuszál, azonban érdemes megemlíteni a főbb hírszerzési formákat, ezek szemléltetésére a Kovács által alkalmazott átfogó ábra megfelelő:⁴

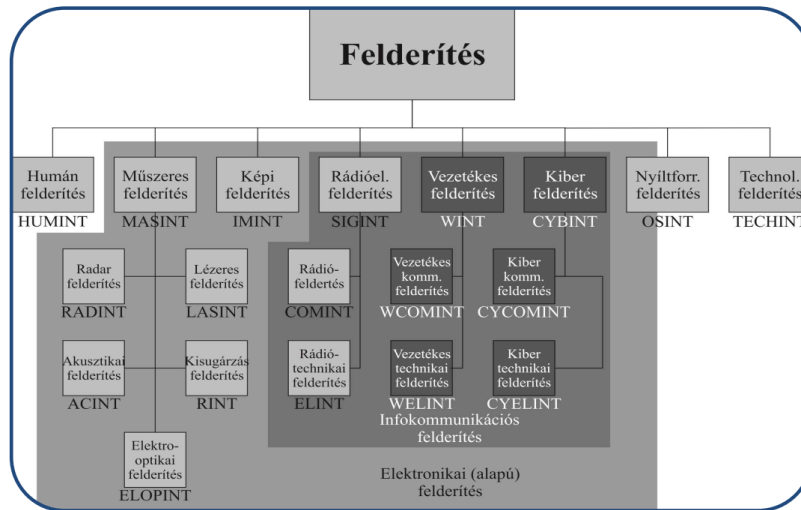
¹ 2013. évi L. törvény az állami és önkormányzati szervek elektronikus információbiztonságáról 1.§ (1) bekezdés 25. pontja.

² Deák Veronika (2018). A nyílt forrású információszerzés szerepe a kibertámadások végrehajtása során. Hadmérnök, 13(3), 391–402. Elérés: <https://folyoirat.ludovika.hu/index.php/hadmernok/article/view/3859> p. 3.

³ Tóth Tivadar (2001). Katonai szervezetek és a nyílt forrású információgyűjtés. Repüléstudományi közlemények XIII. évfolyam 33. szám. Elérhető: https://real-j.mtak.hu/20046/2/rtk_2001_13_2_.pdf#page=66 p. 66.

⁴ Kovács Zoltán (2015): A titkosszolgálatok és az információgyűjtés jövőbeli lehetséges irányai, új technológiák megjelenése és hatásai a titkosszolgálati tevékenységekre in: Boda József. A nemzetbiztonság technikai kihívásai a 21. században. ISBN:9786155527746. p.111.

1. ábra: a felderítés területei. Forrás: Kovács (2015) p.111.



Az Amerikai Egyesült Államok katonai fogalomtára szerint a teljes spektrumú hírszerzési tevékenység (All-Source Intelligence – ASI) kategorizálása az alábbiak szerint történik:

1. Emberi ügynökökön alapuló (Human Intelligence – **HUMINT**);
2. Technikai eszközökön alapuló (Technological Intelligence – **TECHINT**);
3. Nyilvános forrásból történő (Open Source Intelligence – **OSINT**).⁵

2.1.3. A nyílt forrású információgyűjtés (OSINT)

A **nyílt forrású információszerzés (Open Source Intelligence)** terjedelmes tudományos, legfőképpen nemzetbiztonsági és katonai hírszerzésű perspektívából kiinduló irodalmi bázissal rendelkezik, hiszen amellet, hogy már jelentős történetiséggel bír⁶, jelen korunk digitális transzformációja egyre tovább növeli értékét. Ugyanakkor észlelhető, hogy az eddigi elvárásokat kiszolgáló hírszerzési szükséglet kiszolgálása mellett alkalmazási köre kiszélesedett, és pár éve **már kilépett az állami alkalmazás privilégiumából**. Ez a folyamatos és több területet érintő terjeszkedés a témában korábban megfogalmazott definíciós célú gondolatokat is próbára teszi. Az

⁵ Dictionary of United States Military Terms for Joint Usage, Departments of the Army, Navy, and Air Force, Washington, D.C., May 1995, p. 53.

⁶ Vadász Pál (2018). A szemantikus keresés módszerei és alkalmazási lehetőségei a védelmi szférában, a közigazgatásban, illetve a gazdasági életben. Doktori (PhD) értekezés. Nemzeti Közszolgálati Egyetem Katonai Műszaki Doktori Iskola. Budapest, 2018. p. 64.

OSINT témában íródott hazai szakirodalmat szerzők, például Bányász (2015)⁷, Bonnyai és társai (2023)⁸, Csizner (2019)⁹, Solti (2019)¹⁰, Vadász (2018)¹¹ a téma definiálására a Lévay Gábor általi megfogalmazást használják fel, miszerint:

„Az OSINT a katonai felderítés és a hírszerzés rendszerén kívül létező, a publikum (tehát minden egyén) számára nyilvánosan, legális eszközökkel megszerezhető, vagy korlátozott körben terjesztett, de nem minősített adatok szakmai szempontok alapján történő felkutatását, gyűjtését, szelektálását, elemzését-értékelését és felhasználását jelenti.”¹²

Elismerésre méltó, hogy Lévay közel 20 éves magyarázata a jelentősen változó környezet ellenére is időtállóan bizonyul, azonban kritikával illetném abban a tekintetben, hogy álláspontom szerint az OSINT manapság nem a hírszerzés rendszerén kívül létezik, hanem annak szerves, önálló ágát képező része, egyéb iránt aktuálisnak tekinthető a definíció. Erre az álláspontra erősít rá az a megfogalmazás, amelyet Csizner tanulmányában definíciós céllal is idéz a CIA¹³ egykori igazgatójától, Michael Hayden-től, mégpedig az alábbiak szerint:

„A források nyilvánosak. Az érdeklődési körünk nem az. A szakmánkban nem mindig a titkos információk a legértékesebbek. Valójában, sokkal élvezetesebb megoldani egy problémát vagy megválaszolni egy nehéz kérdést olyan információk alapján, amit valaki meggondolatlanul nyilvánosan is hozzáférhetővé tett”.

Úgy vélem, az első fogalom egyszerűen jellemzi magát a folyamatot annak területi határaival, eszközszerével és módszertanaival, míg a második értelmezés inkább a mögöttes filozófiát és

⁷ Bányász Péter (2015). A közösségi média, mint a nyílt forrású információszerzés fontos területe. Nemzetbiztonsági Szemle, 3. sz. (2). p. 23.

⁸ Bonnyai Tünde, Kiss Adrienn, Tóth András (2023). Nagyvállalati biztonságtudatosság és nyílt forrású információszerzés: Kiber- és információbiztonsági tanulmányok, Budapest, Nemzeti Közszerzési Egyetem Ludovika Egyetemi Kiadó. p.84.

⁹ Csizner Zoltán (2019). Az OSINT határai. Nemzetbiztonsági Szemle 7. évfolyam 2. szám, Budapest. <https://doi.org/10.32561/nsz.2019.2.2> p. 20.

¹⁰ Solti István (2019). Az OSINT információgyűjtő eszközeiről. Nemzetbiztonsági Szemle, 7. évfolyam 2. szám. Budapest. DOI: <http://doi.org/10.32561/nsz.2019.2.1> . p.6.

¹¹ Vadász (2018). p.65.

¹² Lévai Gábor (2006). OSINT (Open Source Intelligence) – Nyílt információs hírszerzés. Egyetemi jegyzet, Budapest, Zrínyi Miklós Nemzetvédelmi Egyetem, 2006. p.6.

¹³ Central Intelligence Agency – Központi Hírszerzési Ügynökség (USA).

érzelmi motivációt kívánja megértetni az olvasóval. A két körülírás együttesen már megadja azt értelmezést, amelyre a dolgozat gondolatmenetének vezetése szempontjából szükség van.

2.2. Az OSINT főbb jellemzői

A nyílt forrású információszerzésnek számos olyan tényezője (források, módszerek) van, amelyek meghatározzák tevékenységét és biztosítják önálló tevékenységként való értelmezését. A következő alfejezetekben erre térek ki.

2.2.1. Források és módszerek

Manapság az OSINT több területen jelen van, hiszen a katonai, a nemzetbiztonsági és rendészeti szervek mellett az **üzleti szférában is önálló képességek alakultak ki**.¹⁴ Az állami szervek hírszerzési igényeit a jog által szolgáltatott eszközrendszerrel szolgálják ki, míg utóbbinál egyéb megjelenési módokat találunk. Az üzleti vállalkozások **főként a konkurencia elemzésére**, az újságírók és oknyomozók leginkább **tényfeltárássra**, míg a pénzügyi intézetek **ügyfélazonosításra és a kockázatelemzéshez szükséges információk** feltárássára alkalmazzák azt.¹⁵ Az OSINT forrásait tekintve az azokkal szemben felállított legfőbb **kritérium az, hogy nyíltak legyenek**, tehát bárki számára elérhetőnek. Ezek felsorolását több szakirodalom (pl.: Vida 2013¹⁶, Solti 2019¹⁷) a NATO nyílt forrású adatszerzéssel foglalkozó kézikönyvének¹⁸ tartalmával szemlélteti, amelyek az alábbiak:

OSINT források:

- Hagyományos (nyomtatott és elektronikus) média;
- internet;

¹⁴ Dobák Imre (2019). OSINT – Gondolatok a kérdéskörhöz. Nemzetbiztonsági Szemle 7. évfolyam 2. szám. Budapest. <http://doi.org/10.32561/nsz.2019.2.7> p. 3.

¹⁵ Precognox (2022.03.08.). A hírek és információk megszerzése még nem hírszerzés. Elérhető: <https://precognox.hu/blog/a-hirek-es-informaciok-megszerzese-meg-nem-hirszerzes/> (letöltve: 2023.12.22.)

¹⁶ Vida Csaba (2013). Nyílt forrású adatszerzés (OSINT) In: Kobilka István (szerk.): Nemzetbiztonsági alapismeretek, Nemzeti Közszerzési és Tankönyv Kiadó Zrt. ISBN 978-915-5344-32-9. p. 115-124.

¹⁷ Solti István (2019). Az OSINT információgyűjtő eszközeiről. Nemzetbiztonsági Szemle, 7. évfolyam 2. szám. Budapest. DOI: 10.32561/nsz.2019.2.1. p. 3-18.

¹⁸ NATO Open Source Intelligence Handbook (2001). Elérhető: <https://github.com/lawsecnet/OPSEC/blob/master/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf> (letöltve: 2023.12.22.)

- kereskedelmi online szolgáltatók;
- „szürke irodalom”;
- nyílt események (konferenciák, előadások);
- személyes tapasztalatok (szakértők, megfigyelők stb.);
- kereskedelmi műholdak felvételei;
- tudományos-kutató szervezetek.

A NATO által megfogalmazott felsorolás lassan már negyed évszázadot megélt, így már csak az eltelt idő okán is esedékes lenne újragondolása a jelenleg ismert társadalmi viszonyok tükrében. Jómagam inkább érték egyet azzal az egyszerű megközelítéssel, amelyet Vadász Pál (2018) jelöl meg doktori értekezésében, miszerint a nyílt információkeresés forrásai kétfelé bonthatóak:

- **Klasszikus** (pl.: nyomtatott sajtó, könyvtár, irattár, tv, rádió stb.).
- **Elektronikus** (pl.: a klasszikus források digitális formában elérhető változatai, ingyenes és fizetős adatbázisok, közösségi platformok, honlapok, műholdak adatai, online keresőrendszerek, online térképek, közösségi helyek forgalmi adatai stb.).¹⁹

A NATO felsorolástól korszerűbb, ugyanakkor egy kibertámadás szemszögéből kiinduló megközelítésnek tekinthető Deák (2018)²⁰ összefoglalása az OSINT forrásokról:

- Telefonos információgyűjtés;
- hagyományosan publikált anyagok;
- megfigyelés;
- **internetes információgyűjtés;**
- kukaátvizsgálás;
- rádió- és televíziós adások, média hírek;
- személyes megkeresés.

Deák (2018) felsorolásának kibertámadásban való relevanciáját nem vitatom. Azonban álláspontom szerint napjaink egyik leghatékonyabb hírszerzési felületévé – a digitális transzformációnak köszönhetően – az **internet** vált. Erre a megállapításra, valamint az arra való

¹⁹ Vadász (2018).

²⁰ Deák (2018). p. 9-11.

tekintettel, hogy a dolgozat tartalmi korlátjainak tartása okán a téma pontos lehatárolása szükséges, továbbá azért, mert a dolgozat a nagy nyelvi modellek OSINT képességeit a kibertámadásokkal összefüggésben vizsgálja, a következőkben ebből felsorolásból az **internetes információgyűjtés témája kerül további részletezésre.**

2.2.2. Az internetes nyílt forrású információgyűjtés

A világhálón történő nyílt információgyűjtéssel összefüggésben fontos tudni, hogy az internetnek több olyan rétege is létezik, amelyekről az átlag felhasználóknak tudomásuk sincs. Az általuk érzékelt, úgynevezett **felszíni web (surface web)** csupán a webes tartalmak **4%-át** jelenti, a fennmaradó **96%** a „**deep web**”-en található, amelyek a hagyományos keresőmotorokon keresztül nem érhetőek el. Ennek még apróbb része a „**dark web**”, amely a becslések szerint a teljes webes tartalom **6%-a**. Az itt található weboldalak már csak olyan **speciális eszközökkel és konfigurációs beállításokkal** érhetőek el, amelyeket az online tevékenységek elrejtésére terveztek. Ezek az oldalak legfőképpen az **illegális kereskedelemnek** adnak otthont, illetve olyan nem illegális tartalomnak, amelyeknek **legfőbb mozgatórugója az anonimitás** megőrzése.²¹

A világhálón végrehajtott nyílt forrású információgyűjtésre a terminológia hazai és nemzetközi viszonylatban számos kifejezést alkalmaz, a hírszerzési tevékenység forrásait és módszereit tekintve. Több helyen megjelenik a **web intelligence (WEBINT)** kifejezés, amely a webes forrásokból történő hírszerzést jelenti, valamint a **CYBER-HUMINT** tevékenységre utalás, ami az emberi erővel és forrásból történő hírszerzés (**HUMINT**) kibertérben alkalmazott formája. Ide sorolható még a forrásként közösségi oldalakat felhasználó (**SOCMINT**) információgyűjtő tevékenység is. Tekintettel arra, hogy azonban a lehatárolt kutatási témát a felsoroltak közül egyik terület sem fedi le maradéktalanul, valamint arra a tapasztalásra, hogy a kutatás fókuszába került legújabb tudományos munkák **egységesen az OSINT megnevezést** alkalmazzák, így talán az a legcélszerűbb, ha itt sem térünk el ettől a gyakorlattól.

²¹ Kim Rosengren (2022). Contribution of Open-Source Intelligence to Social Engineering Cyberattacks. Bachelor's Thesis. Turku University of Applied Sciences. Finland. p. 24.

2.2.3. Az aktív és a passzív felderítés

Egy cél érdekében végzett felderítési cselekvés lehet **aktív** vagy **passzív**. A kettő közötti differencia a célszeméllyel vagy a rendszerrel való érintkezésben, valamint az alkalmazott módszertanokban van. Előbbi esetében a gyors eredmény reményében **közvetlen kapcsolat** jön létre, ám a közvetlenség **magas kockázatot** jelent a **rejtve maradási jellemző megőrzésére**. Ezzel a módszertanokkal olyan információkat gyűjthetünk, amelyek specifikusabbak, tehát eredményesebben szolgálják a célokat. A kibertérben ilyen módszerek pl.: a social engineering²², a fiktív közösségi profillal történő információszerzés, a portscannerek és a penetrációs tesztlők eszköztárában lévő sebezhetőség-ellenőrző szoftverek alkalmazása.

A passzív esetben egyfajta **közvetett interakció** jön létre, ahol nem fontos szempont a rövid időn belüli teljesülés, így **mélyebb vizsgálatot** tesz lehetővé és a tevékenység észlelésének **kockázata is alacsonyabb**. Ez a cselekvési forma a **nyílt információk begyűjtésére** irányul. Az ebbe a csoportosításba sorolható kibertéri módszerek például a szervezeti IT infrastruktúra és jellemzőinek feltérképezése, a keresőmotorok által indexelt információk felhasználása.

A valódi OSINT tevékenység véleményem szerint passzív, de tapasztalatom alapján a szakirodalmi állásfoglalás sem egységes ebben a tekintetben, hiszen több esetben bontják magát az OSINT-et is aktív és passzív lehetőségekre. Legtöbbször az eszközök több különböző (aktív-passzív) felhasználási módja okán sem sorolható egyértelműen valamelyik csoportba, tehát az OSINT eszközöket sem tudjuk maradéktalanul klasszifikálni.

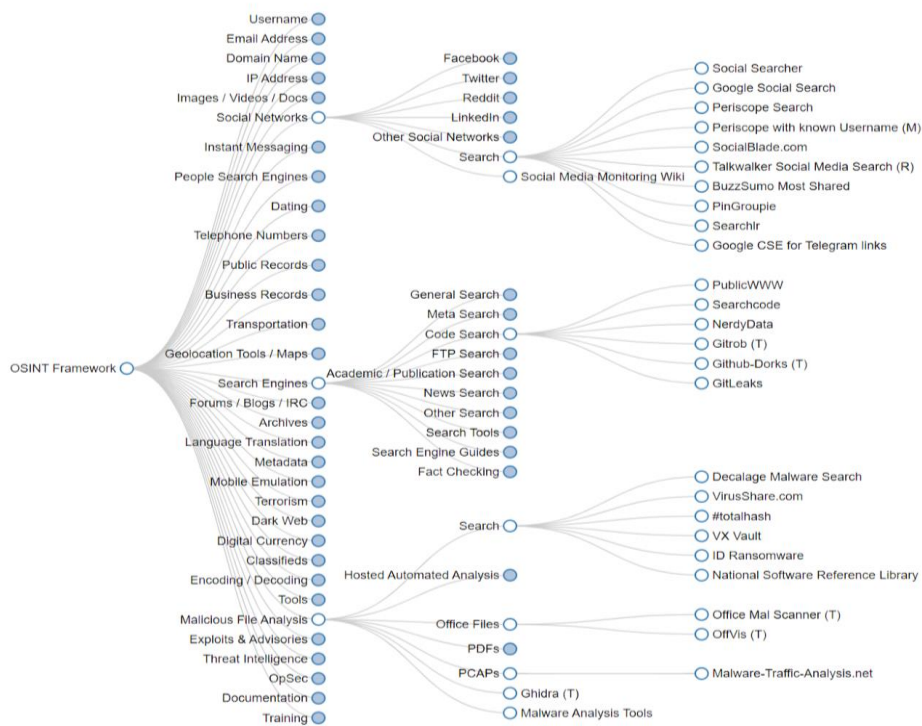
2.2.4. Az „internetes OSINT eszközök”

A OSINT tevékenységhez számos eszköz (pl.: szoftver, webalkalmazás stb.) létezik, amelyek megismerésére **az információbiztonsági indíttatásból** létrejött **OSINT Framework**²³ megfelelő platform. Bár a keretrendszer fő funkciója az ingyenes eszközök és ingyenes OSINT erőforrások felhasználása, jól funkcionálhat a szervezeti kiberbiztonság megteremtésében is.

²³ OSINT Framework. Elérhető: <https://osintframework.com> (letöltve: 2024.01.12.)

Az 2. ábrán láthatjuk a különböző célokhoz rendelt OSINT eszközök ábrázolását, amelyek biztosítják a forrásban elérhető információk célzott és hatékony megszerzését. Hasonlóképpen próbálkozik összefogni az OSINT eszközök színes skáláját a svájci **i-Intelligence** vállalat által készített **Open Source Intelligence Tools and Resources Handbook**, vagyis a nyílt forrású információgyűjtési eszközök és források kézikönyve, amelyet nyomozók és információbiztonsági szakemberek támogatására hoztak létre elsődlegesen.²⁴ Ebben a kézikönyvben is OSINT eszközök csoportokba rendezett listáját találjuk. E két forrás kiváló eszközkészletet biztosít az OSINT műveletekhez.

2. ábra. Az OSINT Framework eszközrendszereinek vizuális ábrázolása.
 Forrás: <https://osintframework.com/>



2.3. Az OSINT megjelenése a Cyber Kill Chain szakaszaiban

A nyílt forrású információgyűjtés a kiberbiztonsági megközelítésből kétpólusú. Egyértelműen kategorizálható, hiszen **fenyegetést jelent** az elektronikus információbiztonságra, ugyanakkor az

²⁴ Open Source Intelligence Tools and Resources Handbook Elérhető: <https://i-intelligence.eu/resources/osint-toolkit> (letöltve: 2024.03.21)

által okozott kockázatok redukálására – **mint eszköz – egyaránt alkalmazható**. A rosszindulatú OSINT elleni védelem hatékony megvalósításához elemeznünk kell a tevékenység jellemzőit és követnünk kell fejlődését.

Deák (2018) tanulmányában három részre bontja a kibertámadásokhoz szükséges információkat:

1. Informatikai rendszerre vonatkozó;
2. célpontra vonatkozó;
3. tervezési, szervezési, végrehajtási információkra.²⁵

Álláspontom szerint bizonyos megközelítésben ez is helytálló, de a dolgozat hipotéziseinek vizsgálatához és bizonyításához alkalmasabb a **Lockheed Martin**²⁶ vállalat által meghatározott kibertámadási lánc vagy láncolat (angoul: **Cyber Kill Chain**²⁷, a továbbiakban: CKC) modellje. A **hét szakaszt** (fázist) tartalmazó megközelítés felhasználása támogatja a kiberbiztonsági szakembereket abban, hogy a kibertámadás folyamatának minden önálló részével szemben megfelelő védelmi intézkedéseket (pl.: megakadályozás, a károk mérséklése) hozhassanak. A CKC szemlélet alkalmazása egyértelműen **javítja a támadás átláthatóságát és segíti elemzési tevékenység hatékonyságát**, ezeket az előnyöket használjuk ki vizsgálatunkhoz is.²⁸

A Cyber Kill Chain fázisai:

1. **Felderítés (reconnaissance)**: A támadó felderíti a célponthoz tartozó olyan információkat, amelyekre ki lehet dolgozni a támadás részleteit. Ilyenek lehetnek például az IP-címek, elektronikus levelezési címek, hálózati információk stb.
2. **Fegyverkezés (weaponization)**: Ebben a szakaszban a támadó a megszerzett tudás birtokában azonosítja a sebezhetőségeket és olyan rosszindulatú szoftvert készít, amely kihasználja azokat.

²⁵ Deák (2018).

²⁶ Egy amerikai globális biztonsági és repülési vállalat, amely széleskörű tevékenységet folytat, többek között a védelmi, biztonsági, információs technológiák területén.

²⁷ A „kill chain” az angol nyelvben egy katonai fogalom, amelyet a támadás szerkezetének azonosítására alkalmaznak. A kibertámadásokkal foglalkozó külföldi tanulmányok gyakran alkalmazzák a „Cyber Kill Chain” modellt a kibertámadási jelenségek különböző szempontú vizsgálata során.

²⁸ Cyber Kill Chain®. Elérhető: <https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html> (letöltve: 2024.01.05.).

3. **Szállítás (delivery):** A támadó rosszindulatú szoftvert eljuttatja a célszámítógéphez. Ez különböző módszerekkel történhet, mint például e-mail mellékletek, weboldalak, USB meghajtók stb. segítségével.
4. **Kihasztnálás (exploitation):** A rosszindulatú szoftver a sikeres szállítást követően a célrendszerben kihasználja sebezhetőséget annak érdekében, hogy valamilyen szintű hozzáférést szerezzen a célrendszerhez.
5. **Telepítés (installation):** A rosszindulatú szoftver a sebezhetőség sikeres kihasználása után telepíti magát a célrendszeren, és legtöbbször létrehoz egy úgynevezett „backdoor”-t²⁹ a támadó számára, amellyel előkészíti a következő fázis érvényesülését.
6. **Futtatás és vezérlés (command & control):** A támadó a backdoort alkalmazva létrehoz egy olyan csatornát, amelyen keresztül a rosszindulatú szoftver és a kompromittált rendszer távoli irányítására képessé válik.
7. **Célok elérése (actions on objectives):** Azon tevékenységek végrehajtása, amelyekkel a támadó eléri kívánt céljait. Ilyenek például az adatlopás, az adatmegsemmisítés vagy fájlok titkosítása pénzbeli követelésért.³⁰

Ki kell térni arra, hogy a CKC keretrendszere **nem fedi le maradéktalanul a kibertámadások típusainak teljes körét**, nem alkalmazható minden variáns elemzésére. A modellt kifejezetten több lépésből álló távoli, strukturált hálózati behatolásokra és célzott támadásokra (pl.: APT³¹) tervezték, az alacsonyabb szintű például egyszerűbb szkriptek, DDos, Ad-hoc jellegű, valamint a személyes végrehajtást igénylő támadások elemzésében nem hatékony. Továbbá – mivel a külső támadói tevékenységre összpontosít – a **belső fenyegetések** vizsgálatában sem célravezető.

Az OSINT **Cyber Kill Chain** modellben való értelmezése során nem csak egy átfogó képet kaphatunk a szakaszokban értelmezhető, **támadói célú OSINT** műveletekről, de ugyanakkor egyidejűleg azonosítható a kiberbiztonságot **támogató OSINT alkalmazási lehetőségek** köre. A következőkben forrásként az OSINT Framework keretrendszerét, valamint Yamin és munkatársai

²⁹ Angol nyelvű kifejezés, jelentése „hátsó kapu” vagy „hátsó ajtó”. Szoftverben (vagy ritkább esetben hardverben) létrehozott olyan funkció, amelynek segítségével az illetéktelen személyek hozzáférnek a programhoz, a számítógéphez vagy annak bizonyos részeihez.

³⁰ Cyber Kill Chain®.

³¹ Advanced Persistent Threat rövidítése. Ez egy kibertámadás típusa, amelyet magasan képzett és jól felszerelt támadók, gyakran állami csoportok vagy nagy szervezetek hajtanak végre.

tanulmányát³² felhasználva a CKC szakaszaiban megjelenő OSINT eszközök és alkalmazásai módszertanaikból példálózunk.

2.3.1. Felderítés (reconnaissance)

A világhálón megjelenő emberi tevékenységek úgynevezett digitális lábnyomokat hagynak maguk után, amelyek megfelelő forrást biztosítanak a CKC **felderítés** szakaszában érvényesülő OSINT tevékenységeknek. Bár az OSINT létjogosultságát (támadó vagy védelmi szándékkal egyaránt) minden CKC szakaszban értelmezhetjük, de azok közül a felderítési fázisban érvényesül a leghatékonyabban. A OSINT és a CKC felderítési folyamatának célja azonos, ami a kívánt információk megszerzése.

A felderítési szakasz során leggyakoribban megszerezhető nyílt információk a **domain részletek, tartományokkal és szerverekkel kapcsolatos adatok, operációs rendszerek és futó szolgáltatások verziói**, amelynek kinyerésére több OSINT eszköz (pl.: DNS Dumpster³³, domainsdb³⁴, ip2whois³⁵, recon-ng³⁶) létezik. Mindemellett jelentős mennyiségű információ szerezhető meg **archívumokból** is, hiszen a **szervezetek gondatlanságból vagy tévedésből közzétesznek érzékeny adatokat** úgy, hogy erről nincs tudomásuk. Ezek hiába kerülnek a későbbiekben eltávolításra, hiszen léteznek olyan szolgáltatások, **amelyek rendszeresen archiválják a nyilvánosan elérhető weboldalakon szerepeltetett információkat**. Az archivált adatbázisokban történő információk megszerzéséhez is léteznek OSINT eszközök (pl.: WayBackMachine³⁷, weleakinfo³⁸, cryptome³⁹).

³² Yamin Muhammad Mudassar, Mohib Ullah, Habib Ullah, Basel Katt, Mohammad Hijji, and Khan Muhammad. 2022. "Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security" Mathematics 10, no. 12: 2054. DOI: <https://doi.org/10.3390/math10122054>

³³ DNS Recon and Research, Find and Lookup DNS Records. Elérhető: <https://dnsdumpster.com> (letöltve: 2024.01.05.).

³⁴ Registered Domain Names Search — Find Registered Domain Names. Elérhető: <https://instantdomainsearch.com> (letöltve: 2024.01.06.).

³⁵ Free WHOIS – Domain Name Lookup. Elérhető: <https://www.ip2whois.com> (letöltve: 2024.01.05.).

³⁶ GitHub — lanmaster53/recon-ng: Open Source Intelligence Gathering Tool Aimed at Reducing the Time Spent Harvesting Information from Open Sources. Elérhető: <https://github.com/lanmaster53/recon-ng> (letöltve: 2024.01.06.).

³⁷ Internet Archive: Wayback Machine. Elérhető: <https://archive.org/web/> (letöltve: 2024.01.06.).

³⁸ We Leak Info — Data Breach Search Engine. Elérhető: <https://weleakinfo.io/> (letöltve: 2024.01.06.).

³⁹ Cryptome. Elérhető: <https://cryptome.org> (letöltve: 2024.01.06.).

A **közösségi oldalakon** megjelenő nyílt információk vonatkozhatnak akár a vizsgált magánszemélyre vagy az őt foglalkoztató szervezet tevékenységére. Az e források kezelésére alkalmas eszköz a Raven⁴⁰ és a Social Searcher⁴¹. Láthatjuk, hogy a felderítési szakasz alatti OSINT tevékenységhez bőven akad forrás és a hozzájuk rendelt eszköztár is rendelkezésre áll.⁴²

2.3.2. Fegyverkezés (weaponization)

Az első szakaszt követően a feltárt információkkal összefüggésben a sebezhetőségek azonosítása történik meg. A sebezhetőségek feltárása is **két részre klasszifikálható**.

Az **aktív megoldások** (pl.: sebezhetőségi szkennerek) tevékenysége **észlelhetővé válik** a szervezet számára, így az OSINT-tel szembeni elvárásoknak csak a **passzív megoldások felelnek meg**. E tevékenység során az **infrastruktúrára vonatkozó specifikumok megállapítása** történik meg a nyíltan elérhető online eszközökkel (pl.: Shodan,⁴³ Censys⁴⁴ stb.), majd ezt követően a releváns **sebezhetőségek azonosítására a megfelelő adatbázisok** (pl.: VulDB,⁴⁵ NVD⁴⁶, CNNCD⁴⁷) szolgáltatnak adatot. A fázis zárásaként a beazonosított sebezhetőségek kihasználására az úgynevezett „**exploit**”-tok is **meghatározhatók nyílt forrású adatbázisokból** (pl.: Exploit DB⁴⁸, Rapid7 Exploit DB⁴⁹, 0 day Today⁵⁰). A felsorolt eszköztípusok és a hozzájuk alkalmazott

⁴⁰ GitHub — 0x09AL/raven: A Raven egy információszerző eszköz, amely egy szervezet alkalmazottainak adatainak összegyűjtésére (pl.: profilozására) használható. Elérhető: <https://github.com/0x09AL/raven> (letöltve: 2024.01.06.)

⁴¹ Social Searcher – Free Social Media Search Engine. Elérhető: <https://www.social-searcher.com> (letöltve: 2024.01.06.).

⁴² Yamin, M.M.; Ullah, M.; Ullah, H.; Katt, B.; Hijji, M.; Muhammad, K. Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security. Mathematics 2022, 10, 2054. <https://doi.org/10.3390/math10122054> p. 6.

⁴³ A Shodan egy speciális keresőmotor, amely az internetre csatlakoztatott eszközöket és szolgáltatásokat keresi és indexeli. Elérhető: <https://www.shodan.io> (letöltve: 2024.01.05.).

⁴⁴ A Censys egy kiberbiztonsági kutatók által fejlesztett keresőmotor, amely az interneten található eszközök és szolgáltatások feltérképezésére és elemzésére specializálódott. Elérhető: <https://censys.com> (letöltve: 2024.01.05.).

⁴⁵ Vulnerability Database (VulDB). Elérhető: <https://vuldb.com> (letöltve: 2024.01.05.).

⁴⁶ Az Amerikai Egyesült Államok kormányának nemzeti sebezhetőségi adatbázisa. National Vulnerability Database. Elérhető: <https://nvd.nist.gov> (letöltve: 2024.01.05.).

⁴⁷ 国家信息安全漏洞平台. (kínai – magyar fordításban: Nemzeti Információbiztonsági Sebezhetőségi Platform) Elérhető: <https://www.cnnvd.org.cn> (letöltve: 2024.01.05.).

⁴⁸ Exploit Database-Exploits for Penetration Testers, Researchers, and Ethical Hackers. Elérhető: <https://www.exploit-db.com> (letöltve: 2024.01.05.).

⁴⁹ Vulnerability and Exploit Database. Elérhető: <https://www.rapid7.com/db> (letöltve: 2024.01.05.).

⁵⁰ 0day.today - Biggest Exploits Database and 0day market. Elérhető: <https://en.0day.today/faq> (letöltve: 2024.01.06.).

eljárásrend együttes, eredményes használata esetén a **támadó létrehozza a rosszindulatú szoftvert**. Ezt követően lép a támadás a következő fázisába.⁵¹

2.3.3. Szállítás (delivery)

Annak ellenére, hogy az információbiztonság iránti tudatosság növekszik, továbbra is az egyik **leginkább hatékony módszer marad a vállalati e-mail címeken keresztül** egy rosszindulatú szoftvekkel megfertőzni egy szervezet informatikai rendszerét. Egy exploit e-mailen történő szállításának címzettjének azonosítására kedvező alkalmazás a Hunter⁵². A szolgáltatást alkalmazva kilistázhatjuk az interneten a célszervezethez tartozó, **keresőmotorok által indexelt e-címeket**. Továbbá – amennyiben a szervezeti e-mail címet használták más weboldalon (pl.: fórum, vendégkönyv stb.) – információt kaphatunk a **felhasználó szokásairól, érdeklődési köréről** is, amely szintén felhasználható információvá válik egy támadás során. Létezik olyan webes alkalmazás (pl.: haveibeenpwned⁵³) is, amely az azonosított e-mail címet lekérdezve **visszajelzést szolgáltat** arra vonatkozóan, hogy az korábban **bármely adatvédelmi incidens során érintett volt-e**. Pozitivitás esetén, az információ birtokában a darkweben⁵⁴ történő keresés további információkhoz, a biztonság szempontjából **kritikus megállapításokhoz is vezethet**.⁵⁵ A **felhasználónevek azonosítása** szintén potenciális veszélyforrásnak minősül, hiszen az alkalmazottak alapvető web és közösségi médián folytatott tevékenysége is feltárható és felhasználható. Erre kiváló OSINT eszköz a NameCHK⁵⁶ vagy a ThatsThem⁵⁷. Nem elhanyagolható az általánosnak tekinthető megállapítás, hogy az informatikai berendezések többsége sajnos az **alapértelmezett konfigurációs beállításokkal vannak üzembe helyezve**, így az alapértelmezett fiókok a hozzájuk tartozó standard jelszavakkal **könnyű hozzáférést**

⁵¹ Yamin et al p. 7.

⁵² Hunter (Email Hunter). Elérhető: <https://hunter.io> (letöltve: 2024.01.07.).

⁵³ Have I Been Pwned. Elérhető: <https://haveibeenpwned.com> (letöltve: 2024.01.07.).

⁵⁴ A darkweb az internet egy rejtett része, amely speciális böngészőkön keresztül érhető el és nem kerül indexelésre a hagyományos keresőmotorok által. Ez a titkosított, anonim környezet mind törvényes, mind törvénytelen tevékenységeknek.

⁵⁵ Yamin et al p. 9.

⁵⁶ Namechk[Username, Domain, and Trademark Search]Username Registration. Elérhető: <https://namechk.com> (letöltve: 2024.01.07.).

⁵⁷ Find People for Free|Get Their Contact Info|ThatsThem. Elérhető: <https://thatsthem.com/> (letöltve: 2024.01.07.).

biztosítanak a rendszerekhez. Ehhez több nyíltan elérhető adatbázis (pl.: Default Password DB⁵⁸, Default-password,⁵⁹ Default Password Lookup⁶⁰) áll rendelkezésre.

A felsorolt példákban láthatjuk, hogy a szállítás szakaszában alkalmazható OSINT tevékenységek legfőképpen saját tevékenységünk digitális lábnyomait, vagy gondatlanságunk következményeit használják ki, amely kockázatok csak a **biztonságtudatosság szemléletének** alkalmazásával redukálhatók.⁶¹

2.3.4. Kihasztnálás (exploitation)

A rosszindulatú szoftverek ellen például a **szignatúrán** és a **viselkedés felismerésén** alapuló **vírusvédelmi** megoldásokkal védekezhetünk.⁶² Előbbi egy folyamatos szkenneléssel keresi a rosszindulatú szoftver jellemzőit a vizsgálat alá vett rendszeren. Ezek a szignatúrák **nyílt adatbázisokból is lekérdezhetők**, de **nemcsak a védelmi, hanem támadói célokat is képesek** szolgálni. A támadók előszeretettel alkalmazzák ezeket szolgáltatásokat (pl.: Virustotal,⁶³ Hybrid Analysis,⁶⁴ Ether⁶⁵) annak érdekében, hogy **teszteljék az általuk írt rosszindulatú szoftverek** aktív védelem alatti pozitívását. A tesztelési eredmények alapján úgy **modifikálják a rosszindulatú szoftvert**, hogy az észrevétlen maradjon olyan környezetekben is, ahol vírusvédelem található.⁶⁶

2.3.5. Telepítés (installation)

A már lefejlesztett rosszindulatú program futtatását a támadó szándékának megfelelően kell végrehajtani. Talán az ebben a fázisban értelmezett OSINT megoldások állnak a legtávolabb a klasszikus megközelítéstől. A **telepítés fázisának valós idejű megvalósításban nem**, hanem leginkább az előzetes felkészülési szakaszban találunk nyílt forrású információszerzési

⁵⁸ Default Passwords|CIRT.net. Elérhető: <https://cirt.net/passwords> (letöltve: 2024.01.07.).

⁵⁹ Default Passwords List. Elérhető: <https://default-password.info> (letöltve: 2024.01.07.).

⁶⁰ Default Password Lookup. Elérhető: https://www.fortypoundhead.com/tools_dpw.asp (letöltve: 2024.01.07.).

⁶¹ Yamin et al. p. 10.

⁶² Mi a vírusvédelem? – Biztonságosinternet.hu Elérhető: <https://www.biztonsagosinternet.hu/hu/tippek/60> (letöltve: 2024.01.08.).

⁶³ Virustotal. Elérhető: <https://www.virustotal.com/> (letöltve: 2024.01.08.).

⁶⁴ Hibrid Analysis. Elérhető: <https://www.hybrid-analysis.com/> (letöltve: 2024.01.08.).

⁶⁵ Ether. Elérhető: https://ether.gtisc.gatech.edu/web_unpack/ (letöltve: 2024.01.08.).

⁶⁶ Yamin et al. p. 11.

vonatkozást. Léteznek olyan tényezők (információk), amelyeket a célrendszeri környezet szolgáltató és **befolyásolhatják** a telepítés sikerességét. A támadóknak ezekre az információra szükségük van a támadás szofisztikálásához.

Ennek a legkönnyebben megvalósítható és legkevésbé kockázatosabb megoldása a **célkörnyezettel azonos virtuális környezet létrehozása (emulálás)**, amely során a **sebezhetőség szimulálható** és az **exploit tesztelhető**. Ez a lehetőség biztosítja a különböző hardveres (pl.: asztali számítógép, szerver, mobil, IoT) eszközök és szoftveres (pl.: operációs rendszer) megoldások emulálását, így a célzott rendszerrel közel azonos környezetben tesztelhető az elvárt működés. Ilyen szolgáltatások (pl.: VMware,⁶⁷ Virtual Box,⁶⁸) közül több elérhető ingyenes vagy térítés ellenében. A manapság oly népszerű okostelefonok is eredményes támadási célnak számítanak. **IOS esetében korlátozottak az emulációs lehetőségek**, de az Android alapúak közül számos harmadik féltől származó megoldás (pl.: Genymotion,⁶⁹ Bluestacks,⁷⁰ Andyroid⁷¹) áll rendelkezésre. A virtuális környezetben már rendelkezésre állnak azok a befolyásoló tényezők, amelyek biztosítják a támadások célkörnyezetre történő szofisztikálását.⁷²

2.3.6. Futtatás és vezérlés (command & control)

A sikeres település után megkezdődik a fertőzött eszköz támadó általi – kívánt mértékű – vezérlése. Ehhez leggyakrabban olyan **kommunikációs csatornákat** hoznak létre, amelyek igyekeznek jóindulatúnak tűnni és az **észlelési küszöbszint alatti mértékben folytatni a kommunikációt**. Ezzel összefüggésben is elérhetőek OSINT források, hiszen léteznek olyan **anonimitást megőrző fórumok**, amelyek tájékoztatást adnak veszélyeztetett rendszerekről. Ilyen információkat vagy értesítéseket a **kiberbiztonsági kutatók és kiberbűnözők egyaránt megosztanak**, hiszen ezzel képesek népszerűsíteni hírnevüket és szakmán belüli elismertségüket. Az ilyen információbázis jól hasznosulhat a kiberbiztonság művelőinek munkájában, hiszen az általuk kezelt rendszereket is érinthetik. De ugyanakkor rosszindulatú céllal megközelítve veszélyes eszközt biztosít a

⁶⁷ VMware Workstation 17 Player. Elérhető: <https://www.vmware.com/nordics/products/workstation-player/workstation-player-evaluation.html> (Utolsó letöltés: 2024.01.14).

⁶⁸ VirtualBox. Elérhető: <https://www.virtualbox.org/> (letöltve: 2024.01.15).

⁶⁹ Genymotion Android Emulator. Elérhető: <https://www.genymotion.com> (letöltve: 2024.01.09.).

⁷⁰ BlueStacks by now.gg. Elérhető: <https://www.bluestacks.com/> (letöltve: 2024.01.09.).

⁷¹ AndY your desktop mobile. Elérhető: <https://www.andyroid.net/> (letöltve: 2024.01.09.).

⁷² Yamin et al. p. 12.

kiberbűnözők részére is. Fontos megemlíteni, hogy egy már **kompromitálódott rendszer** további **kiberbűncselekmények elkövetésére is felhasználható.** Ilyen weboldal a Zone-H⁷³, amely a kibertámadásokat, különösen a weboldalak elleni támadásokat dokumentálja.⁷⁴ Fontos megjegyezni, hogy ezek az információk kifejezetten érzékenyek lehetnek, hiszen az oldal a hackerek „dicsőségfalaként” is szolgál, amely helyzet **etikai és jogi kérdéseket** egyaránt felvet.⁷⁵ Továbbá ide sorolható OSINT eszközök például a Blackweb,⁷⁶ Malware patrol⁷⁷, amelyek a **rosszindulatú szoftverek terjedéséről szolgáltatnak ismereteket.**⁷⁸

2.3.7. Célok elérése (actions on objectives)

Az eddig bemutatott szakaszok sikeres teljesítését követően jutunk el a befejező szakaszhoz, azonban ebben a fázisban **közvetlen és ártó szándékú OSINT tevékenység nehezen azonosítható.** Ami viszont ide köthető, az az incidensek észlelésével kapcsolatban megjelenő szervezetek közti **incidens, avagy fenyegetés hírszerzés** (threat intelligence). Egy észlelt támadás esetén az incidensre vonatkozó információkat megoszthatják a szervezetek egymással, így más szervezetek **fel tudnak készülni** az adott típusú, vagy hasonló incidensre. Az ilyen jellegű kommunikáció fontosságát igazolja, hogy az Európai Unió Hálózat- és Információbiztonsági Ügynökség (ENISA) a CSIRT (Computer Security and Incident Response Team) csoportok nem alapvető, de megelőző szolgáltatásaik közé sorolja a biztonsággal kapcsolatos információk terjesztését is.⁷⁹ Az ilyen információknak van nyílt forrású elérhetősége is, mint például a Trellix⁸⁰, vagy a valós megfigyelések feldolgozásán alapuló MITRE ATT&CK.⁸¹ A felsorolt adatbázisokban olyan információk is megjelenhetnek, amellyel szemben – bizonyos szervezetek vonatkozásában – védelmi intézkedések még nem születtek, így azok a kibertámadók számára eredményesek lehetnek.

⁷³ Zone-h unrestricted information. Elérhető: <https://www.zone-h.org/> (letöltve: 2024.01.09.).

⁷⁴ Megjegyzés: az oldal használatakor érdemes óvatosnak lenni mert az információk felhasználása bizonyos körülmények között kockázatos.

⁷⁵ Yamin et al. p. 12.

⁷⁶ GitHub — maravento/blackweb. Elérhető: <https://github.com/maravento/blackweb> (letöltve: 2024.01.09.).

⁷⁷ Malware Patrol. Elérhető: <https://www.malwarepatrol.net/> (letöltve: 2024.01.09.).

⁷⁸ Yamin et al. p. 13.

⁷⁹ Részletes leírás a CSIRT csoportok létrehozásáról. ENISA. Elérhető: <https://www.enisa.europa.eu/publications/csirt-setting-up-guide-in-hungarian/@@download/fullReport>

⁸⁰ Trellix. Elérhető: <https://www.trellix.com/about/newsroom/stories/research/> (letöltve: 2024.01.14.)

⁸¹ Mitre Att&ck. Elérhető: <https://attack.mitre.org/> (letöltve: 2024.01.14.)

Az **idő kritikus tényező a kiberincidensek kezelésében**, így akinél fennáll a sebezhetőség és huzamosabb időn belül nem intézkedik annak kezelésére, azzal szemben jó eséllyel alkalmazni fogják az ilyen adatbázisból származó ismereteket.

2.4. Védelmi megoldások OSINT-ellen

Nyilvánvaló, hogy az **online jelenlét növeli a szervezet hírnevét**, és közvetlenül befolyásolja annak sikerét. Jelentős hatással van márkaismertségre, az ügyfélszerzésre, a szervezet hírnévmenedzsmentjének, valamint az ügyfélbizalom kiépítésének egy fontos eszköze. Ugyanakkor a **nagy mértékű online jelenlét estén a biztonsági incidensek és az adatszivárgások** kapcsán **kockázatokkal** kell számolnunk. Ezek az **események negatívan befolyásolhatják a szervezet hírnevét**.

Az eddigiekben érintettük az OSINT eszközök hatékonyságát és képességeit, de léteznek olyan védelmi célú megközelítések és ajánlott intézkedések, amelyek csökkentik a digitális lábnyom méretét és hozzájárulnak az online megjelenés egy biztonság tudatosabb formájának létrehozásához. Ehhez ajánlatos egy **digitális lábnyom⁸² stratégiát** bevezetni, amellyel összefüggésben meghozhatók azok az intézkedések, amelyek hozzájárulnak a szervezeti célokhoz.

Amennyiben az OSINT tevékenységek elleni digitális lábnyom csökkentése a cél, a legfontosabb cél **az online jelenlét adatminimalizálása**. Az elv: csak a szükséges adatok közzététele, amely nélkülözi a személyes adatok megosztását. Segíthet egy olyan **adattvédelmi szabályzat bevezetése**, amely figyelemmel van az OSINT elleni védelmi szempontokra, ugyan ezen a területen járulhat hozzá az OSINT technikákra és adattvédelmi **kockázatokra vonatkozó oktatás** is.

A magánélet védelmét erősítő technológiák (**Privacy Enhancing Technologies – PET** és az **Anti-Fingerprinting Privacy Enhancing Technologies – AFPET**) szintén hozzájárulhatnak az online térben jelentkező kockázatok csökkentéséhez, így hozzájárulva a szervezetek biztonságos adatkezeléséhez.

⁸² A digitális lábnyom az egyének és szervezetek által az online világgal való interakció során hátrahagyott adatok egyedi gyűjteménye.

2.5. A kibertéri nyílt forrású információszerzés SWOT elemzése

Az OSINT tevékenység megfogalmazása pár rövid sorban megtörtént, de a gyakorlat ábrázolása a szakirodalmakban helyenként ellentmondó. Konfliktus jelenik meg az alkalmazott keretrendszer határainak kitűzése során **a nyilvánosságra, a törvényességre, a különböző jogrendszerek által eltérően védett személyiségi jogok védelmére** vonatkozólag.

A polgári ágazatban megjelenő üzleti, valamint a katonai és rendvédelmi célú OSINT tevékenységek pontos mezsgyéinek megfogalmazása **elmosódottként jellemezhető**, valamint az alkalmazott definíciók és eszközrendszerek tekintetében **sem pontosan strukturált**.⁸³ Ezeket a problémákat nemzetbiztonsági megközelítéssel, gyakorlati példákkal szemlélteti és elemzi tanulmányában Solti (2019)⁸⁴, míg Csizner (2019)⁸⁵ ennek a jogi, az időbeliségi, a biztonsági és szakértelmi hiányosságok nehézségeit szemlélteti.

Az OSINT alkalmazás során felmerülő számos nehézség és hátráltató tényező, az ismert előnyök tükrében megadja a lehetőséget arra, hogy azokat együttesen megjelenítve és elemezve egy átfogó ismertetést kapjunk.

Vadász (2018)⁸⁶, Dobák (2019)⁸⁷, valamint Qusef és Alkilani (2021)⁸⁸ által készített tanulmányok több helyen említik az OSINT előnyeit és hátrányait, így ezeket az információkat feldolgozva, továbbá a ChatGPT 4.0 által legenerált információkkal⁸⁹ kiegészítve képesek vagyunk egy SWOT analízis⁹⁰ létrehozására, amely praktikusán ábrázolja az OSINT stratégia szempontú elemzését.

⁸³ Vadász (2018). p. 66.

⁸⁴ Solti (2019). p.11.

⁸⁵ Csizner (2019).

⁸⁶ Vadász (2018). p. 67.

⁸⁷ Dobák (2019). p. 4-8.

⁸⁸ Abdallah Qusef & Hamzeh Alkilani (2022). The effect of ISO/IEC 27001 standard over open-source intelligence. PeerJ Comput Sci. <https://doi.org/10.7717/peerj-cs.810>

⁸⁹ OpenAI ChatGPT 4.0 lekérdezés. 2023.12.17. forrás: <https://chat.openai.com/share/e28593b3-4295-4d39-9181-5e05e374827f>

⁹⁰ SWOT mozaikszó: Strengths, Weaknesses, Opportunities, Threats.

3. ábra: a kibertéri nyílt forrású információszerzés SWOT analízise. Forrás: a szerző szerkesztése.

A KIBERTÉRI NYÍLT FORRÁSÚ INFORMÁCIÓSZERZÉS SWOT ANALÍZISE		
	Erősségek (Strengths)	Gyengeségek (Weaknesses)
Belső tényezők	• Hozzáférhetőség • Költséghatékonyság • Adatgazdagság • Időhatékonyság • Rejtve maradás • Bürokrácia-mentesség	• Információk megbízhatósága • Adatdömping • Elemzési képességek • Etikai és jogi kérdések • Védett adatok • Transzparencia-szükséglet
	Lehetőségek (Opportunities)	Fenyegetések (Threats)
Külső tényezők	• Technológiai fejlődés • Szélesebb körű alkalmazhatóság • Közösségi hálózatok és big data folyamatos adatszállítása • Információszivárgás értékelése	• Adatvédelmi és egyéb jogi szabályozások • Dezinformáció és manipuláció • Etikai aggályok • Illegális felhasználás

Erősségek.

Ide sorolható a nyilvánosan elérhető adatok könnyű elérése, **hozzáférhetősége**. A többi hírszerzési formával ellentétben az OSINT **költséghatékonynak** mondható, hiszen kevés költséggel jár a nyílt, ingyenes forrásokból származó adatok megszerzése. Forrását tekintve **adatgazdag**, hiszen a nyílt források széles körű információkat kínálnak különböző témákban.

A többi hírszerzési formával ellentétben az OSINT alapvetően egy **gyors információszerzést** tesz lehetővé. A kibertéri információgyűjtés esetében ez különösen jellemző, hiszen a számos specifikus OSINT eszköz növeli a hatékonyságot, így rövidebb idő alatt képes a cél elérésére. A megfelelő technikával és módszertanokkal tökéletesen álcázható, így alkalmazása **rejtve marad**. Hírszerzési formaként kezelve alkalmazása **mentesül a bürokráciától**, mert engedélyeztetése nem szükséges.

Gyengeségek.

Információk megbízhatósága ennek számít, hiszen nehéz differenciálni a valid információkat és a dezinformációs célból készített tartalmakat. A hatalmas információmennyiség (**adattömping**) – a rendelkezésre álló források tekintetében – feldolgozása kihívást jelenthet. **Elemzési képességek:** az adatok elemzéséhez humán szakértelem és eszközök szükségesen, ezek hiánya problémát jelenthet, így gyengítheti a tevékenység folytonosságát. **Etikai és jogi kérdések** merülhetnek fel az adatgyűjtés során, hiszen a vizsgált adatok magánéleti jogokat sérthetnek. Fontos megemlíteni, hogy a **biztonságtudatos személyek esetében nem hatékony**, így adatvédelmi gyakorlatok alkalmazása gyengíti az OSINT hatékonyságát. Ahhoz, hogy egy hírszerzési tevékenység OSINT-minőségét megőrizze, a teljes tevékenységet átölelő **transzparencia szükségeltetik**. Ismerni kell a teljes folyamatot és a módszertant, hiszen a források mellett az információt értelmező módszernek is „nyílt”-nak kell lennie. Például nem szerepelhet benne olyan rész-tevékenység, amely nem nyilvános forrásokat gyűjt fel.

Lehetőségek.

Az új technológiák tovább javíthatják az adatgyűjtés hatékonyságát és pontosságát, így a **technológiai fejlődés** bizonyosan a tevékenység támogatója lesz a jövőben. A digitális transzformációnak köszönhetően az OSINT egyre inkább alkalmazhatóvá válik **szélesebb körben**, hiszen kiterjedhet az üzleti intelligenciára, a biztonsági elemzésre, a hírszerzésre és egyéb területekre is. A **közösségi hálózatok által generált adatfolyam és a big data jelensége** továbbra is szállítja a nagy adatmennyiséget, amely egy állandósult, aktualitást biztosító, új információ-forrásként jelenik meg az OSINT tevékenységekhez. Az OSINT alkalmas védelmi célú intézkedések támogatásához, hiszen az **információ-szivárgás** egy olyan értékelhető jelenség, amely feltárja az informatikai infrastruktúrák hiányosságait és hibáit.

Veszélyek.

Az **adatvédelmi szabályozások** és az OSINT módszereinek gyakorlati alkalmazása okozta érintkezések továbbra is konfliktusosok. A hamis információk kezelése megnehezítheti a valós adatok azonosítását, így a **dezinformáció és manipuláció** jelentősen veszélyeztetik az OSINT adatforrásainak hitelességét. A módszerrel és eljárásokkal kapcsolatban **etikai és jogi aggályok**

(pl.: a magánélet megsértése) merülhetnek fel az OSINT folyamatokban. A technológia – nyíltsága okán – rosszindulatú szándékkal szabadon alkalmazható és a felhasználásához jelentős források biztosíthatók, így annak **illegális felhasználása** kockázatként jelenik meg.

2.6. A mesterséges intelligencia megjelenése az OSINT-ben

Azt az információt, hogy az OSINT történetiségében mikor lépett be mesterséges intelligencia, könnyen szemléltethetjük el João Rafael Gonçalves Evangelista és munkatársai 2021-ben végzett kutatásának megállapításaival, amely során az OSINT szakirodalmi bázisát **részletes elemzésnek vetették alá**.

Csak a közeli OSINT múlt témáit bemutatva, 2012 és 2014 között a **közösségi hálózatok és közösségi médiák elemzési tevékenysége** jelent meg dominánsan a **szövegbányászat és a szentimentelemzés** technikáival együttesen.⁹¹

2014 és 2016 között az OSINT fő vizsgálati fókusza már a **big data** technológia és az **információbiztonság** területére koncentrálódott, hiszen ekkoriban már **gépi tanulóssal és NLPmegoldásokkal** igyekeztek a **kiberincidensek és social engineering elleni védelmet** kisegíteni.⁹²

2016-tól 2019-ig az OSINT a kiberbiztonság területén jeleskedett, és leginkább a **darkwebbel** kapcsolatos keresési információkra, a **földrajzi helymeghatározásra, a kiberfenyegetések felderítésére** helyezték a **gépi tanulási erőforrásokat**.⁹³

A kutatás megállapításai között volt az is, hogy a vett minta alapján **az OSINT tanulmányok 24% mesterséges intelligenciával is foglalkozott**, valamint azt, hogy ezek közül **a tanulmányok 41%-a kiberbiztonság területét dolgozta fel**.⁹⁴ Ezek alapján láthatjuk, hogy a **dolgozat által vizsgált OSINT és AI témakörök a kiberbiztonságban a legnépszerűbben kutatott területnek számítanak**.

⁹¹ João Rafael Gonçalves Evangelista, Renato José Sassi, Márcio Romero & Domingos Napolitano (2021). Systematic Literature Review to Investigate the Application of Open Source Intelligence (OSINT) with Artificial Intelligence. Journal of Applied Security Research, 16. <https://doi.org/10.1080/19361610.2020.1761737> p. 358.

⁹² João Rafael et al. (2021). p. 358.

⁹³ João Rafael et al. (2021). p. 358.

⁹⁴ João Rafael et al. (2021). p. 362.

3. A NAGY NYELVI MODELL (LLM)

3.1. Mesterséges intelligencia (MI)

A **mesterséges intelligencia (MI)**, vagy **angolul artificial intelligence (AI)** manapság egy kifejezetten népszerű terület, de a fogalom nem újkeletű, hiszen jelentős történetisége van. Talán hihetetlennek tűnik, de az **első MI eredményt 1943-ra** jegyzi a tudomány. Ekkor történt, hogy Warren McCulloch és Walter Pitts az agyi neuronokról szóló alapszintű fiziológiai ismereteikhez társították a matematikai elemzés módszertanát, így megalkotva **a legelső neuronmodellt**, amely nélkül ma nem beszélhetnénk korszerű MI megoldásokról. A következő években sorra jelentek meg olyan kutatások és szabályok (pl.: Hebb-tanulás 1949-ben), vagy műszaki megoldások (pl.: Marvin Minsky és Dean Edmonds Snarc elnevezésű neurális számítógépe 1951-ben) amelyek ezekre az ismeretekre alapoztak. Azonban talán a **legteljesebb elképzelés Alan Turing-hoz köthető**, akinek 1950-ben megjelent megfogalmazásai (pl.: **Turing-teszt, a gépi tanulás, megerősített tanulás**) manapság is gyakran említett kifejezésnek számítanak a MI terminológiában.⁹⁵

A mesterséges intelligencia kialakulásában tett történelmi utazást bizonyára számos érdekes esemény tenné színessé, de terjedelmi korlátok okán ennek bemutatásától ez a dolgozat eltekint. Azonban mivel a nagy nyelvi modellek MI megoldásoknak minősülnek, így a bemutatást nem nélkülözhetjük. Az említett színes történetiségnek köszönhetően **több MI definíciót alkottak** az idők során, de a technológia rohamos fejlődése okán előbb-utóbb elavulttá váltak, így **nincs egy egységesen elfogadott közülük**.

Az MI definíciós kísérletei **három perspektívából alkothatnak maradandó gondolatot**, ezek a **filozófiai, a technológiai és a jogi**, melyek közül a technológiai megközelítések **nem tekinthetők konzisztensnek**, hiszen a gyorsvonatként száguldó innovatív technológiát meghatározni műszaki paraméterekkel **maradéktalanul és időtállóan megfogalmazni lehetetlen küldetés**. Azonban a **filozófiai megközelítések permanensnek bizonyultak**, és manapság népszerűnek is mondhatók. Erre példa **John. R. Searle 30 éves definíciója**, aki az MI magyarázatát a kognitív állapot elérése

⁹⁵

Mesterséges Intelligencia Elektronikus Almanach. Elérhető:
http://project.mit.bme.hu/mi_almanach/books/aima/ch01s03 (letöltve: 2024.01.21.)

mentén kétfelé, **erős és gyenge MI-re osztotta.**⁹⁶ Ez a megközelítés azonban **számomra nem kellően részletes**, hiszen a jelenleg gyakorlatban megjelenő megoldások differenciálását véleményem szerint magától értetődőbben kellene végezni. Meggyőződésem, hogy a jogi deklaráció adhatja meg azt a **pontosságot, egyértelműséget, interdiszciplinaritást és időtállóságot** amelyre jelenleg a szükség kívánczik. A következőkben ezen kísérletekből példálózunk.

Az amerikai jog a következő kísérletet teszi (magyar nyelvre fordítva):

*„A **"mesterséges intelligencia"** kifejezés olyan gépi alapú rendszert jelent, amely egy adott, ember által meghatározott célrendszerre vonatkozóan képes valós vagy virtuális környezetet befolyásoló előrejelzéseket, ajánlásokat vagy döntéseket hozni. A mesterséges intelligencia rendszerek gépi és emberi alapú bemeneteket használnak a következőkhöz:*

- 1. érzékelik a valós és virtuális környezeteket;*
- 2. az ilyen észleléseket automatizált módon történő elemzéssel modellekké absztrahálják; és*
- 3. a modell következtetések felhasználásával információs vagy cselekvési lehetőségeket fogalmazznak meg.”*⁹⁷

A hazánkban érvényes európai jogalkotás akarata napjainkban ölt alakot a témában. Talán az eddig legteljesebb kezdeményezés az **Európai Parlament és a Tanács Rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok** (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról szóló rendelete. A Rendelet a következőképpen határozza meg az MI-t:

*„A **mesterséges intelligencia rendszer (MI-rendszer)** olyan szoftver, amelyet a felsorolt technikák és megközelítések közül egy vagy több alkalmazásával fejlesztettek, és amely az ember által meghatározott célkitűzések adott csoportja tekintetében olyan kimeneteket (például tartalmat,*

⁹⁶ John R. Searle (1980). Minds, brains, and programs. Behavioral and Brain Sciences. 1980;3(3):417-424. <https://doi.org/10.1017/S0140525X00005756> 417.p.

⁹⁷ 15 U.S. Code § 9401 – Definitions (3)

előrejelzéseket, ajánlásokat vagy döntéseket) képes generálni, amelyek befolyásolják azt a környezetet, amellyel kölcsönhatásba lépnek.⁹⁸

A technikák:

- a. *Gépi tanulási megközelítések, ideértve a felügyelt, a felügyelet nélküli és a megerősítő tanulást, a módszerek széles skálájának, többek között a mélytanulásnak az alkalmazásával;*
- b. *Logikai és tudásalapú megközelítések, beleértve a tudás megjelenítését, az induktív (logikai) programozást, a tudásbázisokat, a következtetőmotorokat, a(z) (szimbolikus) érvelést és a szakértői rendszereket;*
- c. *Statisztikai megközelítések, Bayes-féle becslés, keresési és optimalizálási módszerek.*⁹⁹

Anélkül, hogy a két definíció között sarkalatos elemzői párhuzamot vonnánk, észlelhetjük, hogy számos jellemzőben **azonos a két jogalkotói akarat**, hiszen az **emberi cél, a környezettel való kölcsönhatás és a gépiesség** egyaránt alkalmazott megfogalmazások mindkét esetben. A továbbiakban ebben a jelenségben vezetjük át az olvasót a vizsgált téma második eleméig, vagyis a nagy nyelvi modellekig.

3.2. A gépi tanulás módszerei (ML)

Az előző fejezetben láthattuk, hogy a bemutatott MI definíció a **gépi tanulást (machine learning – ML)** további alkategóriákra bontva említi, amelyek a **felügyelt (supervised learning – SL)**, a **felügyelet nélküli (unsupervised learning – UL)**, valamint a **mélytanulás (deep learning – DL)**. Arthur Samuel szerint gépi tanulás **a számítástechnika azon részterülete, amely anélkül képes tanulni, hogy programoznánk azt.**¹⁰⁰ Ezek az algoritmusok statikus programutasításokat követve a jelentős adattömegekből (bemeneti minta) képesek modellt építeni (tanulni) és előrejelzéseket készíteni. A gépi tanulás szorosan kapcsolódik a **számítógépes statisztikákhoz**, valamint

⁹⁸ Az európai parlament és a tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról. 3. cikk. 1. Elérhető: <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX:52021PC0206>

⁹⁹ Az európai parlament és a tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról.

¹⁰⁰ Andres Munoz (2012). Machine Learning and Optimization. Elérhető: https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf

értelmezésében néha **összemosódik az adatbányászattal**. Az **adatelemzés területén** alkalmazott gépi tanulási módszerek pedig olyan modellek és algoritmusok, amelyek képesek előrejelzéseket produkálni, ezt nevezik **prediktív analitikának**.¹⁰¹

A gépi tanulás esetében két elterjedt alkalmazási módszerről beszélhetünk, amelyek a már említett **felügyelt és a felügyelet nélküli tanulás**. A megjelenés tekintetében ezek közül előbbi a jelentősebb (körülbelül 70%), míg utóbbi 10-20% arányban jelenik meg, ugyanakkor meg kell említeni a **félíg felügyelt és a megerősített tanulási módozatokat**, amelyek ritkábban alkalmazott módszerek.¹⁰²

A **felügyelt tanulás** során **címkezett példák segítségével** képzik ki a modellt úgy, hogy a **kimenet előre definiált**. Ebben az esetben az algoritmus úgy tanul, hogy két készletet (bemenet és kimenet) összeveti és megalkotja a modellt annak érdekében, hogy a címkezetlen adatok tekintetében is hatékony legyen. Ehhez leginkább olyan módszereket alkalmaz, mint például a **regresszió**,¹⁰³ a **predikció**,¹⁰⁴ a **gradient boosting**.¹⁰⁵ A felügyelt tanulást leginkább olyan célokra használják, ahol a múltbéli adatok feldolgozásával valószínűsíthetőek bizonyos jövőbeli események.¹⁰⁶

A **felügyelet nélküli tanulást** azonban leginkább akkor alkalmazzák, amikor a **struktúrálatlannak tűnő adattömegből** kell kitalálnia az algoritmusnak, hogy milyen minták találhatók benne. Címkezett adatok nélkül a **cél az adatokban található egyfajta szabályosság vagy struktúra megállapítása**. Példaként említve ez a módszer kiválóan működik pénzügyi tranzakciós adatokon, hiszen hasonló jellemzőkkel rendelkező ügyfelek csoportjai azonosíthatók, amelyek így felhasználhatók további szervezeti célok (pl.: marketing tevékenység) érdekében. Az

¹⁰¹ Pariwat Ongsulee (2017). Artificial intelligence, Machine learning and Deep learning, 15th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, Thailand. <https://doi.org/10.1109/ICTKE.2017.8259629>

¹⁰² Pariwat (2017).

¹⁰³ A regresszió egy statisztikai módszer, amelyet arra használnak, hogy megvizsgálják a változók közötti kapcsolatokat.

¹⁰⁴ A predikció, vagy előrejelzés egy olyan folyamat vagy technika, amelynek célja kimenetek, eredmények vagy adatok jövőbeli állapotának meghatározása korábbi adatok vagy ismert információk alapján.

¹⁰⁵ A gradiens erősítés egy gépi tanulási technika, amely a döntési fák sorozatát használja előrejelzések készítésére és a prediktív teljesítmény javítására.

¹⁰⁶ Pariwat (2017).

alkalmazott módszerek az **önszerveződő térképek**¹⁰⁷, a **legközelebbi szomszéd analízis**,¹⁰⁸ **K-közép klaszterezés**¹⁰⁹ és a **szinguláris felbontás**.^{110 111}

A **félíg felügyelt tanulást (semi-supervised learning – SSL)** többnyire ugyanazon tevékenységekre alkalmazzák, mint a felügyelt tanulást, azonban a modellképzéshez **címkezett adatokat csak kis mértékben**, míg a **címkezetlen adatokból nagy mennyiségben** használnak fel. Ez a módszer annyiban tér el a felügyelt tanulástól, **hogy gazdaságosabbá válhat a kifejlesztése abban az esetben**, ha az adatok címkézése jelentős költségekkel jár.¹¹²

A **megerősített tanulás (reinforcement learning – RL)** egy szintén önálló kategória, amely megoldás **leginkább a robotikában és a játékokban** (pl.: sakk) hasznosul. Ebben az esetben az algoritmus próbálgatással fedezi fel azt, hogy melyik művelet eredményei adják a legnagyobb jutalmat, vagyis a megerősítést. A módszernek három összetevője van: az **ágens** (a döntéshozó), a **környezet** (amivel az ágens kapcsolatba léphet), és az **akciók** (amit az ágens megtehet). A cél ebben a tekintetben az, hogy az ágens olyan döntéseket hozzon, amelyek maximalizálják a várható jutalmat.¹¹³

A **mélytanulás (deep learning)** vagy más néven **mély gépi tanulás**, egy olyan speciális gépi tanulás, amely **egynél több mesterséges neurális hálózatot** és hozzá kapcsolódóan gépi tanulási algoritmusokat tartalmaz. A **mesterséges neuronokról** korábban (4.1. fejezet) már esett szó, alapjaiban nem új technológia, ám a jelenkor számítási kapacitás-lehetőségeivel és egyéb algoritmikákkal társítva az **egyik legdinamikusabban fejlődő terület**. A mélytanulás során a legegyszerűbb esetben két neuroncsoport lehetséges, melyből az egyik a bemeneti jelet kapja, a másik pedig kimeneti jelet küld, ám egy mély hálóban a bemenet és kimenet között számos réteg

¹⁰⁷ Az önszerveződő térkép (SOM), más néven Kohonen-térkép, egyfajta mesterséges neurális hálózati technika, amelyet nagydimenziós adatok alacsonyabb dimenziókban történő feltérképezésére és megjelenítésére használnak. Forrás: <https://houses-apartments.com/hu/design/parametric-and-computational-design/what-is-a-self-organizing-map/>.

¹⁰⁸ A legközelebbi szomszéd analízis a térbeli rendezettség vizsgálatának gyakorta alkalmazott matematikai statisztikai eszköze. Forrás: https://hu.wikipedia.org/wiki/Legk%C3%B6zelebbi_szomsz%C3%A9d_anal%C3%ADzis

¹⁰⁹ A K-közép klaszterezés egy felügyelet nélküli tanulás algoritmus, amely erejének és egyszerűségének köszönhetően a legszélesebb körben használt megoldás.

¹¹⁰ A szinguláris érték felbontás (SVD, Singular Value Decomposition) egy matematikai módszer, amelyet elsősorban lineáris algebra területén használnak.

¹¹¹ Pariwat (2017).

¹¹² Pariwat (2017).

¹¹³ Pariwat (2017).

neuron található, amelyek **valójában nem neuronokból állnak, de a felépítést könnyebb megértenünk, ha így gondolunk rá.** A mélytanulás tulajdonképpen az adatok **reprezentációjának**¹¹⁴ megtanulásán alapul. Ezt a folyamatot úgy képzelhetünk el, hogy ha a bemenet egy kép, annak számos jellemzője reprezentálható, akár pixelenkénti intenzitás vektorai alapján, vagy absztraktabb módon alakzatok vagy régiók halmazaként. Az említett algoritmusok, amelyek lehetnek felügyelték és nem felügyelték ezeket a reprezentációkat igyekeznek megtanulni. Az ilyen architektúrákat előszeretettel alkalmazzák olyan területeken, mint a számítógépes látás, beszédfelismerés, hangfelismerés, biometrikus azonosítás és **a természetes nyelvfeldolgozás.**¹¹⁵

3.3. Természetes nyelvfeldolgozás (NLP)

A soron következő tématerületek bemutatása előtt illedelmes érinteni a mesterséges intelligencia egyik azon ágát, amely nélkül manapság nem lennének nagy nyelvi modellek. A nyelvészetnek **két ága van, a számítógépes nyelvészet és az elméleti nyelvészet.** A számítógépes nyelvészet olyan algoritmusok kifejlesztésével foglalkozik, amelyek a természetes nyelvet alkalmazzák bemenetként.¹¹⁶

A természetes nyelvfeldolgozás (NLP) egy tudományos és technológiai alapú kutatási terület, amely **az emberi nyelvek reprezentációjának** és automatikus elemzésének számítási technikáit foglalja magában. Az emberi nyelvek szövegszerű elemzése, annak mély megértése még a gépek számára sem egyszerű, hiszen abban számos olyan nehezítő tényező merül fel, mint például a nagy mennyiségben megjelenés, az összetettség, a többértelműség. Az automatizált NLP célja, hogy egy adott időkorláton belül a **nyelvi szövegek bőséges mennyiségében rejlő nagy tudástartalmat, bölcsességet hatékonyan és nagy pontossággal felfedezze.** Egy erre alkalmas program kifejlesztése eléggé problematikus, de az internet hatalmas mennyiségű erőforrásként használható az NLP rendszerek alkotására, alkalmazására.¹¹⁷

Az NLP-ek alkalmazási területei többnyire az alábbiak:

¹¹⁴ Az az eljárás, amely során a nyers adatokat olyan formátumba alakítjuk át, amelyet a mély neurális hálózatok hatékonyan tudnak feldolgozni.

¹¹⁵ Pariwat (2017).

¹¹⁶ K.R. Chowdhary (2020). Fundamentals of Artificial Intelligence. Springer India 1st ed. ISBN: 81-322-3972-5. <https://doi.org/10.1007/978-81-322-3972-7>. p. 603-640.

¹¹⁷ K.R. Chowdhary (2020).

- automatikus nyelvi fordítás,
- információ keresés (IR, vagyis information retrieval),
- információ kinyerése (IE, vagyis information extraction),
- kérdések és válaszolás (QA, vagyis question and anwering),
- nagy méretű szövegek indexelése, keresése,
- szövegek automatikus összefoglalása,
- szövegek klasszifikációja,
- szöveggenerálás/dialóguselemzés.¹¹⁸

Az NLP legegyszerűbb megközelítései a **szintaktikai és a szemantikai elemzés** módszertana. A **szintaktikai elemzést** a nyelvtan segítségével végzik el, meghatározzák a vizsgált mondat szerkezetet, valamint az igék alanyát és tárgyát. Ha a szerkezet meghatározása megfelelő, akkor további operátor-operandus kapcsolatok meghatározásával érkezőnk el a szintaktikai elemzés kimeneteléhez.¹¹⁹

A **szemantikai elemzés** célja azonban annak a meghatározása, **mit jelent és mit tartalmaz** az adott mondat, túllépve a szavak önálló értelmezésének szintjén. Leginkább arra törekszik, hogy megértse a kifejezések és a mondatok mögötti **kontextusbeli kapcsolatokat és a rejtett információkat**. A szemantikai elemzés a kérdő és felszólító mondatok esetében körülményesebb, ugyanakkor az olyan feladatokban, mint a gépi fordítás, az összefoglaló készítés, a kérdésmegválaszoló rendszerek és az érzelmek elemzése **kulcsfontosságú megközelítés**.¹²⁰

A felsoroltakon kívül még említhetnénk számos elemzési megközelítést (pl.: pragmatikaikai, diskurzus, szövegkohéziós stb.), de a terjedelmi korlátok okán ezektől eltekintünk.

3.4. A nagy nyelvi modellek felépítése, működése és képességei

Az NLP történeti fejlődése színes, hiszen átível a statisztikai nyelvi modellezésen, a neurális nyelvi modellezésen, majd az előre tanított nyelvi modelleken (Pretrained Language Model, a továbbiakban: PLM), és elér a manapság nagy népszerűségnek örvendő **nagy nyelvi modellekig**

¹¹⁸ K.R. Chowdhary (2020).

¹¹⁹ K.R. Chowdhary (2020).

¹²⁰ K.R. Chowdhary (2020).

(Large Language Model, a továbbiakban: LLM). A **hagyományos nyelvi modellek (LM)** eleinte csak feladatspecifikációra fejlesztett felügyelt rendszerek voltak, addig az őket követő PLM modelleket már önfelügyelt¹²¹ környezetben, nagy szövegtörzseken¹²² tanítják.¹²³ A nagyobb PLM-ek már több paramétert¹²⁴ tartalmaztak, mint elődjeik, és természetesen a **képzésükhöz szükséges adatállomány** is jelentősen megnőtt. Ennek köszönhetően a modellek teljesítménye tovább növekedett és egy **új elnevezéssel már nagy nyelvi modellekként hivatkozunk rájuk**,¹²⁵ hiszen ezek már valóban képesek a koherens kommunikációra és e mellett többféle feladatra alkalmazhatók.¹²⁶

*A nagy nyelvi modellek olyan mélytanulási algoritmusok, amelyek nagy mennyiségű paramétert és képzési adatot használnak a szövegek megértéséhez és előrejelzéséhez. A generatív mesterséges intelligencia-alapú modell az egyszerű szöveggeneráláson kívül számos természetes nyelvi feldolgozási feladatot képes elvégezni, beleértve a tartalom átdolgozását és fordítását is.*¹²⁷

Típusait osztályok szerint csoportosíthatjuk:

- **Encoder (kódoló):** többnyire csak nyelv megértésére alkalmasak, és olyan funkciókat látnak el, mint az osztályozás vagy a szentimentelemzés. Ilyen architektúra például a BERT (*Bidirectional Encoder Representations from Transformers*).
- **Decoder (dekódoló):** rendkívül eredményesek tartalom generálásában (pl.: történetírás). Ilyen architektúra a GPT-3 (*Generative Pre-trained Transformer*).

¹²¹ Az önfelügyelt tanulás során a modell olyan feladatokat kap, amelyek a meglévő adatokból generálhatók.

¹²² A "korpusz" a nyelvészetben és a szöveges adatokkal dolgozó területeken egy nagy, strukturált szöveggyűjteményt jelent, amelyet tudományos kutatásokhoz, nyelvi modellek betanításához vagy egyéb nyelvvel kapcsolatos feladatokhoz használnak.

¹²³ Jacob Devlin, Ming-Wei Chang, Kenton Lee & Kristina Toutanova (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. DOI: <https://doi.org/10.48550/arXiv.1810.04805>

¹²⁴ A paraméterek a modellben lévő súlyokat jelentik, amelyek a tanulási folyamat során állítódnak be annak érdekében, hogy a bemeneti szöveg alapján pontos kimenet generálódjon.

¹²⁵ C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, vol. 21, no. 1, p. 5485–5551.

¹²⁶ Blaise Agüera y Arcas (2022). Do large language models understand us. Daedalus. vol. 151, no. 2, p. 183–197. https://doi.org/10.1162/daed_a_01909.

¹²⁷ Michael McDonough (2024.01.11.). Large Language Model. Elérhető: <https://www.britannica.com/topic/large-language-model> (letöltve: 2024.01.31.)

- **Encoder-decoder (kódoló-dekódoló):** az ilyen architektúra típusok egyaránt egyesítik mindkét típus összetevőit és kiválóak olyan feladatokban, mint a fordítás és az összefoglalás. Ilyen architektúra például a T5 (*Text-to-Text Transformer*).¹²⁸

Az utóbbi években LLM megjelenések jelentősen emelkedő tendenciát mutatnak, amelyeknek oka a **rendkívüli számítási kapacitások rendelkezésre állása**, a **nagyméretű képzési adatok elérhetősége**, de vitathatatlan, hogy ebben a technológiában az **igazi áttörést a transzformátorok alkalmazása hozta el**.¹²⁹

3.4.1. A transzformátor architektúra

A transzformátorok a jelenkori LLM-ek a mély tanulási architektúrák **egyik önálló osztályát képviselik**, amelyeket transzformátorhálózatoknak is neveznek. A transzformátormodell **egy neurális háló**, amely a **kontextust és jelentést tanul meg azáltal, hogy követi a kapcsolatokat a szekvenciális adatokban**. A transzformátorok több rétegből (transzformátorblokkból) állnak, amelyeknek vannak **önfigyelő (self-attention)**, **előrecsatoló (feed-forward)** és **normalizáló(normalizaton)** rétegei. Ezek a rétegek együtt dolgoznak annak érdekében, **hogy megértsék a bemenetet és előre megjósolják a várható kimenetet**.¹³⁰ A architektúra működését először a Google szakemberei mutatták be az „*Attention Is All you Need*” című tanulmányukban 2017-ben. Az ekkor publikált új megközelítés azonnal felhívta magára a figyelmet, hiszen a különböző fordítási teszteken csúcspontszámokat ért el annak ellenére, hogy a képzési időigénye és a költsége töredéke volt a korábbi típusoknak.¹³¹ A stanfordi egyetem kutatócsoportja 2021-es publikációban már **alapmodellnek nevezte a transzformátorokat**, mert véleményük szerint **paradigmaváltást** hajtanak végre az AI közegben.¹³²

¹²⁸ Large Language Models. Elérhető: <https://www.nvidia.com/en-us/glossary/large-language-models/> (letöltve: 2024.02.07.).

¹²⁹ A. Chernyavskiy, D. Ilvovsky, and P. Nakov (2021): Transformers:“the end of history” for natural language processing?” in Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part III 21. Springer, p. 677–693.

¹³⁰ Large Language Models. Elérhető: <https://www.nvidia.com/en-us/glossary/large-language-models/> (letöltve: 2024.02.07.).

¹³¹ Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin (2017). Attention Is All You Need. <https://doi.org/10.48550/arXiv.1706.03762>.

¹³² Rishi Bommasani és mtsai (2021): On the Opportunities and Risks of Foundation Models. Center for Research on Foundation Models, Stanford Institute for Human-Centered Artificial Intelligence, Stanford University. Elérhető: <https://arxiv.org/pdf/2108.07258.pdf>.

A transzformátor technológia hatékonyságát **két innovatív megoldásnak** köszönheti, amelyek a **pozícionális kódolás és az önfigyelés**. Előbbi egy korszerű megoldás a **szekvenciális** (szavak sorrendjében történő) **betáplálás** helyett, amely során a **szekvencián belüli előfordulást használja** bemenetként. Az **önfigyelés** a bemeneti adatok feldolgozásakor azokat a bemenet többi részével összefüggésben **fontosságuk alapján súlyozza**, amely így megadja a lehetőséget arra, hogy **ezekre összpontosítson a modell** és nem kell a teljes bemenetre maximális figyelmet fordítania. A két módszer együttes használata lehetővé teszi az **összefüggések hatékony analizálását**.¹³³

3.4.2. Fejlesztési és üzemeltetési szükségletek

A nagy nyelvi modelleket legtöbbször **nagy méretű korpuszokon, felügyelet nélküli tanulással** képzik ki. A modell képzési elv teljesen azonos a mélytanulással kapcsolatban leírtakkal, a címkézetlen adattömegben igyekszik mintákat felfedezni. A széleskörű képzési folyamatnak köszönhető, hogy az LLM modellt **nem szükséges konkrét feladatra képezni**, hanem az alapjaiban **képes több, eltérő típusú felhasználást kiszolgálni**. Annak ellenére, hogy ezek képességek elismerésre méltóak, esetenként azért hibáznak is a modellek. Azonban a fejlesztők és vállalatok eltökéltek az iránt, hogy a modelleket „megszelídítsék” és teljes mértékben a kívántak szerint működjenek. Ennek érdekében további olyan eszközöket és módszereket fejlesztenek, mint a **prompt-tuning**¹³⁴, a **fine-tuning**¹³⁵ és az **adapterek**^{136 137}.

Az LLM fejlesztés és üzemeltetés jelentős erőforrásokat igényel, amelyek közül a legszükségesebbek:

- jelentős tőkebevonás;

¹³³ Large Language Models. Elérhető: <https://www.nvidia.com/en-us/glossary/large-language-models/> (letöltve: 2024.02.07.).

¹³⁴ Virginia Adams, Zenodia Charpy and Yi Dong (2022.07.12.). Adapting P-Tuning to Solve Non-English Downstream Tasks. Elérhető: [www.developer.nvidia.com/blog/adapting-p-tuning-to-solve-non-english-downstream-tasks/](https://developer.nvidia.com/blog/adapting-p-tuning-to-solve-non-english-downstream-tasks/) (letöltve: 2024.02.07.).

¹³⁵ A fine-tuning (finomhangolás) során a modell egy adott feladatspecifikus adatkészleten történő további tréningjét végzik el, hogy a modell jobban teljesítsen bizonyos konkrét feladatokban.

¹³⁶ Az adapterek alkalmazása a modell eredeti architektúrájába kis, tréningezhető rétegek beépítését jelenti. Ezek az adaptorrétegek lehetővé teszik a modell finomhangolását specifikus feladatokra anélkül, hogy a modell teljes architektúráját újra kellene tréningezni.

¹³⁷ Large Language Models. Elérhető: <https://www.nvidia.com/en-us/glossary/large-language-models/> (letöltve: 2024.02.07.).

- nagy adatkészletek;
- műszaki szakértelem;
- nagyszabású számítási képességre képes informatikai infrastruktúra.

A tőkebevonás mértékének durva becsléséhez segítségül jöhet néhány szakirodalmi információ. 2020-ban az Open AI kutatótudósai közzétettek egy tanulmányt a GPT-3 nevű modell részletezéséről.¹³⁸ A publikált adatok alapján 12 millió dollárra becsülték a képzési költségeket.¹³⁹ A szükséges informatikai infrastruktúra becsléséhez költségpéldázat lehet az MT-NLG-ről (Megatron-Turing Natural Language Generation) szóló információk elemzése. Ezek szerint a modell képzése egy Selene szuperszámítógépen¹⁴⁰ történt, és minden modellreplika betanítása 560 db DGX A100 GPU-t ölelt fel, amelyeknek darabontként a piaci ára több, mint 200 ezer dollár. Fontos tudni, hogy amennyiben a jelentős számítási feltételek adottak, egy LLM betanítása akkor is heteket vagy **hónapokat vehet igénybe**. A Yalm 100B betanítási időigénye 65 napba telt egy 800 db A100 GPU felhasználásával és DeepSpeed (a tanítást segítő keretrendszer) segítségével. A nagy mennyiségű számítási műveletek irreálisan hosszú betanítási időt eredményezhetnek, így a tanítás során külön figyelmet kell fordítani az algoritmusok, a szoftverek és a hardverek **együttes optimalizálására is**.¹⁴¹

3.4.3. Képzési adatkészletek általános típusai

Az előző fejezetben említést tettünk az LLM létrehozásának szükségleteiről. A modellek képzéséhez további fundamentum a hatalmas adatkészlet. Ebben a fejezetben röviden kerül bemutatásra, hogy milyen típusú és forrású adatokból tanítható hatékonyan a nagy nyelvi modell.

¹³⁸ Tom B. Brown és munkatársai. (2020). Language Models are Few-Shot Learners. DOI: <https://doi.org/10.48550/arXiv.2005.14165>.

¹³⁹ Kyle Wiggers (2020.07.01.). OpenAI's massive GPT-3 model is impressive, but size isn't everything. Elérhető: <https://venturebeat.com/ai/ai-machine-learning-openai-gpt-3-size-isnt-everything> (letöltve: 2024.02.08).

¹⁴⁰ Harangi László (2020.08.17.): Egy hónap alatt összeszerelte az NVIDIA a világ hetedik leggyorsabb szuperszámítógépét. Elérhető: <https://www.pcwplus.hu/pcwpro/egy-honap-alatt-osszeszerelte-az-nvidia-a-vilag-hetedik-leggyorsabb-szuperszamitogepet-283131.html> (letöltve: 2024.02.12.).

¹⁴¹ Paresh Kharya and Ali Alvi (2021.10.11). Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model. Elérhető: <https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/> (letöltve: 2024.02.12.).

Az alapvető források:

1. **Nyilvános adatkészletek.** Különbféle forrásból származó, szabadon hozzáférhető adathalmaz, amely számos **szöveges információt tartalmaz**. Ezeket legtöbbször tudományos intézmények vagy kormányzati szervek osztják meg nyilvánosan (pl.: weboldalak, könyvek, cikkek). Előnyük a sokféleségükben és méretükben rejlik.
2. **Domain-specifikus adatkészletek.** Szintén nyilvános adatkészletek, amelyeket egy adott területre vagy iparágra szabtak. Fontos szerepet játszanak az LLM-ek adott témakörben elérhető teljesítményének javításában, hiszen úgy lettek összeállítva, hogy tartalmazzák az **egyedi nyelvi mintákat és a szakmai terminológiát**. Ilyen adatkészletekkel **javítható az LLM pontossága** olyan egyedi témákban, mint például a jog, az orvostudomány, vagy a pénzügy.
3. **Felhasználók által generált tartalmak.** Ez a tartalomkör a közösségi médiákban megjelent **bejegyzésekből, fórumbeszélgetésekből, termékismertetőkből, blogbejegyzésekből** származó, változatos nyelvi stílust és kifejezést tartalmazó szöveget jelent. Hatalmas mennyiségével biztosítja az LLM-ek számára a folyamatosan frissülő adatkészletet, biztosítva a naprakész információkat és nyelvi formátumokat.
4. **Licenszelt adatkorpuszok.** **Megbízható forrásból származó, strukturált** adatgyűjtemény, amelynek használatát jogi megállapodás szabályozza. Leginkább kiadóktól és adatszolgáltatóktól szerezhetőek be, szövegeket, képeket, hangfelvételeket tartalmaznak. LLM képzés szempontjából **értékesek és minőségiek**, és a szerzői jogok tekintetében **garantálják a jogbiztonságot**. Ugyanakkor felhasználásuk költséges és korlátozott. Ilyenek angol nyelven a Corpus of Contemporary American English (COCA), vagy a British National Corpus (BNC), magyarul a Magyar Nemzeti Szövegtár (MNSZ), vagy digitális könyvtárak és egyéb könyvtári adatbázisok.
5. **Kódtárak.** Programozási **forráskódok**, az ehhez fűződő dokumentumok és egyéb megjegyzések széles köre, amelyek leginkább olyan felületen elérhetők el, mint például a GitHub¹⁴², GitLab¹⁴³ és Bitbucket¹⁴⁴. A kódtárak elemzésével az LLM-ek megtanulják a

¹⁴² GitHub. Elérhető: <https://github.com/> (letöltve: 2024.03.02.).

¹⁴³ GitLab. Elérhető: <https://gitlab.com/> (letöltve: 2024.03.02.).

¹⁴⁴ Bitbucket. Elérhető: <https://bitbucket.org/> (letöltve: 2024.03.02.).

programozással kapcsolatos szövegezés megértését és generálását, így jártasságot szerezve a szoftverfejlesztés terén.

- 6. Szintetikus adatkészlet.** Egy technikai megoldással létrehozott adathalmaz. **Mesterséges adatokat hoznak létre** olyan esetekben, amikor a természetes adatok szűkösek, érzékenyek, vagy elérésük költséges. Létrehozásukhoz olyan algoritmusokat alkalmaznak, amelyek a valós adatok jellemzőivel rendelkező adatot állítanak elő. Előnyük van **adatvédelem, méretezhetőség és sokszínűség tekintetében**, azonban nem mindig valóságshűek, a **generáló algoritmusok fejlesztése és finomhangolása időigényes**.¹⁴⁵

A manapság elérhető LLM-ek képzési adatkészletei általában minden fentebb felsorolt csoportból tartalmaznak **legalább egy típusú forrást**. Azonban az adatkészlettel további teendők vannak, hogy az tanításra alkalmassá váljon. Aszerint, hogy milyen célú felhasználásra tervezték a modellt, annak alapján tervezik és szervezik össze a képzési adatkészletet is. Általánosságban elmondható, de a zavarosabb adatkészleteket esetében szükségszerű, hogy a képzési adatokat felkészítésnek vetik alá: **átalakítják, osztályozzák, tisztítják, deduplikálják, pontozzák és priorizálják, elvetik az alacsony minőségű adatkészletet**. A folyamat előkészítésének részeként a modell működési célját tekintve kiválasztják a legcélszerűbb és/vagy a legminőségibb adatkészletet és ennek megfelelően súlyozzák azt a képzési folyamatban úgy, hogy az adatkészlet változatossága azért megmaradjon.

3.4.4. Nyílt adatkészletek gyűjtési eszközei

Az előző fejezetben bemutatott képzési adatkészletek közül a nyíltan elérhető információs adatkészletek előnyösek abból a szempontból, hogy ezek az információk (legfőképpen szövegek, képek és strukturált adatok) helyezik időben adekváttá az LLM interakciókat. Azonban ezeket a hatalmas méretű adathalmazokat – a modellképzés előtt – **feltérképezni és megszerezni szükséges**. Manapság ehhez több automatizált eszköz elérhető, a webes szolgáltatások információinak összegyűjtésére kiváló eszköz például a Scrapy¹⁴⁶. Fontos továbbá megemlíteni a

¹⁴⁵Prominent Data Collection Methods and Tools for LLMs. Elérhető: <https://www.turing.com/resources/data-collection-methods-and-tools-for-llms> (letöltve: 2024.02.14.).

¹⁴⁶ Scrapy. An open source and collaborative framework for extracting the data you need from websites. Elérhető: <https://scrapy.org/> (letöltve: 2024.02.15.).

Common Crawl¹⁴⁷ nyílt forrású archívumát, ami egy nonprofit szervezet által fenntartott ingyenes, hatalmas mennyiségű adattár. A Common Crawl hónapról hónapra végig pásztázza az internetet, és létrehoz nagy strukturált adathalmazt, amely weboldalak milliárdjait tartalmazza. A kutatás során megállapításra került, hogy ez az adatkészlet több nagy nyelvi modell betanításának **fundamentuma**.

3.4.5. A prompt engineering

Az LLM-ek esetében a természetes nyelv funkcionál alapvető interfészként, és annak alkalmazásával történik meg az alapvető kezelés. A **prompt** maga a modellnek tett utasítás, és annak különböző minőségi és mennyiségi alkalmazása eltérő kimeneteket eredményez. A technológiában rejlő lehetőségek maximális potenciáljának kihasználására alakult ki a „**prompt engineering**”, amely tulajdonképpen a hatékony utasítás-kiadás módszertanát jelenti.

A megfelelő utasításnak megvannak az alapvető szükséges paraméterei:

1. **Egyértelműség:** ajánlatos világosan kiadni az utasítást és kerülni a bonyolult, szakzsargont használó megfogalmazásokat.
2. **Konkrétság:** a rövid karakterszámú bevitel helyett inkább „hosszú ívű”, tartalmukban konkretizált szövegezés ajánlott, amelyben lehetőség szerint teljeskörűen, kontextusban testet öltve jelenjen meg a kívánt álláspont.

A **prompt engineering** névvel ellentétben **nem nevezhető mérnöki tevékenységnek**, de bizonyos, hogy az erre vonatkozó tudás ismerete **elengedhetetlen** eleme az LLM hatékony használatának. A tevékenység támogatására használható weboldal a PromptHub¹⁴⁸, ahol jelentős mennyiségben érhetőek el előregyártott promptok a kívánt célokhoz.¹⁴⁹

¹⁴⁷ Common Crawl. Common Crawl maintains a free, open repository of web crawl data that can be used by anyone. Elérhető: <https://www.commoncrawl.org/> (letöltve: 2024.02.15.).

¹⁴⁸ Prompthub. Elérhető: <https://www.prompthub.us>.

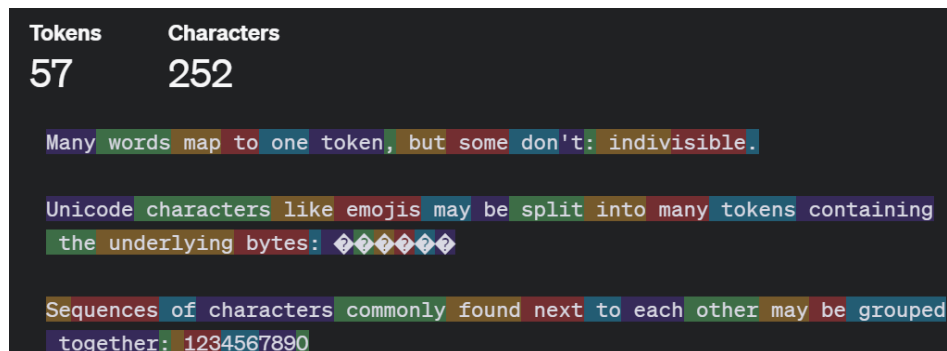
¹⁴⁹ Isabelle Nguyen (2023.06.15.). The Beginner's Guide to LLM Prompting. Elérhető: <https://haystack.deepset.ai/blog/beginners-guide-to-llm-prompting> (letöltve: 2024.02.15.).

3.4.6. A tokenizáció

A tokenizálás a **bemeneti és kimeneti szövegek** felosztása kisebb egységekre, amelyeket így az LLM képes hatékonyabban feldolgozni. A **tokenek lehetnek szavak, szótöredékek, számok, önálló karakterek, szimbólumok vagy írásjelek**. Ez egy praktikus folyamata a szövegszerű adatok kezeléséhez, hiszen csökkenti a számítási erőforrás-szükségletet és a memóriefelhasználást. Többféle tokenizáció létezik, az **LLaMA tokenizerje a mondatok darabjait dolgozza fel**, míg **az OpenAI GPT-4** modellje például a „**Byte-Pair Encoding (BPE)**” módszert alkalmazza. Utóbbi a leggyakrabban előforduló karakter vagy byte párokat egyetlen tokenbe vonja össze. Általánosságban alkalmazható az a megállapítás, hogy **egy token körülbelül 4 karakternyi szövegnek** felel meg az angol nyelv esetében.¹⁵⁰ A 4. számú ábrán láthatunk egy vizuális szemléltetést a GPT-3.5 és a GPT-4 tokenizálására.

4. ábra: a GPT-3.5 és a GPT-4 tokenizálásának szemléltetése.

Forrás: <https://platform.openai.com/tokenizer>.



3.4.7. Képességek és a főbb alkalmazási területek

Az LLM kifejezetten jó képességekkel rendelkezik az **érvelésben, a tervezésben, a döntéshozatalban, a kontextuson belüli értelmezésben** és a **zero shot**¹⁵¹-ok kezelésében. A zero shot műveletek eredményeinek különlegessége, hogy az LLM modelleket nem kifejezetten erre a típusú feladatra fejlesztették, de **mégis emberi szintű eredményeket produkálnak**. A technológia

¹⁵⁰ Michael Humor (2023.09.10.). Understanding “tokens” and tokenization in large language models. Elérhető: <https://blog.devgenius.io/understanding-tokens-and-tokenization-in-large-language-models-1058cd24b944> (letöltve: 2024.02.15.).

¹⁵¹ A zero-shot azt jelenti, hogy a modell olyan feladatok elvégzésére is képes, amelyekre nem látott kifejezetten példákat a tanulás során.

még korai életszakaszban jár, de óriási potenciált rejt a jövőre nézve, hiszen ahogy nőnek a modellek méreteikben, egyre jobbak képességeikben is. A szövegalapú (text-based) tartalmak tekintetében az alábbiak a legfőbb alkalmazási módok:

1. **Generatív funkció** (pl.: történetírás, marketing tartalom gyártás).
2. **Összefoglalás** (pl.: jogi fogalmazások, értekezletek jegyzőkönyveinek összefoglalása).
3. **Fordítás** (pl.: természetes nyelvek között, text to code¹⁵²).
4. **Klasszifikáció** (pl.: toxicity classification¹⁵³, szentiment analízis).
5. **Chatbot** (pl.: kérdések és válaszok, virtuális asszisztens).¹⁵⁴

A legfőbb alkalmazási területek az alábbiak:

- **Egészségügy:** orvoskutatók LLM-t képeznek ki az egészségügyi tankönyvekből, kutatási dokumentumokból és a betegek elektronikus nyilvántartási adataiból.
- **Ügyfélkezelés:** chatbotokban való alkalmazás kiváló és kielégítő ügyfélélményt produkál.
- **Szoftverfejlesztés:** az LLM-ek alkalmazása remekül hasznosul a szoftverfejlesztési folyamatokban.
- **Marketing szolgáltatások:** a vásárlói visszajelzések feldolgozása, majd kiértékelése és ez alapján a tartalomképzés (generálás) remekül hasznosul a tartalomgyártásban.¹⁵⁵

3.4.8. A multimodal képességek megjelenése

Visszacsatolva a gépi tanulás során kifejtettükre, meg kell említenünk, hogy attól, hogy az LLM-ek **alapértelmezetten egy adatomódban működnek** (szöveg), nem csak ez az egy adatbeviteli lehetőség létezik. Hasonlóan bemenetek lehetnek a további adatomódok, mint például a **képi** (objektumfelismerés, képosztályozás), vagy a **hang** (beszédfelismerés) is. Manapság az LLM-ek fejlődése olyan gyors ütemben zajlik, hogy már nem példa nélküli a nagy nyelvi modellekbe – a

¹⁵² Ez a képességlehetővé teszi a modell számára, hogy emberi nyelven megfogalmazott utasításokat vagy leírásokat programozási kóddá alakítsanak át. A felhasználók leírhatják a kívánt műveletet vagy problémát természetes nyelven, és a modell képes generálni a megfelelő programozási kódot a megoldáshoz.

¹⁵³ A toxikusság szerinti osztályozás arra irányul, hogy a modell felismerje és osztályozza a szövegek káros, sértő vagy egyébként toxikus tartalmát. Ez a folyamat magában foglalja a szövegek elemzését és a potenciálisan toxikus elemek, például gyűlöletbeszéd, zaklatás, előítéletek vagy egyéb káros nyelvhasználat azonosítását.

¹⁵⁴ Large Language Models. Elérhető: <https://www.nvidia.com/en-us/glossary/large-language-models/> (letöltve: 2024.02.07.)

¹⁵⁵ Large Language Models.

szövegtípus mellett – további adatmódzatok reprezentálási képességének a beépítése. Az így létrejött rendszereket már **nagy multimodális¹⁵⁶ modellnek (Large Multimodal Model – LMM)** nevezi a terminológia, hiszen **nem csak szöveges bemenetet tudnak reprezentálni és kimenetet generálni, hanem képeket, hangokat és videókat is.** Fontos megemlíteni, hogy **nem minden multimodális rendszer LMM,** csak azok, amelyek nyelvi modellkomponenssel rendelkeznek.¹⁵⁷

A multimodális jelző az alábbiak közül egyet vagy többet jelent:

1. A **bemeneti és kimeneti mód különböző** (pl.: szöveg-kép, kép-szöveg).
2. A **bemenet multimodális** (pl.: szöveget és képet egyaránt tud reprezentálni).
3. A **kimenet multimodális** (pl.: szöveget és képet egyaránt tud generálni).¹⁵⁸

A LMM az utóbbi évek fejlesztésének eredménye, úgymond a **legújabb generációja.** További lehetőséget nyújt abban a tekintetben, hogy az **összefüggések megértése már eltérő adattípusok között is létrejöhet,** valamint a különböző adattípusok együttes generálásának lehetősége is új perspektívákat nyithat meg. Ezek a lehetőségek már egy sokoldalúbb alkalmazást eredményeznek, amelyből egy **magasabb szintű kreativitás származtatható.**¹⁵⁹ A dolgozat kizárólag az LLM-ekre fókuszál.

3.5. Nagy nyelvi modellek napjainkban

2022 novemberében a ChatGPT megjelenése és a vele járó népszerűség felforgatta a világot. Képességeivel megdöbbsentette a társadalom egyszerű tagjait, hatása elérte a szoftverfejlesztő vállalatokat is. Számos szervezet gondolta újra innovációs politikáját és dinamikusabban fordult az AI fejlesztések irányába. Ezzel az időszakkal egyidejűleg és ezt követően már számos új LLM bemutatása történt meg. A megjelent modellek egy része technológiai nagyvállalatok tulajdona és **előfizetés szükséges** hozzáférésükhöz, másik része **nyílt forráskódú** (open-source), így bárki szabadon használhatja. A 5. ábrán láthatjuk az LLM-ek megjelenésének idővonalát, ahol a vonal

¹⁵⁶ A szó jelentése és fogalma: olyan kommunikációs forma vagy eszközrendszer, amely többféle érzékszervi csatornát használ. Forrás: www.nonprofitmedia.hu. (letöltve: 2024. 02.18.).

¹⁵⁷ Chip Huyen (2023.10.10). Multimodality and Large Multimodal Models (LMMs). Elérhető: https://huyenchip.com/2023/10/10/multimodal.html#part_1_understanding_multimodal (letöltve: 2024. 02.18.).

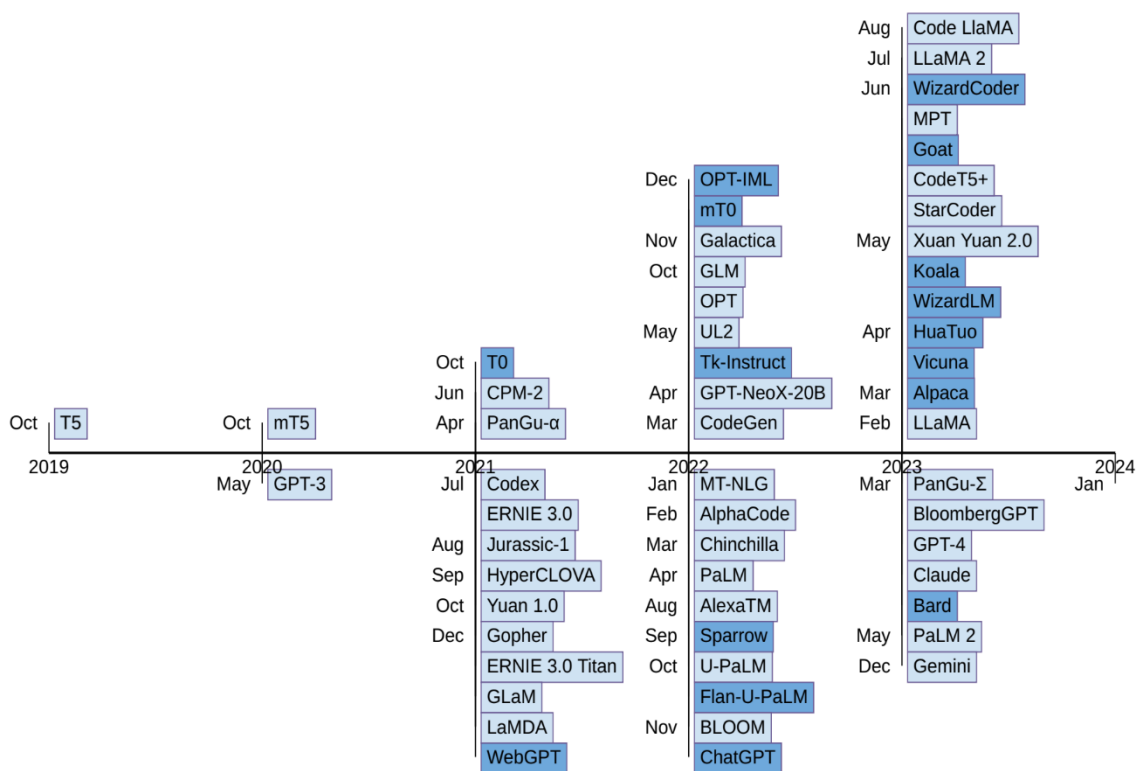
¹⁵⁸ Chip Huyen (2023).

¹⁵⁹ Chip Huyen (2023).

feletti modellek nyílt forráskódúak, míg az alsók szolgáltatók által biztosított nagy nyelvi modellek. Utóbbi csoportot gyakran nevezik „hostolt” vagy „hostolható” modelleknek, hiszen üzemeltetésüket a szolgáltató végzi és elérésük általában webes felületen, vagy API-n keresztül történik. A világokékkal jelölt rendszerek **előre tanított modellek**, míg a sötétkékek további finomhangoláson is átesett, úgynevezett „**instruction-tuned**”¹⁶⁰ modellek.

5. ábra: az utóbbi években megjelent nagy nyelvi modellek időrendben.

Forrás: Humza Naveed et al. a comprehensive overview of large language models.



Láthatjuk, hogy az utóbbi években a különböző modellek megjelenésének gyakorisága, valamint a nyílt forrású megoldások mennyisége egyaránt növekszik. A következő alfejezetekben röviden és a teljesség igénye nélkül példálózunk az utóbbi időszakban bemutatott nagy nyelvi modellekből.

¹⁶⁰ Az "instruction-tuning" egy későbbi fázis a modell fejlesztésében, amely speciális finomhangolást jelent, ahol a modellt arra képezik, hogy hatékonyan válaszoljon felhasználói utasításokra vagy direktívákra. Ez különösen fontos azokban az esetekben, amikor a modellt olyan feladatokra használják, ahol az utasítások értelmezése és azoknak megfelelő válaszok generálása kulcsfontosságú.

Kitérünk azok fejlesztőire, méreteikre¹⁶¹ és elérhetőségükre, valamint a velük kapcsolatos specifikumokra.

3.5.1. Hostolt nagy nyelvi modellek

Név: GPT-4. **Fejlesztő:** OpenAI. **Méret:** nem ismert. **Elérhetőség:** API hozzáférés.

Jelenleg az Open AI legfrissebb multimodális modellje és népszerűsége okán érintjük.¹⁶² Bemutatása 2023. március 14-én történt meg. A GPT-4 (Generative Pre-train Transformer) transzformátor architektúrán alapul és képes bemenetként reprezentálni szöveget és képeket egyaránt. Kiváló teljesítményt nyújt számos feladatban, beleértve a szakmai orvosi és jogász vizsgákat is. Mérete nem ismert, de korábbi modelleszaládjának legnagyobb verziója, a 2020-ban bemutatott GPT-3 175 B paraméterrel rendelkezett.¹⁶³ A GPT-3 megjelenésekor a legnagyobb paraméterrel rendelkező LLM volt és képességei magas értékeléseket értek el, különösen kiváló volt a „few-shot”¹⁶⁴ feladatokban, amely során a prompt néhány megadott példát adnak meg, majd ez alapján kell tartalmat generálni a modellnek. A GPT-4 korábbi verzióival ellentétben már képes már körülbelül **50 oldalnyi szöveges formátumot is reprezentálni**. Jelenleg az API-n keresztüli hozzáférése ChatGPT Plus előfizetéssel elérhető.¹⁶⁵

Név: PaLM 2. **Fejlesztő:** Google. **Méret:** 540 B. **Elérhetőség:** API hozzáférés.

A PaLM (Pathways Language Model) szintén transzformátor architektúra, de nem egy hagyományos jellegű LLM kezdeményezés eredménye. A Google Pathways¹⁶⁶ kutatási kezdeményezésén alapul, amelynek célja egy olyan AI megoldás létrehozása, amely nem csak kizárólag egy, hanem több, **eltérő típusú felhasználásra képes**, valamint gyorsan tanulja **meg új**

¹⁶¹ A modellek méreteinél az angol terminológia által alkalmazott „B”, vagyis billion, jelzést alkalmazzuk. 1 B = 1 milliárd paramétert jelöl.

¹⁶² GPT-4. Elérhető: <https://openai.com/research/gpt-4> (letöltve: 2024.02.12.).

¹⁶³ Nagesh Singh Chauhan (2022.05.04.). OpenAI GPT-3: Understanding the Architecture. Elérhető: <https://www.theaidream.com/post/openai-gpt-3-understanding-the-architecture> (letöltve: 2024.02.12.).

¹⁶⁴ A „few-shot-learning” módszert az LLM-ek modellek tanításánál alkalmazzák. Segítségével megtanítják a modelleket olyan feladatok hatékony megoldására, amikor csak néhány példa alapján kell/lehet megoldani a feladatot.

¹⁶⁵ Suleman Kazi & Adel Elmahdy (2023.10.17). Top Large Language Models (LLMs): GPT-4, LLaMA 2, Mistral 7B, ChatGPT, and More (2023). Elérhető: <https://vectara.com/blog/top-large-language-models-llms-gpt-4-llama-gato-bloom-and-when-to-choose-one-over-the-other/#gpt4> (letöltve: 2024.02.12.).

¹⁶⁶ Jeff Dearn (2021.10.28.). Introducing Pathways: A next-generation AI architecture. Elérhető: <https://blog.google/technology/ai/introducing-pathways-next-generation-ai-architecture/>.

feladatok végrehajtását. A PaLM kezdeti eredményeinek bemutatása 2022 áprilisában történt meg,¹⁶⁷ majd 2023 márciusában megnyitották az API hozzáféréseket. A PaLM2 nyilvános bemutatását már 2023 májusában végrehajtották.¹⁶⁸ A PaLM egyik figyelemreméltó képessége, hogy már első megjelenésekor képes volt magyarázatokat generálni olyan komplex logikát igénylő feladatok esetében, mint a viccek magyarázata.¹⁶⁹ A PaLM2-nek specifikus változatai is vannak, mint a Med-PaLM 2, amely az élettudományokra és az orvosi információkra van finomhangolva, vagy a **Sec-PaLM**,¹⁷⁰ **amely a kiberbiztonsági elemzésekre összpontosít.**¹⁷¹ Jelenleg a PaLM2 API csak az USA-ból elérhető.

Név: MT-NLG. **Fejlesztő:** Nvidia / Microsoft. **Méret:** 530 B. **Elérhetőség:** API hozzáférés.

A Turing NLG 17B¹⁷² és a Megatron-LM¹⁷³ utódjaként is nevezett MT-NLG-t (Megatron-Turing Natural Language Generation) 2021. októberében hozták nyilvánosságra, és megjelenésekor meghaladta a GPT-3 eredményeit. A szintén transzformer alapú Megatron-LM architektúráját és a DeepSpeed¹⁷⁴ optimalizálási környezetét használja. Célirányosan arra lett alkotva, hogy **koherens és kontextuálisan releváns szöveget alkosson.**¹⁷⁵ Nagy pontosságú feladat-végrehajtásra képes olyan természetes nyelvi feladatok során, mint az előrejelzés, szövegértés és az ész érvelés.¹⁷⁶ Az MT-NLG hatalmas mennyiségű paraméterszámának köszönhetően unikumnak számít, ugyanakkor megfigyelések szerint a modell **erősen sztereotip és erős torzításokat** vett fel a képzési

¹⁶⁷ Aakanksha Chowdhery et al.(2022): PaLM: Scaling Language Modeling with Pathways. Google Research. <https://doi.org/10.48550/arXiv.2204.02311>.

¹⁶⁸ Sean Michael Kerner (2023): Pathways Language Model (PaLM). Elérhető: <https://www.techtarget.com/whatis/definition/Pathways-Language-Model-PaLM> (letöltve: 2024.02.12.).

¹⁶⁹ Suleman Kazi & Adel Elmahdy (2023).

¹⁷⁰ Sunil Potti (2023.04.25.). Supercharging security with generative AI. Elérhető: <https://cloud.google.com/blog/products/identity-security/rsa-google-cloud-security-ai-workbench-generative-ai> (letöltve: 2024.03.05.).

¹⁷¹ Sean Michael Kerner (2023).

¹⁷² Corby Rosset (2020.02.13.). Turing-NLG: A 17-billion-parameter language model by Microsoft. Elérhető: <https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/> (letöltve: 2024.02.12.).

¹⁷³ GitHub — NVIDIA/Megatron-LM. Elérhető: <https://github.com/NVIDIA/Megatron-LM> (letöltve: 2024.02.12.).

¹⁷⁴ Egy könnyen használható mélytanulás-optimalizáló szoftvercsomag, amely példátlan sebességet biztosít mint a modell képzéshez, mind a következtetésekhez. Elérhető: <https://github.com/microsoft/DeepSpeed> (letöltve: 2024.02.12.).

¹⁷⁵ Suleman Kazi & Adel Elmahdy (2023).

¹⁷⁶ Kharya & Alvi (2021).

adatokból.¹⁷⁷ Elviekben fejlesztők részére API hozzáférést biztosít az fejlesztő, de ezt a kutatás során nem sikerült hiteles forrásból megerősíteni.

3.5.2. Nyílt forráskódú nagy nyelvi modellek

Név: Llama 2. **Fejlesztő:** Meta AI, **Méret:** 5B-70B. **Elérhetőség:** nyílt forráskódú

A Meta Llama családjának ígéretes modellje 2023 júliusában lett bemutatva.¹⁷⁸ A családban számos variánsa elérhető, a **Llama Chat** például a természetes nyelvi feldolgozásra, míg a **Code Llama** pedig kódolásra lett tanítva. Utóbbi kiváló lehetőség a kódolás hatékonyságának növelésére, leginkább azért, mert két további almodellel (Python¹⁷⁹, Instruct¹⁸⁰) rendelkezik,¹⁸¹ amelyek további specifikus alkalmazást tesznek lehetővé. A Llama a nyílt forráskódú modellek közül **a legjobb eredményekkel** rendelkezik. A modellek kutatási cézzal ingyenesek, valamint mivel OS (Open Source) modellről van szó, számos további kiegészítő szkript elérhető a modellhez, ami finomhangolási lehetőségeket biztosít a saját használatú alkalmazásra.¹⁸²

Név: YALM 100B. **Fejlesztő:** Yandex **Méret:** 100 B. **Elérhetőség:** nyílt forráskódú

Ez a modell a Yandex orosz multinacionális informatikai vállalat által 2022. júniusában közzétett GPT architektúra szövegszerű tartalmak generálására és feldolgozására.¹⁸³ A modell alapja MT-NLG-nél is említett nyílt forráskódú transzformátor a Megatron-LM. Apache 2.0 licenc alatt jelent meg, amely kutatási és kereskedelmi felhasználást is lehetővé tesz. Betanítása 1,7 TB online szövegből, könyvekből és számtalan egyéb forrásból álló korpuszon történt, **főként orosz nyelvű tartalommal** (a teljes adatkészlet 75%-a).¹⁸⁴

¹⁷⁷ Kharya & Alvi (2021).

¹⁷⁸ Hugo Touvron és mtsai (2023): Llama 2: Open Foundation and Fine-Tuned Chat Models. GenAI, Meta. DOI: <https://doi.org/10.48550/arXiv.2307.09288>.

¹⁷⁹ Kifejezetten a Python programozási nyelvre alkotva azzal, hogy képzés Python kódokkal lett súlyozva.

¹⁸⁰ Az Instruct almodell leginkább a természetes nyelv és a kódolás közötti kapcsolatok kezelésére lett tervezve.

¹⁸¹ Meta (2023.08.23): Introducing Code Llama, a state-of-the-art large language model for coding. Elérhető: <https://ai.meta.com/blog/code-llama-large-language-model-coding/> (letöltve: 2024.02.12.).

¹⁸² Top Large Language Models (LLMs): GPT-4, LLaMA 2, Mistral 7B, ChatGPT, and More (2023).

¹⁸³ Mikhail Khrushchev (2022.06.23.). Yandex Publishes YaLM 100B. It's the Largest GPT-Like Neural Network in Open Source. Elérhető: <https://medium.com/yandex/yandex-publishes-yalm-100b-its-the-largest-gpt-like-neural-network-in-open-source-d1df53d0e9a6> (letöltve: 2024.02.12.).

¹⁸⁴ Yandex YaLM-100B. Elérhető: <https://github.com/yandex/YaLM-100B> (letöltve: 2024.02.12.).

3.5.3. A dark webbel kapcsolatos nagy nyelvi modellek

A dolgozat témájára tekintettel az általános LLM lehetőségek mellett ki kell térnünk a rosszindulatú, illegális tevékenységekre is, hiszen egyrészt OSINT tevékenységek rosszindulatú célzattal is történhetnek, másrészt a kutatás fókuszába kerültek olyan modellek is, amelyeknek megjelenése a dark webhez köthető. A következőkben röviden bemutatásra kerül a DarkBERT, ami nem illegális használatra készített LLM, de erősen kötődik a közeghez, valamint olyan LLM modellek kerülnek említésre, amelyek **a dark web mélyéről** érhetőek el, és egyértelműen **ártó szándékkal fejlesztették** azokat.

DarkBERT

A 3.2.2. fejezetnél már említést tettünk a darkweb jellemzőiről, valamint az 4.4.3. fejezetnél tárgyaltuk a LLM tanítási adathalmazainak típusait. Érdekes kísérletnek bizonyult egy japán AI alapú, kiberfenyegettség-hírszerzésre specializálódott vállalat, az S2W¹⁸⁵ szakemberei és dél-koreai kutatók közös próbálkozása, amely során a **dark webről indexelt tartalom**ból készített **körpuszon tanítottak be** egy BERT¹⁸⁶ modellt. Az így létrejött, DarkBERT elnevezésű LLM-et összehasonlítottak további alap-modelltípusok (BERT, RoBERTa¹⁸⁷) képességeivel, valamint kiberfenyegetések kezelésével kapcsolatos információk kezelésével. Kutatásuk bizonyította, hogy a DarkBERT bár alacsony különbséggel, de **jobb eredményt ért el** az alapmodellekhez képest, valamint jobban eligazodott a dark web-specifikus feladatokkal¹⁸⁸ kapcsolatban és **kiberbiztonsági fenyegetések azonosításában**.¹⁸⁹

¹⁸⁵ DarkBERTダークウェブで訓練された最初のAI. (japán – magyar fordításban: „DarkBERT az első sötét weben edzett mesterséges intelligencia) Elérhető: <https://s2wjapan.com> (letöltve: 2024.02.12.).

¹⁸⁶ A Google kutatói által 2018-ban kifejlesztett BERT az első LLM-ek egyike. Megdöbbentő eredményeivel elismertette a transzformer technológiát.

¹⁸⁷ „Robustly Optimized BERT Approach” rövidítése, a RoBERTa egy BERT-változat, amelyet a Meta hozott létre a Washingtoni Egyetemen együttműködésben.

¹⁸⁸ A kiberbiztonsági fenyegetések mellett szemantikus kereséseket végeztek a kábítószer-eladásokkal kapcsolatban is, amely szintén hatékonyabbnak bizonyult az alapmodellhez képest.

¹⁸⁹ Youngjin Jin, Eugene Jang, Jian Cui, Jin-Woo Chung, Yongjae Lee & Seungwon Shin (2023): DarkBERT: A Language Model for the Dark Side of the Internet. DOI: <https://doi.org/10.48550/arXiv.2305.08596>.

WormGPT:

2023. július 13-án a SlashNext kiberbiztonsági vállalat kutatói blogbejegyzést tettek közzé, amelyben egy dark webes hackerfórumon eladásra hirdetett LLM eszköz bemutatását osztották meg.¹⁹⁰ A WormGPT a GPT modellek „**blackhat**” **alternatívája** és kifejezetten rosszindulatú kiber tevékenységek segítésére képezték ki. Egy viszonylag elavultnak gondolt GPT-J¹⁹¹ nyelvi modellen alapuló illegális kibertéri információkat tartalmazó korpuszon tanított LLM-megoldás. A vállalat szakemberei BEC¹⁹² támadásokkal kapcsolatos tartalmak készítésében vizsgálták generatív képességeit, és az **eredmények nyugtalanítóak voltak** abban a tekintetben, hogy rendkívül meggyőző és **stratégiai szempontból kifinomult e-mail üzenetet készített**. Az eredmény a ChatGPT képességeivel közel megegyezőnek minősítették, ugyanakkor a WormGPT esetében nincsenek etikai korlátok, sem security layerek¹⁹³ A WormGPT-t a megadott információk alapján kiberbiztonsági kutatók tevékenység-támogatására készítették és minden tartalomgenerálás esetében felhívja a figyelmet az etikai irányelvek betartásának fontosságára, de ez mit sem gyengíti azt a megállapítást, hogy hatékonysága **a rosszindulatú tevékenységek támogatásában valósul meg** leginkább.¹⁹⁴

Bár a WormGPT megjelenése sem tehető nagyon régre, ezt követően más, hasonló célú LLM modell jelent meg (pl.: **FraudGPT**), és valószínűleg ez a jelenség ezzel nem ért véget. Az előzőekben egy pillanatnyi betekintést nyújtottunk abba, hogy az LLM modellek egyes esetei milyen hatást gyakorolnak a kiberbiztonságra, a továbbiakban a modellek kiberbiztonsági aspektusaira térünk ki.

¹⁹⁰ Daniel Kelley (2023.07.13). WormGPT – The Generative AI Tool Cybercriminals Are Using to Launch Business Email Compromise Attacks. Elérhető: <https://slashnext.com/blog/wormgpt-the-generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks> (letöltve: 2024.03.12.).

¹⁹¹ GPT-J. Transformers documentation. Elérhető: https://huggingface.co/docs/transformers/model_doc/gptj (letöltve: 2024.03.12.).

¹⁹² Business Email Compromise a kiberbűnözés egy olyan fajtája, ahol a csaló e-mailt használ arra, hogy valakit pénz küldésére vagy bizalmas vállalati adatok kiszolgáltatására csaljon ki. Forrás: <https://www.microsoft.com/en-us/security/business/security-101/what-is-business-email-compromise-bec>.

¹⁹³ A security layerek a LLM modellek azon védelmi megoldásai, amelyek megakadályozzák a szoftver rosszindultú alkalmazását.

¹⁹⁴ Charlie Osborne (2023.10.24.). WormGPT: What to know about ChatGPT's malicious cousin. Elérhető: https://www.zdnet.com/article/wormgpt-what-to-know-about-chatgpts-malicious-cousin/#google_vignette (letöltve: 2024.03.12.).

3.5.4. Valós idejű információgyűjtési képesség

A ChatGPT-t üzemeltető OpenAI 2023 őszén lehetővé tette népszerű modelljén az internetes böngészés lehetőségét a **Bing integrálásával**, ezzel egyidejűleg a ChatGPT is integrálásra került a Bing keresőjébe. Bár előbbi módosítást már ugyanazon év nyarán is végrehajtották béta verzióként, azonban akkor problémák merültek fel, mert a Bing modullal **végzett keresések képesek voltak bejutni a különböző online fizetős szolgáltatások mögé**, majd megosztani az ott tárolt tartalmakat.¹⁹⁵ ¹⁹⁶ Az OpenAI eltávolította a Bing modult a ChatGPT-ből, majd a hiba javítását követően újra visszahelyezte. A történetekből láthatjuk, hogy egy olyan élenjáró és tőkeerős, kifejezetten AI megoldásokat fejlesztő vállalat esetében is **incidens történt** a szolgáltatással összefüggésben. A valós idejű információkeresés LLM képességével nincs egyedül az OpenAI, hiszen a **Google Gemini LLM-je a Google Search** keresőmotorját használja. A fontos megállapítás ezzel kapcsolatban az, hogy a jelenleg **hostolt LLM megoldások már alkalmazzák a valós idejű információgyűjtés képességét**.¹⁹⁷

Sofia Eleni Spatharioti és munkatársai egy 2023-ban végzett kísérletükben összehasonlították a **hagyományos és az LLM alapú információ kereséssel** kapcsolatos hatékonyságot. Megállapították, hogy a felhasználóknak **egyre több LLM eszköz elérhető keresési céllal**, és az így végzett keresések **gyorsabbak, kevesebb lépésből elérhetőek és a résztvevők kedvezőbb értékeléssel látták el az LLM alapú kereséseket**. Az LLM képességeket meghaladó feladatban azonban jelentős visszaesést mértek, hiszen téves **információt szolgáltatott**, így a felhasználók **bizalma is sérült**. Bizonyos, hogy az LLM alapú keresés **meg fogja változtatni az emberek internet információkeresési gyakorlatát**. A feltárt megállapítások azt sugallják, hogy a keresés hatékonysága és az azzal kapcsolatos társadalmi elégedettség valószínűleg növekedni fog.¹⁹⁸

¹⁹⁵ Kevin Okemwa (2023.06.04.) ChatGPT pauses Bing integration to stop people from bypassing paywalls. Elérhető: <https://www.windowcentral.com/software-apps/chatgpt-pauses-bing-integration-to-stop-people-from-bypassing-paywalls> (letöltve: 2024.03.13.).

¹⁹⁶ Holupredictions (2023). It looks like you can use ChatGPT to bypass paywalls. Elérhető: https://www.reddit.com/r/ChatGPT/comments/14j8q1u/it_looks_like_you_can_use_chatgpt_to_bypass/ (letöltve: 2024.03.13.).

¹⁹⁷ Reece Rogers (2023.10.15). How to Use ChatGPT's 'Browse With Bing' Tool—Plus 6 Starter Prompts. Elérhető: <https://www.wired.com/story/chatgpt-browse-with-bing-internet-connected-openai/> (letöltve: 2024.03.13.).

¹⁹⁸ Sofia Eleni Spatharioti, David M. Rothschild, Daniel G. Goldstein, Jake M. Hofman (2023). Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment. Microsoft Research, New York City. DOI: <https://doi.org/10.48550/arXiv.2307.03744> p.15.

Az említett hostolt modellekkel végzett valós idejű nyílt forrású információkeresés lehetősége a **H1 hipotézis** felvetésével összefüggésben van. Fontos felvetni azt a megállapítást, hogy közvetlenül, végponttól-végpontig **nagy halmazi információt kerestetni** a ChatGPT-vel vagy a Geminivel (a Google legújabb modellje) **nem lehet**, hiszen a számítási **kapacitás-kihasználás korlátozása**, valamint a **gyors idejű output követelmények okán** a válaszadások nagyon „kapzsiak”, csak minimális mennyiségű információt gyűjtenek össze. Tehát **nem hatékonyak**, de alkalmazhatóak önálló OSINT eszközként.

3.6. A nagy nyelvi modellek kiberbiztonsági sebezhetőségei

A ChatGPT megjelenése óta rendszeresen látnak napvilágot olyan sajtóhírek, amelyek a nagy nyelvi modellek „meghackeléséről” szólnak, valamint ez a téma népszerű tudományos körökben is. Ahogy minden szoftveres rendszerrel szemben **fentáll a rosszindulatú beavatkozás kockázata**, ez a jelenség az LLM-ek viszonylatában is **értelmezhető**. A nagy nyelvi modellek kiberbiztonsági aspektusainak teljes körű vizsgálata érdekes téma. Annak mélyreható tanulmányozása bizonyára egy önálló tudományos dolgozat terjedelmét kitenné, de ez a dolgozat erre nem vállalkozik.

Azonban, hogy a megfelelő védelmi stratégiákat kidolgozhassuk, külön vizsgálati alapot kell képezzen az LLM modellek **kiberbiztonsági gyengeségei is**. Az alábbiakban egy rövid összefoglalót kívánok adni a Open Web Application Security Project (OWASP) szakértői által meghatározott LLM sebezhetőségekről, amelynek top 10-es listája az alábbi.¹⁹⁹

1. Prompt injection

Ez az LLM elleni kibertámadások leggyakoribb típusa, amely úgy manipulálja az LLM működését, hogy az **figyelmen kívül hagyja a rendszer biztonsági mechanizmusait** és a hacker utasítását hajtsa végre. Ennek érdekében olyan speciális utasítást készítenek (promptot), amely **eltéríti a nagy nyelvi modellt az alapértelmezett működésétől**. Lehet közvetlen és közvetett. Előbbi rögtön az

¹⁹⁹ Nihad Hassan (2024.03.06). Explore mitigation strategies for 10 LLM vulnerabilities. Elérhető: <https://www.techtarget.com/searchEnterpriseAI/tip/Explore-mitigation-strategies-for-LLM-vulnerabilities>.

LLM bemenetét használja fel az **injektáláshoz** (közvetlen kapcsolat), utóbbi viszont olyan tartalom (pl.: weboldal) reprodukálásra irányítja az LLM-et, **amelyben káros kód van elhelyezve**.²⁰⁰

2. Bizonytalan kimenetkezelés

Ez a fajta támadás akkor jön létre, amikor támadók úgy manipulálják a modellt, hogy annak **generatív képessége rosszindulatú céllal vezérelve, ellenőrzés nélkül** valósuljon meg. Ezzel félrevezethető információk, de akár rosszindulatú scriptek is generálhatók, amelyek felhasználhatók például olyan további támadásokhoz, mint például a cross-site scripting (XSS).²⁰¹

²⁰²

3. Adatmérgezés

A modellek a tanítási és az feldolgozási fázisokban egyaránt jelentős mértékű adatot használnak fel. Ezekben a folyamatokban a **nagy adathalmazok** (pl.: szöveges korpuszok, nyílt forrású adatok, pénzügyi, marketing és értékesítési adatok) **manipulálása** ahhoz vezethet, hogy az LLM hibás mintákat tanul be és/vagy hibásan generál tartalmat, így jelentősen sérül a **modell integritása és megbízhatósága**.²⁰³

4. DoS

Denial-of-service attack, vagyis **szolgáltatás-megtagadási támadás**. A szolgáltatásmegtagadás egy **releváns kiberbiztonsági fenyegetés**, tekintettel arra, hogy a nagy nyelvi modelleknek **jelentős hardver és szoftvererőforrásra van szükségük**, így amennyiben a támadók nagy mennyiségű forgalommal árasztják el az LLM infrastruktúráját, **az a rendelkezésre állás megszűnéséhez vezet**. Ugyanakkor olyan **lekérdezések** is végrehajthatók a modell használata közben, amelyek **túlterhelik** azt, így **csökkentve a teljesítményt** és az elérhetőséget.²⁰⁴

5. Ellátási lánc sebezhetősége

²⁰⁰ Nihad Hassan (2024.03.06).

²⁰¹ A számítógépes sebezhetőség egy fajtája: Tulajdonképpen egy szimpla injection megoldás, amely során egy rosszindulatú kód beillesztése történik meg egy weblapra, amelyet futtatni fog a megtámadott fél.

²⁰² Nihad Hassan (2024.03.06).

²⁰³ Nihad Hassan (2024.03.06.).

²⁰⁴ Nihad Hassan (2024.03.06.).

Természetesen a LLM infrastruktúra is számos, több gyártótól származó szoftverből (külső könyvtár és pluginek) és hardverből épül fel, így a hosting szolgáltatáshoz vagy az önálló futtatáshoz **alapvető komponensek sebezhetőségei** is kihasználhatók egy kibertámadásban. Az ellátási lánc sebezhetőségei közé sorolhatjuk a **tanítóadatokban elhelyezett sebezhetőséget** is. A biztonsági réseket kihasználva manipulálható a rendszer működése.²⁰⁵

6. Érzékeny információ szivárgása

A nagy nyelvi modellek külső beavatkozás nélkül is oszthatnak meg **érzékeny információkat** (pl.: tanítási adatok részleteit, architektúrára jellemzőket.). Továbbá a képzési adatok – kiváltképp a nyílt forrású információkból alkotott korpuszok – tartalmazhatnak **harmadik félre vonatkozó különleges adatokat** (pl.: személyes adatok, jelszavak) amelyek nyilvánosságra hozása adatvédelmi incidenst indukálhat.²⁰⁶

7. Nem biztonságos bővítmény kezelés

Az nagy nyelvi modellek képességeinek bővítése céljából **több beépülő modult** vezérelnek, például a ChatGPT esetében a **Bing szolgáltatója a valós idejű információgyűjtést**. Azonban ezek a modulok tovább növelik a kibertámadások kockázatait, hiszen a modulok külső kapcsolaton keresztül elérhető és manipulálható a nagy nyelvi modell.²⁰⁷

8. Túlzott meghatalmazás

A modellekkel kapcsolatban jelen van egy igen erős fenyegetettség, ami tulajdonképpen a rendszerből eredő **biztonsági résnek is minősíthetünk**. Amennyiben az adott modell **túl széleskörű hozzáféréssel rendelkezik** (pl.: kapcsolt szolgáltatások, beépülő modulok, jelentős kibertéri kapcsolat, fennáll a lehetősége és veszélye annak, hogy akár önállóan, akár manipuláció hatására váratlan kimenetelként a modell rosszindulatú, de akár **támadó célú műveleteket hajthat végre** és jelentős károkat okozhat.²⁰⁸

²⁰⁵ Nihad Hassan (2024.03.06.).

²⁰⁶ Nihad Hassan (2024.03.06.).

²⁰⁷ Nihad Hassan (2024.03.06.).

²⁰⁸ Nihad Hassan (2024.03.06.).

9. Túlzott függőség

Ez a sebezhetőség nem is a LLM-ben keresendő, hanem az alkalmazó emberi jellemzőiből, viselkedéséből. Ha **túlzott a bizalom a modell irányába**, akkor az számottevő **hatást gyakorol az emberi döntéshozatalra**. Ha a modell kimenetele nem kerül ember általi kritikus elemzésre, valamint mivel a modellek gyakran „hallucinálnak” vagy hibás eredményeket generálnak, ebből a kettősből **dezinformáció, vagy sebezhetőség** is keletkezhet. Például egy **text to code** képesség alkalmazása esetén a modell által **megírt kód tartalmazhat sebezhetőségeket**, ami felhasználás esetén kiberbiztonsági sérülékenységet hoz létre.²⁰⁹

10. Modell lopás

A LLM-ek **kiberbiztonsága kiemelten fontos**, hiszen óriási források (anyagi, szakmai, infrastrukturális) felhasználása árán sikerül létrehozni egy jól működő modellt. Azonban, ha sikertelenül védjük modellünket, **illetéktelenek férhetnek hozzá és a működésre vonatkozó információkat szerezhetnek meg**, de akár le is **klónozzhatják** a modellt. Az elloptott modell jelentős gazdasági és piaci veszteséget okoz a szervezetnek, valamint a későbbiekben akár **rosszindulatú tevékenységekre is felhasználható**.²¹⁰

3.7. Nagy nyelvi modellek SWOT analízise

Yupeng Chang és munkatársai az LLM-ek képességeinek értékelési módszerét vizsgálták tanulmányukban, amely eljárás során részletezik azok erősségeit és gyengeségeit. Az általuk tett megállapítások felhasználásával, valamint az eddigiekben részletezett és saját információim alapján – az OSINT tevékenységek vizsgálatához hasonlóan – SWOT analízis készült a nagy nyelvi modellekről a kiberbiztonság aspektusára vonatkozó jellemzők kiegészítésével.²¹¹

²⁰⁹ Nihad Hassan (2024.03.06.).

²¹⁰ Nihad Hassan (2024.03.06). Explore mitigation strategies for 10 LLM vulnerabilities. Elérhető: <https://www.techtarget.com/searchEnterpriseAI/tip/Explore-mitigation-strategies-for-LLM-vulnerabilities>.

²¹¹ Yupeng Chang et al (2023). A Survey on Evaluation of Large Language Models. <https://doi.org/10.48550/arXiv.2307.03109>.

6. ábra: Az LLM modellek SWOT elemzése. Forrás: a szerző szerkesztése.

NAGY NYELVI MODELLEK SWOT ANALÍZISE (kiberbiztonsági jellemzőkkel kiegészítve)		
Erősségek (strengths)		Gyengeségek (weaknesses)
Belső tényezők	• Aritmetikai és logikai következtetés. • Döntéshozatal. • Folyamatos fejlődés. • Folyékony szövegalkotás. • Kontextuális megértés. • NLP teljesítmény • Nyelvi megértés. • Teljesítmény. • Többnyelvű képességek. • Valós idejű információ-gyűjtés.	• Adatfüggés. • Hallucináció. • Értelmezhetőség. • Gyenge NLI²¹² teljesítmény. • Hiteltelenség. • Jelentős költségek. • Toxicitás. • Képzési idő. • Korlátozott ismeretkezelés. • Korlátozott képességek. • Nem optimális környezet. • Szolgáltató függőség. • Túlzott meghatalmazás • Bővítmény-kezelés • Érzékeny információ-szivárgás • Bizonytalan kimenetkezelés
Lehetőségek (opportunities)		Fenyegetések (threats)
Külső tényezők	• Finomhangolás. • Globális szerepvállalás. • Rugalmasság. • Sokoldalúság. • Új területek forradalmasítása. • Valós idejű információkeresés.	• Adatszivárgás. • Elévülés. • Etikai. • Polgári jogi kihívások. • Prompt injection. • Rosszindulatú felhasználás. • Szabályozási kihívás. • Társadalmi megítélés. • Modell lopás. • Ellátási lánc sebezhetősége. • DoS. • Adatmérgezés.

A SWOT mátrixban több olyan jellemző lett szerepeltetve, amelyeknek magyarázatát ebben a főfejezetben már megtettem, így a továbbiakban csak a legfontosabb, illetve a nem egyértelmű tulajdonságok részletezése fog megtörténni.

Erősségek (strengths)

Az LLM-ek rendkívül hatékonyak a **folyékony szövegalkotásban** (pl.: tartalomkészítés, párbeszédalkotása.) Figyelemre méltó **NLP képességekkel rendelkeznek** (pl.: szemantika,

²¹² Az NLI (Natural Language Inference), azaz természetes nyelvi következtetés. Az NLI egy olyan folyamat vagy feladat, amely során a rendszer megpróbálja megállapítani, hogy egy adott állítás logikailag következik-e egy másik szöveges állításból, ellentmondásban áll-e vele, vagy semleges viszonyban van vele.

szintaktika, szentimentelemzés stb.), hiszen a legmagasabb pontszámokat érik el ezekben a feladatokban. **Aritmetikai és logikai következtetési** képességeik is erősségnek említhetőek. **Kontextuális megértésre** képesek, hiszen nagyobb méretű szöveges tartalom teljeskörű értelmezésében, kivonatolásában elismert teljesítményük. A **nyelvi feldolgozási feladatokban** is kiválóak, jól teljesítenek a természetes (pl.: nyelvi fordítások) és programozási nyelvek feldolgozásában és generálásában (pl.: text to code képességek). Remekül használhatók hatalmas mennyiségű adatokkal összefüggő **döntéstámogatásra**. Minden iterációval egyre pontosabbak és hatékonyabbak lesznek, így biztosítva a **folyamatos fejlődést**. **Többnyelvű képességek:** a többnyelvű adatkészleten végzett képzésnek köszönhetően globálisan alkalmazhatók. Egyes modellek korlátozottan képesek **valós idejű információk gyűjtésére**, amely további lehetőségeket biztosít az alkalmazhatóságnak.

Gyengeségek (weaknesses)

Gyenge NLI teljesítmény. Bár a legfrissebb modellek egyre jobb eredményeket érnek el a témában, véleményem szerint még mindig a gyengeségekhez kell sorolnunk az emberi ellentmondások nehézkes kezelését (pl.: nem racionalitáson alapuló véleménykülönbségek azonosítása). Szintén ide sorolandó az események közötti szemantikai hasonlóság felismerésében, az absztrakt gondolkodás terén és az összetett kontextusban való értelmezés során megjelenő **korlátozott képesség**. **Hallucináció:** hajlamos a zavarodottsága, és a hibákra. A **nem optimális környezetben** való érvényesülés. Nem latin betűs szöveg esetén, vagy korlátozott erőforrásokat tartalmazó nyelvi környezetben a teljesítmény nem hatékony. **Káros tartalmak felerősítése és elfogultság (toxicitás):** hajlamosak befogadni, terjeszteni és potenciálisan felerősíteni a tanítási adatkészletben található káros tartalmakat. **Hiteltelenség:** a hallucináció miatt nem minősíthető hiteles forrásnak. Hostolt modellek esetén olyan szolgáltató függőség okozta problémákkal is számolni kell, mint az információbiztonsági kockázatok és a CIA²¹³ követelmények sérülésének kockázata. Professzionális modellek esetében **jelentős költségek** lépnek fel a létrehozás és üzemeltetés alatt egyaránt, így kisebb szervezetek részére nehezen hozzáférhető. A képzési folyamatok jelentős számítási kapacitást igényelnek, és még annak biztosítása során is **időigényes**. **Értelmezhetőség:** hátrány, hogy működésük nehezen érthető, kihívást jelenthet. **Korlátozott**

²¹³ Az információbiztonság három fő alapelve (confidentiality, integrity and availability). Bizalmasság, sértetlenség, rendelkezésre állás.

ismeretkezelés: az általános modellek mély szakmai tudást igénylő szakterületi specifikumok esetében finomhangolás nélkül nem eredményesek. **Adatfüggés:** a képzés sikeressége függ a felhasznált adatok mennyiségétől és minőségétől.

Lehetőségek (opportunities)

A nyílt forrású LLM-ek **sokoldalúan felhasználhatók** és feladataira **finomhangolhatók**. Az eltérő szervezetek saját szakmai tevékenységük támogatására képezhetik modelljüket. A hostolt modellek esetében az API adta lehetőségeknek köszönhetően könnyedén társíthatók különböző alkalmazásokhoz (pl.: chatbotok stb.), így egy **rugalmas felhasználást** tesz lehetővé használatuk. Ahogy egyre több területen kezdik el használni az LLM lehetőségeket, úgy **forradalmasítja meg** azokat, korábbi technológiákat vált fel és eljárásokat tesz hatékonyabbá (pl.: ügyfélszolgálati chatbotok támogatására, vagy más alkalmazásokba integrálható fordítási megoldások). Néhány újonnan bemutatott LLM modell képes a **valós idejű információkeresésre**, amely képesség új perspektívákat nyit meg az OSINT területén is. **Globális szerepvállalás:** a nyelvi akadályok leküzdésében az LLM kulcsfontosságú szereplő lehet a nemzetközi kapcsolatok színterén, bizonyosan számos feladatot fog átvenni a jövőben ezen a területen.

Fenyegetések (threats)

Az LLM-re ható külső fenyegetések köre is színesnek mondható. A legnagyobb fenyegetésként a **prompt injection**²¹⁴ jegyzik a legtöbb helyen, hiszen az LLM modellek érzékenyek a promptokra, különösen a rosszhindulatú promptokra. Az LLM-ek esetében az adatszivárgás kockázatos fenyegetés. A hostolt modellek esetén az utasítások adása során bevitt adatbevittelt követően az adatok kezelésének folyamata **nem transzparens**, így akár olyan **adatszivárgási incidens** jöhet létre, aminek megtörténtéről nincs valós időben tudomásunk. Manapság az LLM-ek esetében a feldolgozott adatokkal kapcsolatos **szerzői és tulajdonjogi tárgyban** indított keresetek kimenetele jelenleg még bizonytalan. A nyílt forrásokon végzett tanítások olyan **etikai problémákat** hoznak a felszínre, amelyek a későbbiekben jogi aktusokká válnak és bírósági szakaszig jutnak. Számos technológia, valamint az LLM-ek esetében is tény, hogy villámgyors elterjedésükkel a jog ereje

²¹⁴ A prompt injection során a felhasználó vagy másik program úgy manipulálja a bemeneti parancsot vagy utasítást (a promptot), hogy a rendszer nem várt vagy nem kívánt műveleteket hajtson végre. Ez különösen releváns az LLM modellek esetében.

nem bír lépést tartani, így a **szabályozatlanságból** eredő problémák is számottevővé válhatnak. Említeni kell még az MI-k **társadalmi megítélését**, hiszen mivel működése a laikusoknak nehezen érthető, valamint társadalmi és gazdasági hatása még kialakulóban van, így úgy ítélik meg alkalmazását, hogy veszélyezteti az emberek munkahelyeit. Fennáll a kockázata a rosszindulatú felhasználásnak is, hiszen nyílt forráskódú modellek esetében bárhogyan, hostolt modellek esetén a security layerek megkerülésével káros célokra (információgyűjtés, félrevezető információk generálása, kibertámadások) is alkalmazhatók. Mindemellett a már említett rendkívül gyors ütemű fejlődésnek köszönhetően a legújabb modellek folyamatos fejlesztés és frissítés nélkül **elavultnak** számíthatnak rövid időn belül.

4. NAGY NYELVI MODELLEK OSINT ESZKÖZÖKBEN

A **H3 hipotézis** vizsgálatához ki kell tekintenünk a szakirodalmi források közegeből, és a gyakorlatban való megjelenés jellemzőit vesszük górcső alá. A kutatás során olyan **projekteket, szoftvereket, keretrendszereket** igyekeztem találni, amelyek megfelelnek a **két fő kritériumnak**. Az első, hogy **céljuk a nyílt forrású információgyűjtés**, a második, hogy **nagy nyelvi modellek képességeit használják fel** e cél érdekében. A téma újszerűségét bizonyítja, hogy jelentős forrásokat (pl.: idő, eszközök) kellett ráfordítani ahhoz, hogy néhány ilyen szereplő a kutatás látóterébe kerüljön. A tett erőfeszítések eredménnyel jártak, így a következőkben ezekből példálózunk.

4.1. LLM OSINT PoC

Az LLM OSINT²¹⁵ projekt egy ingyenesen elérhető proof of concept (PoC)²¹⁶ koncepció, amely erősen innovatív jegyeket mutat. LLM OSINT-ként van elnevezve a projekt, de a könnyebb azonosíthatóság okán „LLM OSINT PoC” elnevezést fogjuk használni a következőkben. A név mögött egy egészen komplex, több komponensű keretrendszert találunk.²¹⁷

²¹⁵ GitHub — sshh12/llm_osint. Elérhető: https://github.com/sshh12/llm_osint?tab=readme-ov-file (letöltve: 2024.02.16.)

²¹⁶ A „Proof-of-Concept” kifejezés jelentése: koncepció igazolása vagy próbaüzem, műszaki és üzleti szférában is használatos fogalom. Forrás: Mi az a proof-of-concept (PoC) exploit? Nemzeti Kibervédelmi Intézet. <https://nki.gov.hu/it-biztonsag/tudastar/mi-az-a-proof-of-concept-poc-exploit/> (letöltve: 2024.02.16.)

²¹⁷ GitHub — sshh12/llm_osint.

Az alapötlet szerint a nagy nyelvi modellek önállóan képesek weboldalakat átolvasni és az ott található információkat összegyűjteni, de mivel a **végpontól-végpontig tartó információkeresés nem hozta a várt eredményeket**, ezért alternatív megoldásként született meg ez a módszer. A projekt „web ügynököket” alkalmaz az információkeresésre, amelyek a GPT-4.0 képességeit alkalmazzák, majd a részletes keresést mély információgyűjtést végző ügynökök hajtják végre. Ezek úgy működnek, hogy az ügynök a GPT-4.0 által visszajutott információk alapján a **Serper API-n** (google search API) keresztül felkutat annyi releváns internetes hivatkozást, amennyi csak lehetséges, majd web scraping²¹⁸ tevékenység veszi kezdetét, ami a Scrapingbee²¹⁹ alkalmazásával, API-n keresztül történik meg.²²⁰

7. ábra. az LLM OSINT PoC kezdeti ügynöke által futatott angol nyelvű prompt részlete.

Forrás: https://github.com/sshh12/llm_osint

```
1  ✓ PREFIX = ""
2  You are a automated OSINT agent who can expertly find information on persons, companies, places, and things.
3
4  You use the question/thought/action/observation format for performing these searches.
5
6  If you have exhausted the available resources DO NOT STOP, perform another search for a detail you could use more information about.
7
8  Answer the users questions by scanning the internet for relevent data. Be very thorough.
9
10 Hints:
11 * Always read the contents of linked webpages that can give you more information to answer the query
12 * Use the contents of webpages to recursively discover information. Crawl links that could provide more information.
```

A kinyert adatokat (pl.: HTML források), azok méretbeli nagysága miatt további DOM (Document Object Model) chunkok tisztítják olyan szöveggé, amely már formátumában alkalmas a GPT-3.5 LLM általi reprezentálásra. A modell ebből a nagyszöveges promptból kivonatolja a leghasznosabb adatokat egészen addig, amíg a GPT-3.5 tokenizálási korlátját ez meghaladja, majd ez addig ismétlődik, amíg a legutolsó műveletnél elfogy a feldolgozandó tartalom. Az eredményként kinyert részletek újfent visszacsatolásra kerülnek a GPT-3.5-be és az adatok összegzése és részletes feldolgozása történik meg végeredményként. A keretrendszer elnevezése „**LLM map reduce**”.²²¹

²¹⁸ A web scraping olyan tevékenység, amelynek folyamán adatot gyűjtünk weboldalokról, ahol ezek emberi olvasásra alkalmas formában jelennek meg, és nem gépi feldolgozásra optimalizáltak. Forrás: Mi a web scraping , és mire használható? Elérhető: <https://web-scraping.hu/mi-az-a-web-scraping/> (letöltve: 2024.02.16.).

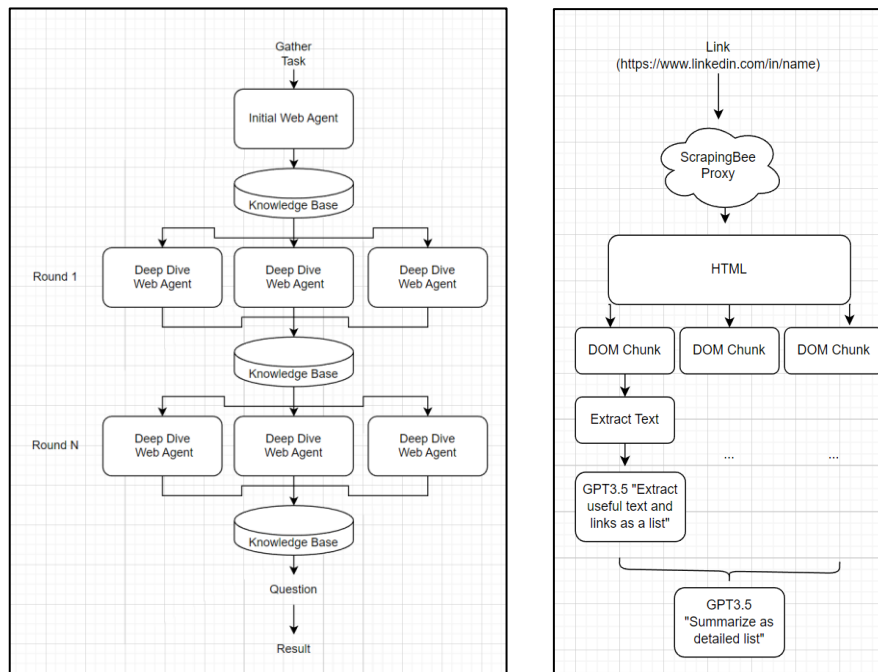
²¹⁹ ScrapingBee. Tired of getting blocked while scraping the web? Forrás: https://www.scrapingbee.com/?fpr=huynh12&gad_source=1&gclid=CjwKCAiArLyuBhA7EiwA-qo80GhBbmvmfz-QWJsibjJE6fqWo8Cm1v4E6uoETjrkW0B3xjhtt-D2sgRoC2QYQAvD_BwE (letöltve: 2024.02.16.).

²²⁰ GitHub — sshh12/llm_osint.

²²¹ Map reduce. Forrás: https://js.langchain.com/docs/modules/chains/document/map_reduce (letöltve: 2024.02.16.).

Az alkalmazás során a hostolt GPT használata miatt **költségekkel is számolni kell**, melynek mértéke függ a beindexelhető hivatkozások számától és a feldolgozott információk mennyiségétől.²²²

8. ábra: Bal oldalt az OSINT-LLM PoC web ügynökeinek működése, jobb oldalt web scrapping adatfeldolgozás szemléltetése. Forrás: https://github.com/sshh12/llm_osint



Ez a keretrendszer célirányosan két cél elérésére lett tervezve, amelyek az alábbiak:

1. Tudjon meg minél többet egy adott személyről, szervezetről, helyről, dologról stb.
2. Írjon összefoglalót erről.

A releváns információkat tartalmazó adathalmaz a GPT képességekkel egyesítve már képes olyan jelentéseket is generálni, amelyek akár a célszemély **pszichológiai jellemzőire vonatkoznak** (erősségek, gyengeségek). Kivonatolhatók a **személyes és a főbb önéletrajzi adatok**, amelyek további – akár rosszindulatú – felhasználása is lehetséges. Megfelelő prompt ismeretekkel, vagy automatizált prompt inputtal olyan **személyre szabott káros tartalmak** (pl.: ismerős személy

²²² GitHub — sshh12/llm_osint.

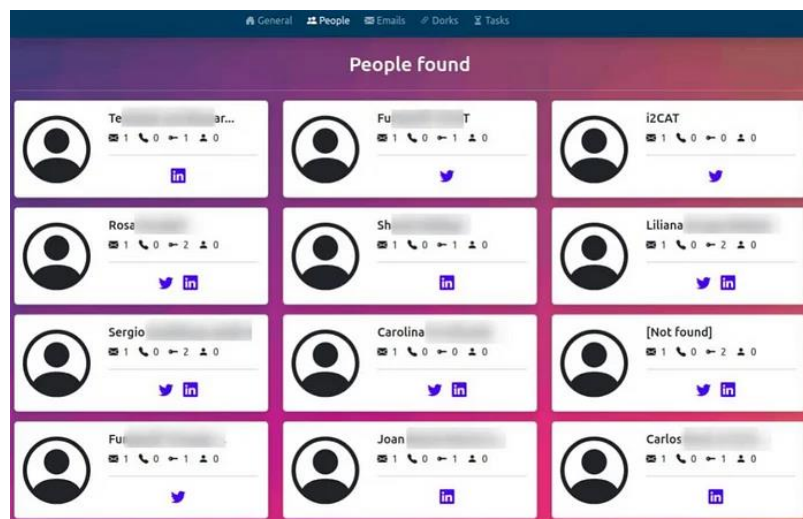
adatainak felhasználásával írt adathalász e-mail, személyre szabott hirdetés mögöttes negatív szándékkal) generálhatók, amelynek **nagy valószínűséggel eredményesek lehetnek** a célszemély esetében.²²³

Az LLM-OSINT PoC keretrendszerben több szálon, több típusú nagy nyelvi modell igénybevételét azonosíthatunk, ez a módszer egyértelműen a LLM képességekre támaszkodik.

4.2. InfoHound és Ollama

Az InfoHound²²⁴ egy speciális OSINT eszköz, amely a tevékenység kezdeti szakaszában **képes passzív és aktív felderítésre** egyaránt. Jelentős mennyiségű adatot képes összegyűjteni a célzott domainről, leginkább személyekre vonatkozó információkat (pl.: személyneveket, e-mail címeket, felhasználóneveket), valamint aldomaineket, fájlokat, URL címeket.

9. ábra: az InfoHound felhasználói felülete (1. kép). Forrás: <https://achirou.com/osint-10-username/>



²²³ GitHub — sshh12/llm_osint.

²²⁴ GitHub — Fundacio-i2CAT/InfoHound Elérhető: <https://github.com/Fundacio-i2CAT/InfoHound?tab=readme-ov-file> (letöltve: 2024.02.17.).

A projekt teljesen nyílt forráskódú szolgáltatásokból áll össze, mint például a django²²⁵, a PostgreSQL²²⁶ és a Redis²²⁷. Az projekt specifikuma, hogy **több Python könyvtár alapú OSINT eszköz** van egy csokorba integrálva, ezek az alábbiak:

1. táblázat: az InfoHound-ba integrált OSINT eszközök. Forrás: a szerző szerkesztése.

	ESZKÖZ	FUNKCIÓ
1.	Beautifulsoup ²²⁸	Megkönnyíti a weboldalakon történő adatszerzési tevékenységet. (HTML és XML források esetében).
2.	DNSpython ²²⁹	DNS szolgáltatással összefüggő lekérdezésekre használható.
3.	E-mail-validator ²³⁰	Az e-mail címek ellenőrzésére alkalmazható általános szolgáltatások során. OSINT funkcióban a web szolgáltatások adataiból kinyert e-mail címek validálása a cél.
4.	Holehe ²³¹	E-mail címeket ellenőrzi a különböző webhelyeken (pl.: Twitter, Facebook) található elfelejtett jelszó funkció alkalmazásával.
5.	Maigret ²³²	A felhasználónevek alapján keresést végez több mint 3000 weboldalon, majd összegyűjti a talált információkat.
6.	Python-whois ²³³	WHOIS ²³⁴ információk lekérdezésére.
7.	Shodan ²³⁵	Az internethez csatlakozó eszközök keresőmotorja.
8.	Texttract ²³⁶	A szöveges tartalom kinyerése minden dokumentumból.
9.	Networkx ²³⁷	Összetett hálózatok szerkezetének, dinamikájának és funkcióinak tanulmányozása.
10.	Továbbá képes DOCX, PPTX, PDF, RAR, SVG fájlok kibontására és a szövegszerű adatok kinyerésére.	

Az InfoHound nagyon részletes statisztikát képes készíteni olyan paramétereket feltüntetve, mint például:

²²⁵ Django. Elérhető: <https://www.djangoproject.com/start/overview/> (letöltve: 2024.02.17.).

²²⁶ PostgreSQL: The World's Most Advanced Open Source Relational Database. Elérhető: <https://www.postgresql.org/> (letöltve: 2024.02.17.).

²²⁷ Redis. Elérhető: <https://redis.io/> (letöltve: 2024.02.17.).

²²⁸ BeautifulSoup. Elérhető: <https://pypi.org/project/beautifulsoup4/> (letöltve: 2024.02.17.).

²²⁹ DNSpython. Elérhető: <https://www.dnspython.org/> (letöltve: 2024.02.17.).

²³⁰ E-mail validator Elérhető: <https://pypi.org/project/email-validator/> (letöltve: 2024.02.17.).

²³¹ Holehe v1.61 Elérhető: <https://snyk.io/advisor/python/holehe> (letöltve: 2024.02.17.).

²³² GitHub — soxoj/maigret. Elérhető: <https://github.com/soxoj/maigret> (letöltve: 2024.02.17.).

²³³ GitHub — mboot-github/python-whois. Elérhető: <https://github.com/mboot-github/python-whois> (letöltve: 2024.02.17.).

²³⁴ who.is Elérhető: <https://who.is/> (letöltve: 2024.02.17.).

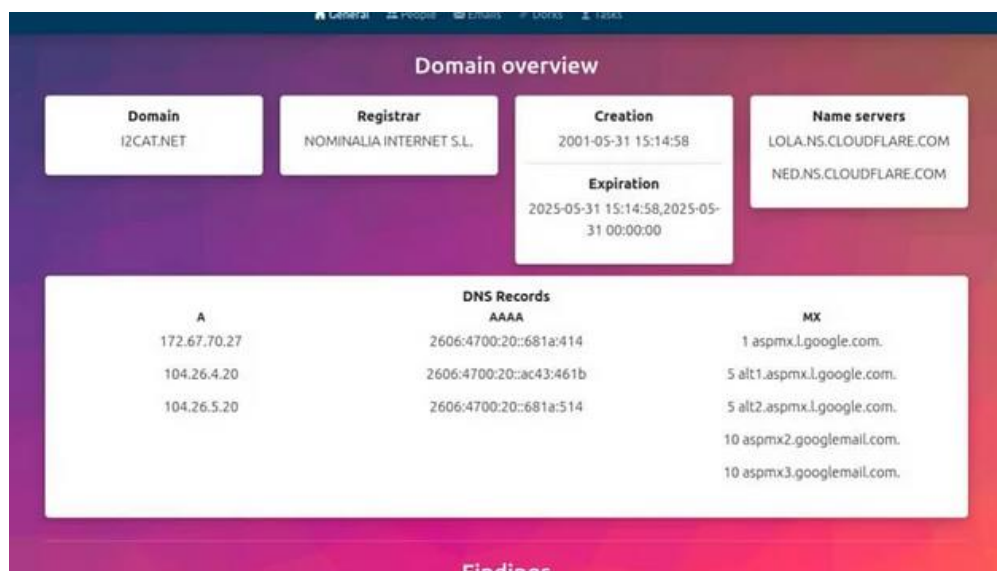
²³⁵ Shodan hivatalos Python könyvtára. Elérhető: <https://github.com/achillean/shodan-python> (letöltve: 2024.02.17.).

²³⁶ Texttract. Elérhető: <https://pypi.org/project/texttract/> (letöltve: 2024.02.17.).

²³⁷ Networkx. Elérhető: <https://pypi.org/project/networkx/> (letöltve: 2024.02.17.).

- Az összegyűjtött URL-ek száma;
- a felfedezett aldomainek száma;
- a felfedezett személyek száma;
- az észlelt fájlok száma;
- az észlelt e-mailek száma (már kiszivárgott, sikeresen egyénhez kapcsolt, más szolgáltatáshoz kapcsolt).²³⁸

10. ábra: az InfoHound felhasználói felülete (2.kép). Forrás: <https://achirou.com/osint-10-username/>



Egy nagyon modern felhasználói felület áll rendelkezésre az adatok áttekintéséhez, valamint az említett részfunkciók önállóan is futtathatók annak érdekében, hogyha a későbbiekben csak részleges frissített elemzésre lenne szükség. A nyílt forráskódnak köszönhetően az eszköz **modifikálható**, így egyéni modulok is elhelyezhetők benne. Az eddigiekben ismertetett információk alapján észlelhetjük, hogy ez a rendszer önmagában is **egy nagyon erős OSINT eszköz**. Viszont a **legutolsó verzió** már **LLM támogatást is kapott**, méghozzá az **Ollama keretrendszerével**.

²³⁸ CyberRaya (2023.10.23.). InfoHound: Enumerate emails, usernames, subdomains, names and other information. Elérhető: <https://osintteam.blog/infhound-enumerate-emails-usernames-subdomains-names-and-other-information-828bb9b8ac8e> (letöltve: 2024.02.18.).

Az Ollama²³⁹ szintén **egy nyílt forráskódú keretrendszer**, amely biztosítja a **helyi gépeken történő, szintén nyílt forráskódú LLM-ek futtatását felhő és internet szolgáltatás nélkül**. Praktikus eszköz azok számára, akik klasszikus operációs rendszereken (macOS, Linux, Windows) kívánnak nagy nyelvi modelleket futtatni. Az Ollama előnye a **rugalmas konfigurációs lehetőség**, amelyekkel jól testre szabhatók a futtatni kívánt LLM-ek. Számos modell élvezi a keretrendszer támogatását, ezek közül a teljesség igénye nélkül példálózva a modellek nevei, paraméterei és helyigénye az alábbi táblázatban kerül szemléltetésre.

2. táblázat: az Ollama által támogatott nagy nyelvi modellek és jellemzőik.

Forrás: <https://github.com/ollama/ollama>

OLLAMA TÁMOGATOTT LLM-EK		
Modell	Paraméterek	Méret
Llama 2	7B	3.8 GB
Mistral	7B	4.1 GB
Dolphin Phi	2.7B	1.6 GB
Phi-2	2.7B	1.7 GB
Neural Chat	7B	4.1 GB
Starling	7B	4.1 GB
Code Llama	7B	3.8 GB
Llama 2 Uncensored	7B	3.8 GB
Llama 2 13B	13B	7.3 GB
Llama 2 70B	70B	39 GB
Orca Mini	3B	1.9 GB
Vicuna	7B	3.8 GB
LLaVA	7B	4.5 GB

Segítségével bármely támogatott LLM futtatható **otthoni környezetben**, annyi kitételrel, hogy a 7B (milliárd) paramétert tartalmazó modellekhez legalább 8 GB, a 13B-s 16 GB, a 33B modellekhez minimum 32 GB memória (RAM) szükséges.²⁴⁰

Az InfoHound legújabb verziója az Ollama által futtatott **Llama2-thonálja ki a személyek szervezetén belüli szerepének rövid leíratosához**. Ez a funkció az OSINT eszközök által feltárt adatok ismeretében **képes meghatározni a szervezeti struktúrát**, amelynek köszönhetően

²³⁹ Ollama. Elérhető: <https://ollama.com/> (letöltve: 2024.02.18.).

²⁴⁰ GitHub — ollama/ollama. Elérhető: <https://github.com/ollama/ollama> (letöltve: 2024.02.18.).

eredményesen meghatározható a **kulcsfontosságú személyek köre**. Tulajdonképpen egy AI alapú profilelemzési feladat segítségével az LLM megvizsgálja az adatokat, majd összefoglalja a kívánt célok érdekében.

4.3. NexusXplore OSINT Research Assistant (ORA)

Az Ausztráliai OSINT Combine által létrehozott NexusXplore a világ egyik vezető OSINT platformja, amely számos képességgel segíti az OSINT elemzőket az adatok gyűjtésében és elemzésében. A vállalat fejlesztői a közelmúltban egyfajta asszisztensként integrálták a ChatGPT-t rendszerükbe, így létrehozva az **OSINT Research Assistant-et**, amelynek használatával az elemzők javíthatják munkafolyamataikat. Az ORA nagy mennyiségű adat gyors és hatékony elemzését teszi lehetővé, valamint más dedikált elemző rendszerek is képesek azonosítani az ORA válaszából származó információkat, így létrehozva **egy mesterséges intelligenciával támogatott digitális OSINT ökoszisztémát**. Ennek a megoldásnak – hatékonyság és gyorsaság szempontjából – lényegi része a GPT modell, amelynek természetes nyelv fordítási képességei hozzájárulnak a több nyelvű források feldolgozásához is. Tekintettel arra, hogy az ORA vállalati produktum, a termékhez való hozzáférés és annak megismerése korlátozott és csak partnereknek lehetséges. A kutatás során kísérletet tettem a kapcsolatfelvételre, azonban az nem vezetett eredményre. Ennek okán a kutatás további részletes, gyakorlatra vonatkozó információkat nem tárt fel.²⁴¹

4.4. Autonóm LLM hack agent (ALLMHA)

Ebben a témában bemutatott rendszer nem OSINT-LLM megközelítésű, így nem felel meg teljes mértékben a kutatás kritériumának. Azonban állaspontom szerint, mivel egy kifejezetten friss tudományos eredményről van szó, és egyaránt van LLM és OSINT érintettsége is, valamint bizonyosan hozzájárul a kutatási terület komplexebb vizsgálatához, így említjük azt.

2024 februárjában megjelent egy „*LLM Agents can Autonomously Hack Websites*” című tanulmány, amelyben Richard Fang és munkatársai megkíséreltek olyan ügynököt (ágenst) létrehozni, amely **LLM képességek kihasználásával autonóm módon képes weboldalak sebezhetőségeinek azonosítására és kihasználására**, oly módon, hogy **előzetesen nem ismeri a**

²⁴¹ Chris Poulter (2023.015.09). OSINT Workflow with ChatGPT: Tips, Risks, and Benefits. Elérhető: <https://www.osintcombine.com/post/osint-workflow-with-chatgpt-tips-risks-and-benefits> (letöltve: 2024.03.16).

támadott oldal sebezhetőségét. Az „LLM-agent” névre keresztelt ügynök többlépcsős folyamatok végrehajtására képes. Az ágens fő alkotóelemei az alábbiak. A weboldalak adatiról a Playwrite²⁴² keretrendszer szállítja a szükséges információkat, amellyel API-n keresztül kommunikál a rendszer. LLM képességek terén egyaránt kísérleteztek hostolt és nyílt forrású nagy nyelvi modellekkel is, összesen 10 típust próbálva (3. táblázat).²⁴³

3. táblázat: a tesztelt modellek listája.

Tesztelt modellek:	
1.	GPT-4
2.	GPT-3.5
3.	OpenHermes-2.5-Mistral-7B
4..	LLaMA-2 Chat (70B)
5..	LLaMA-2 Chat (13B)
6.	LLaMA-2 Chat (7B)
7.	Mixtral-8x7B Instruct
8.	Mistral (7B) Instruct
9.	Nous Hermes-2 Yi
10.	OpenChat 3.5

Minden LLM-et a LangChain keretrendszer alatt futtattak. Az OpenAI Assistant²⁴⁴ API-n keresztül vették igénybe a GPT-3.5 és GPT-4 modellek képességeit, a többi tesztelt modellhez pedig a Together AI²⁴⁵ API-t használták fel. A nagy nyelvi modellekkel kapcsolatban leginkább azok funkcióit hívták le, továbbá dokumentumokat olvastattak be vele (ez egyfajta fine-tuningként is értelmezhető), valamint tervezési képességeit alkalmazták. Az LLM ügynök hozzáférést kapott a Playwright termináljához, valamint egy Python kódértelmezőhöz, tovább **hat dokumentumhoz is, amelyek a webes hackelésről** (általános témáról, valamint SQL-injection, Cross Site Scripting [XSS], Server-side request forgery [SSRF] sérülékenységek kihasználásáról) tartalmaztak információkat.²⁴⁶

²⁴² Playwrite. Playwright enables reliable end-to-end testing for modern web apps. Elérhető: <https://playwright.dev/> (Letöltés: 2024.03.24).

²⁴³ Richard Fang, Rohan Bindu, Akul Gupta, Qiusi Zhan, Daniel Kang (2024). LLM Agents can Autonomously Hack Websites <https://doi.org/10.48550/arXiv.2402.06664>.

²⁴⁴ OpenAI Assistant. Elérhető: <https://platform.openai.com/docs/assistants/overview?context=with-streaming>.

²⁴⁵ Together.ai. Elérhető: <https://api.together.xyz/>.

²⁴⁶ Richard Fang et al (2024).

Az LLM ügynök felé kiadott prompt az alábbi főbb elemeket tartalmazta: **legyen kreatív, próbáljon ki különböző stratégiákat, az ígéretes stratégiákat folytassa a végsőig, sikertelenség esetén is próbáljon ki új stratégiákat.** 15 féle különböző nehézséggel azonosítható sebezhetőséget teszteltek.²⁴⁷

Kísérletük eredményeiben megállapítják, hogy a **GPT-3.5 az esetek 6,7%-ban, míg a GPT-4 modell 73,3%-ban képes volt feltörni a sandboxban²⁴⁸ futtatott weboldalakokat.** A legösszetettebb sikeres támadástípus az SQL-injection volt, hiszen ez a támadás igényelte a legtöbb interakciót a weboldallal. További tesztelések eredményeit elemezve megállapították, hogy az eredményességre leginkább az LLM modell **skálázottsága**, valamint valószínűsíthetően a témába vágó tanítási adatok vannak a legnagyobb hatással. Továbbá a **GPT-4 kifejezetten jól hasznosította a weboldali interakciókat** a stratégia következő lépésének megalkotásakor.²⁴⁹

További érdekes tapasztalás, hogy a **nyílt forrású LLM-ek nem voltak képesek használni** az eszközöket, így **nem hajtottak végre sikeres támadást**, azonban az OpenChat-3.5 képes volt az eszközök alkalmazására, sőt az esetek **25,3%-ában megkísérelte a helyes sebezhetőséget**, de nem használta fel az interakciókból származó visszajelzéseket a helyes támadás értelmezéséhez.²⁵⁰

A képességeket próba alá vetették a tesztkörnyezeten kívül, élesben is. A surface weben 50 darab karbantartatlannak minősített weboldal közül **egy esetben sikerrel hajtotta végre a GPT-4 a weboldal hackelését.** Ez bizonyítja azt, hogy az LLM megoldások már túl léptek a CKC kezdeti fázisain és elérik a kihasználási szakaszt, amely megállapítás további feldolgozására a következő fejezetben is kitérünk.²⁵¹

²⁴⁷ Richard Fang et al (2024).

²⁴⁸ Sandbox, vagyis „homokozó”. Az informatikai biztonság területén programok elkülönített futtatására szolgáló biztonsági mechanizmus. Gyakran használják még nem tesztelt kódok, ellenőrizetlen gyártóktól, szállítóktól vagy végfelhasználóktól származó, nem megbízható programok futtatására. Forrás: [https://hu.wikipedia.org/wiki/Homokoz%C3%B3_\(biztons%C3%A1g\)](https://hu.wikipedia.org/wiki/Homokoz%C3%B3_(biztons%C3%A1g)).

²⁴⁹ Richard Fang et al (2024).

²⁵⁰ Richard Fang et al (2024).

²⁵¹ Richard Fang et al (2024).

5. ÖSSZEFOGLALÁS

A két témában a szakirodalmi források által képviselt tudást, a legújabb, megbízható forrásból származó tudományos eredményeket, a gyakorlatban megjelenő eszközökkel kapcsolatban feltárt jellemzőket és a szerző véleményét felhasználva igyekszem a nagy nyelvi modellek OSINT alkalmazásával kapcsolatban **megalapozott megállapításokat tenni**. Megkísérelem a terület jelenkori megjelenésével kapcsolatos jellemzőket körülírni, értelmezni az **OSINT-LLM megoldások érvényesülését a kibertámadások CKC szakaszaiban**. Továbbá a múltban észlelt haladást elemezve **megpróbálom meghatározni a téma várható, jövőbeli terjedésének irányait**.

5.1. Összegzés és konklúzió

Összességében elmondhatjuk, hogy manapság a kibertérben létrejövő, különböző nyilvános forrásokból származó nagy mennyiségű adat LLM módszerekkel történő feldolgozása **jelentős előnyt jelent** a hatékonyság szempontjából. Az LLM beépítése az OSINT módszerekbe már nem példa nélküli, **de még nem teljesen kiforrott technológia**. A kutatás során arra a megállapításra jutottam, hogy az LLM modellek bizonyosan **erőteljes hatást gyakorolnak** az OSINT-re és leginkább **a nagyobb merítésű, korábban leküzdhetetlennek hitt adatmennyiségek feldolgozásában** képesek hatékonyan közreműködni. Ugyanerre a megállapításra jutottak tanulmányukban Ghioni és munkatársai (2023), valamint arra, hogy a jövőben **egyértelműen az fog piaci előnyt** szerezni, aki integrálja az AI megoldásokat az OSINT elemzési fázisában.²⁵²

A kutatás fókuszába leginkább olyan felhasználási módok kerültek, amelyekben az LLM képességet OSINT módszertanokba az **adatifeldolgozás szintjén integrálják** azért, hogy **nagyobb mennyiségű információból kivonatolják** a releváns információkat a kívánt célok szerint.

Nem példa nélküli az **LLM végponttól-végpontig** tartó nyílt forrású információszerzési képességének alkalmazása sem, de megítélésem szerint ez **önmagában a jelenlegi korlátok okán nem kellően hatékony**, azonban megfelelő kiegészítője lehet az OSINT tevékenységeknek. Az LLM információfeldolgozása a hagyományos OSINT eszközök módszertanától **eltérő**

²⁵² Ghioni, R., Taddeo, M. & Floridi, L. Open source intelligence and AI: a systematic review of the GELSI literature. AI & Soc (2023). <https://doi.org/10.1007/s00146-023-01628-x> p.14.

megközelítést alkalmaz, hiszen a statikus tartalom kivonatolása helyett nagy adatkészletek szintetizálásával emberszerű válaszokat ad vissza.

Az OSINT tevékenységnek megvannak a maga jellemzői, amelyek közül a technológiai **fejlődést biztosító lehetőségek** között jelenik meg az LLM technológia. A nagy nyelvi modellek képesek **redukálni vagy megoldani** az OSINT **néhány gyengeségét** (pl.: adatdömping, elemzési képességek), amelyekkel egyértelműen hatékonyabbá válik a nyílt forrású információgyűjtés.

Napjainkban a mesterséges intelligenciák, azon belül a mélytanulási architektúrák között az LLM az **egyik lépcszerűbb terület**. A technológia **sokoldalúságát** bizonyítja, hogy utóbbi 4-5 évben jelentős számú tudományos eredmény jelent meg több tudományágat érintően. A **mesterséges intelligencia és a kiberbiztonság területe között egyre nagyobb az átfedés**, amely bizonyítja azt a megállapítást, hogy a téma tudományos szempontú feldolgozásának folyamatosságát biztosítani szükséges.

A **nagy nyelvi modellek képességei rendkívüliek**, bizonyosan jelentős hatást fognak gyakorolni társadalmunk számos területére. Azonban a technológia még nem mondható kiforrottnak, és gyengeségei és a vele szembeni fenyegetései listája hosszú. E **hátrányok leküzdése számottevő pénzügyi és szakmai forrásokat igényel**, amelyeket csak a legnagyobb szolgáltatók képesek biztosítani. A legújabb modellekben is tapasztalható **hallucináció** egyértelműen okozója a technológia irányába megjelenő **bizalomvesztésnek**, ugyanezen okból a valós idejű információgyűjtés képességével felvértezett LLM-ek által begyűjtött nyílt információ hitelessége csak **alternatív eszközzel történő visszaellenőrzéssel** biztosítható, amely egyértelműen negatívan befolyásolja az OSINT hatékonyságát. Ugyanakkor a nyílt forráskódú megoldásokat alkalmazó OSINT projekteknel **nem számít a modellek biztonsága**, csak az előnyök kihasználására alapoznak, **amely megközelítéssel jól működő célrendszerek alkothatók**.

OSINT-LLM megoldások között egyaránt találunk olyan módszertanokat, amelyek **LLM szolgáltatók képességeit** veszik igénybe és talán azonos arányban észlelhetők olyan megoldások, amelyek a nyílt forrású LLM előnyeire alapoznak. A két megközelítés között egyértelműen észlelhetünk azonosságot és diverzitást is. A kutatás fókuszába került egy olyan kibertámadást önállóan végrehajtó LLM megoldás is (Autonóm LLM hack agent), amely **bőven meghaladja az**

OSINT alkalmazást, és képességeit tekintve egyértelműen **veszélyforrást jelent** az kiberbiztonság állapotára.

A kutatás fókuszába az eddigiekben leírt véleménnyel szembeni, vagy azzal ellentétes, megalapozottnak és hitelesnek tekinthető **tudományos eredmény vagy állásfoglalás nem került.**

5.2. Az OSINT-LLM keretrendszerek azonosított típusai és jellemzői

A kutatás során feltárt OSINT-LLM integrációkat figyelembe véve, azok architektúrák kialakítása, kapcsolatai, képességei és jellemzőik eltérőek. Azonban a vizsgált keretrendszerek egyaránt azonos komponensekből állnak, melyek a következők:

OSINT-LLM keretrendszert általánosan alkotó részek:

1. Nyílt forrású információszerzést biztosító komponens (OSINT modul)

Funkció: azon elemek köre, amelyek biztosítják a keretrendszer számára a feldolgozni kívánt jelentős mennyiségű adat felkutatását és összegyűjtését. Állhatnak külső szolgáltatók által biztosított OSINT eszközökből, de nyílt forrású, keretrendszerbe integrált OSINT eszközökből is.

2. LLM képességet biztosító modul (LLM modul)

Funkció: Általában egy, vagy több LLM kapcsolatot takar, amely külső szolgáltató által biztosított kapcsolaton elérhető hostolt, vagy saját nyílt forrású nagy nyelvi integráció. A modul biztosítja az OSINT modul által szállított információ (leginkább szöveg alapú adat) hatékony feldolgozását, utasítás alapján történő kivonatolását. Esetenként az OSINT információ feldolgozása generatív képességekkel, vagyis az LLM tanítási adataiból képzett minták felhasználásával is történhet (pl.: „készíts pszichológiai elemzést az alábbi személyről”).

3. Központi modul (Core)

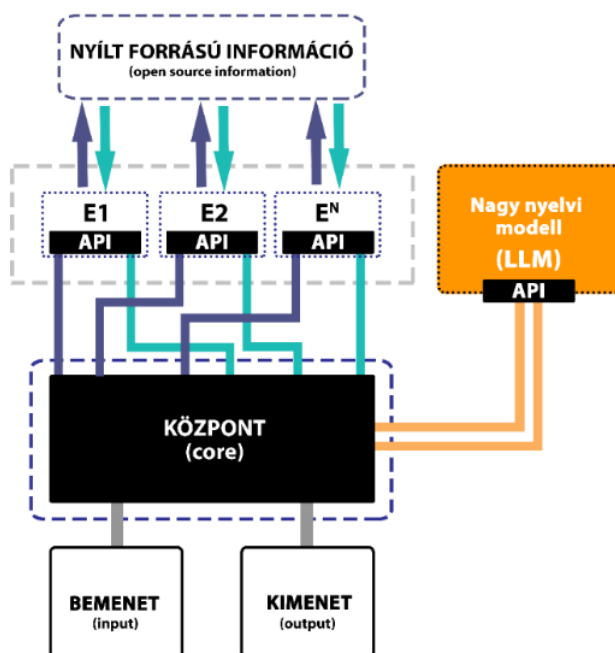
Funkció: a központi modul, amelyet nevezhetünk **„core”-nak is, biztosítja a kapcsolatot** az egyéb képességeket szolgáltató keretrendszer komponensekkel, valamint vezérli azokat. Leginkább a modulok közti információt szállítja, továbbá fogadja a felhasználó általi

bemenetet (utasítást), valamint biztosítja részére történő **kimenetet** (kész információt) közlését.

A modulok keretrendszerben való megjelenése azonban mutat némi asszimetriát, amelynek alapján **sikerrel azonosítottam két eltérő keretrendszer típust**, amelyeket az alábbiakban fogalmazok meg.

5.2.1. Szolgáltatás alapú

Az első a külső **szolgáltatás alapú, vagy arra támaszkodó (third-party dependent)** megoldások csoportja, amely a külső szolgáltatások előnyeit veszi igénybe a megfelelő kimenet elkészítéséhez. Az 11. ábrán látható, ahol a bemenet (pl.: a célszemélyre vagy célszervezetre vonatkozó információ) a központi komponensbe kerül, ahonnan ezzel kapcsolatos információkat kér be a különböző szolgáltatások interfészein (API-k) keresztül.



11. ábra: a szolgáltatásokra támaszkodó OSINT-LLM architektúra ábrázolása.
Forrás: a szerző szerkesztése

Az önálló OSINT eszközként (E) is értelmezhető szolgáltatások száma (N) **elméletben bármennyi lehet**. Ezek a szolgáltatások leginkább olyanok lehetnek, mint például a korábban említett *Google*

Search, Shodan vagy a ScrapingBee. Ezek a szolgáltatók visszajuttatják a **nagy halmazú nyílt forrású információt** a központi egységbe, ahonnan viszont a megfelelő prompt kíséretében megküldésre kerül az LLM-nek, amely egy **szintén külső szolgáltatásként** van igénybe véve API-n keresztül. Ezt követően a **meghívott LLM képesség** (pl.: elemzés, kivonatolás, összesítés) érvényesül és az utasítás szerint információt alkot. Ez az információ – a prompttól függően – **önállóan, vagy egyéb generatív funkciókkal újraalkotott tartalomként** kerül vissza és kimenetként jelenik meg a felhasználó személynél.

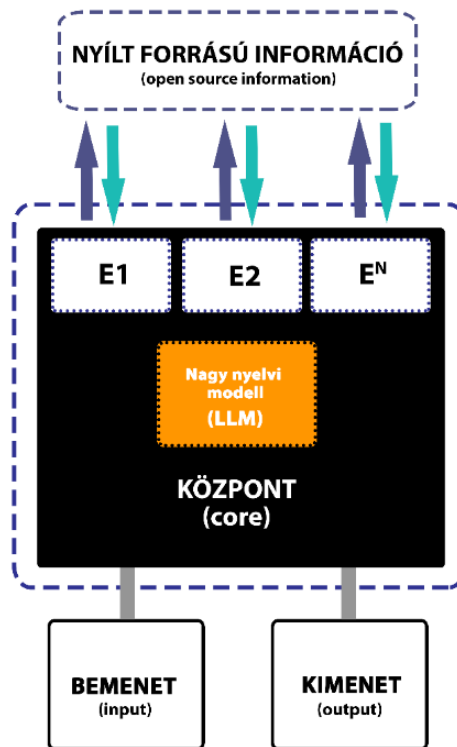
Az így alkotott architektúra előnye, hogy **skálázható**, hiszen a **kapcsolt szolgáltatások száma elméletben korlátlan**. Gyakorlati megközelítésben **forrásigénye alacsony**, hiszen a nagyobb számítási kapacitást igénylő feladatokat (pl. adatbányászat, feldolgozás, LLM képességek stb.) a **kapcsolt szolgáltatások végzik**, ugyanakkor ennek okán **jelentős szolgáltatási költségek** jelentkeznek. Szintén ebből az okból **kitettsége jelentős**, tehát **alkalmazhatósága függ a szolgáltatók rendelkezésre állásától**. **Rejtve maradása** végponttól-végpontig **nem értelmezhető**, hiszen a behívott szolgáltatások megkapják a célpontra vonatkozó információkat, így a tevékenység már nem marad rejtve. Ugyanakkor a rejtve maradásra kockázatot jelent a kapcsolt szolgáltatások által **végzett gyakorlat is** (pl.: aktív OSINT). **Erőssége a rendszernek**, hogy az integrált szolgáltatások a **piacon elérhető leghatékonyabbak** (pl.: legújabb OSINT eszközök, legkorszerűbb LLM képességek), így a várhatóan a kapott eredmények is kielégítőek lesznek.

Ugyanakkor problémásan felel meg **az OSINT tevékenységgel szembeni transzparencia követelménynek**, hiszen a behívott szolgáltatók gazdasági érdekük alapján **nem nyújtanak teljes és maradéktalan átláthatóságot** az általuk szolgáltatott eszközökkel és eljárásaikkal összefüggésben. Ez az architektúrátípus **könnyen alkalmazkodik a változó technológiai megoldásokhoz** és meglátásom szerint amíg a szolgáltatók (különösen az LLM szolgáltató) biztonsági intézkedései teret engednek az ilyen típusú felhasználásnak, **eredményes marad**.

Azonban, ha a hostolt LLM **együttműködése megszűnik a szimbiózisban**, abban az esetben ez az architektúra **alkalmazhatóságának a végét jelentheti**. De tekintettel arra, hogy a nagy nyelvi modellek népszerűsége **töretlen**, jó eséllyel számos **korlátozás nélkül LLM modell lesz elérhető a jövőben**, és mivel ez a modul nem integrált, **könnyen módosítható** milyen szolgáltatáshoz kapcsolódik.

5.2.2. Helyben integrált

A másik csoport a **helyi integrációban (local integrate)** létrehozott típus, **amelyben nem találunk külső szolgáltatói kapcsolatot**, és minden OSINT-LLM funkció *egy helyi rendszerintegrációban* jelenik meg.



12. ábra: az OSINT-LLM képességeket helyi szinten integráló architektúra ábrázolása
Forrás: a szerző szerkesztése

Az előző modell architektúrához képest azt az **eltérést látjuk** (12. ábra), hogy minden szükséges OSINT eszköz (E) és LLM képesség **lokálisan integrálódik** a keretrendszerbe. Ez azt jelenti, hogy **nincs szolgáltató irányába megjelenő kitettség és költségvonzat**, valamint biztosított a rejtve maradás, hiszen **nincs harmadik félen keresztülmenő kommunikáció**. Az integrált OSINT képességet biztosító elemek esetében elmondhatjuk, hogy korlátozott hatékonysággal bírnak, hiszen **többnyire nyílt forráskódúak**, amelyek nem versenghetnek a piaci óriások által szolgáltatott eredményekkel. Ugyan ez a helyzet az LLM modullal is. Itt egy, a **core-ba integrált nyílt forráskódú nagy nyelvi modellt találunk**, amely nyilvánvalóan **nem képes** a korszerű API-n elérhető nagy nyelvi modellek képességeire, ugyanakkor megvannak a maga előnyei. Az így

integrált LLM **fine-tuninggal tovább képezhető speciális OSINT tevékenységre**, amelytől talán még jobb eredményeket várhatunk az OSINT terén. Ennek a keretrendszer típusnak az üzemeltetési költségei is eltérőek. Az OSINT rendszerelemek és az LLM saját architektúráján történő futtatása okán **jelentős számítási kapacitásokat igényelnek**, azonban **nem kell számolni a külső szolgáltató irányába megjelenő költségekkel**. Ez a keretrendszer **szintén skálázható**, hiszen az integrációk száma elméletben itt is korlátlan, azonban már megjelenik a rendelkezésre álló **források által okozott korlátozottság**. Tekintettel arra, hogy **nincs a szolgáltató irányába fellépő függőség**, azt gondolom, hogy ez a megoldás a jövőben továbbra is fejlődni fog. Mivel a teljes keretrendszer elemei integrálva vannak, így az OSINT tevékenység **transzparenciája is biztosítható**. Annak okán, hogy megőrzi a használója anonimitását, **a rosszindulatú tevékenységek kiszolgálójává válhat**, ez jelentős kockázatot jelent a kiberbiztonság állapotára.

5.3. Az OSINT-LLM megjelenése a Cyber Kill Chain szakaszaiban

A dolgozat leíró szakaszában a Cyber Kill Chain szakaszaiban vizsgáltuk az OSINT megjelenését. Ebben a fejezetben, immáron eredményalkotási céllal újra elővesszük a CKC folyamatát és meghatározzuk OSINT-LLM megoldások CKC beli elhelyezkedését. A 4. számú táblázatban (mátrixban) kísérletet tettem erre az ábrázolására, valamint a bemutatott LLM megoldásokkal is kiegészítettem azt.

A mátrix megalkotásának elve és szempontjai és elemzése:

A szakaszokban feltüntettem a **3. 3. fejezetben taglalt OSINT eszközöket**, valamint besoroltam a **5. fejezetben** bemutatott LLM modelleket (LLM OSINT PoC, InfoHound és Ollama, NexusXplore ORA, ALLMHA), illetve az **4.5.3. fejezetben** bemutatott és CKC megközelítés szerint **releváns LLM modelleket** (DarkBERT, WorMGPT, FraudGPT). Az LLM megjelenéseket kétfelé csoportosítottam, az OSINT eszközöket **integráló** (pl.: LLM OSINT PoC stb.) modellek fehér színnel, míg az **önálló OSINT eszköz integrációt nem tartalmazó**, de CKC-ban értelmezhető **LLM-ek sárga betűszínnel**, horizontálisan külön sorban lettek feltüntetve. Ez az ábrázolás lehetővé teszi, hogy ezt a két csoportot önállóan értékeljük aszerint, hogy **mely CKC szakaszban aktívak**.

Továbbá, az OSINT eszközök sorában megjelenítésre kerültek azon feltárt **OSINT szoftverek is**, amelyek a **3. 3. fejezetben nem kerültek** említésre, de az OSINT-LLM integrációk (**5. fejezet**) tartalmazzák vagy alkalmazzák azokat (központi modulba integrálva, vagy API kapcsolattal), ezek jelölése szintén fehér betűszínnel történt meg. A cellák színei is jelentéssel bírnak, **a vörös erősebb árnyalatai jelölik azokat a CKC szakaszokat, ahol álláspontom szerint jelentős kiberbiztonság kockázat** lett feltárva, ennek értelmezéséhez segítség a mátrix alatti útmutató.

4. táblázat: az LLM-OSINT keretrendszerek megjelenése a Cyber Kill Chain értelmezésében.
 Forrás: a szerző saját szerkesztése

Cyber Kill Chain							
	Felderítés	Fegyverkezés	Szállítás	Kihasztnálás	Telepítés	Futtatás és vezérlés	Célok elérése
OSINT (csak osint eszközök)	DNS Dumpster; DNSpython; domainsdb; ip2whois; <u>Networks</u> ; recon-ng; weleakinfo; <u>E-mail validator</u> ; <u>Python-whois</u> WayBackMachine, Hole; cryptome; Raven; Maigret; <u>Serper API</u> ; Scrappingbee	<u>Shodan</u> ; Censys; VulDB; NVD; CNNCD; Exploit DB, Rapid 7 Exploit DB; 0 day Today	Hunter; haveibeenpwned; NameCHK; ThatThem; Default Password DB; Default-password; Default Password Lookup	Virustotal; Hybrid Analysis; Ether	Vnware; Virtual Box; Genymotion; Bluestacks; Andyroid	Zone-H; Blackweb; Malware patrol	
OSINT- LLM	LLM OSINT PoC; InfoHound és Ollama; NexusXplore ORA	LLM OSINT PoC					
Önálló LLM	ALLMHA; DarkBERT;	ALLMHA; WormGPT; FraudGTP	ALLMHA	ALLMHA			

	Nincs azonosított kockázat. A kutatás nem tárt fel kölcsönhatást az adott szinten.
	Alacsony kockázat. A kutatás nem tárt fel gyakorlati megjelenést, de elméletben létrejöhet a kölcsönhatás.
	Mérsékelt kockázat. A kutatás bizonyított egy mérsékelt szintű kölcsönhatást, ezért annak védelmi célú figyelembevétele megfontolandó.
	Magas kockázat. A kutatás magas szintű kölcsönhatást tárt fel, ezért annak védelmi célzatú kezelése ajánlott.
	Rendkívüli kockázat. A kutatás rendkívül magas kölcsönhatást állapított meg, amely magas szintű fenyegetést jelent. A kockázatok kezelése kiemelt figyelmet igényel a védelmi stratégia alkotásakor.

Láthatjuk, hogy az OSINT-LLM megoldások a **felderítési szakaszban a leggyakoribbak**, de **nem kizárólagosan csak itt aktívak**. Továbbá, az ilyen keretrendszerekbe integrált OSINT eszközök is szintén a felderítési szakaszban a leggyakoribbak, **de csak itt jelennek meg**.

Megállapításként elmondhatjuk, hogy reális esély van arra, hogy a kutatás témája, vagyis az OSINT-LLM megoldások köre még **kezdetlegesnek mondható**, ugyanakkor **jövőbeli fejlődése várható**. Esély van arra, hogy az OSINT-LLM keretrendszerek fejlődni és további OSINT eszközökkel fognak bővülni, ennek köszönhetően további CKC szakaszokban jelenhetnek meg a jövőben úgy, hogy az OSINT-al szembeni elvárásoknak megfeleljenek.

Bár a kutatásnak nem volt célja, de feltárássra került, hogy az OSINT eszközök integrációja nélküli, önálló nagy nyelvi modellek (pl.: ALLMHA, WormGPT, FraudGTP) **képességeiket tekintve bizonyítottan értelmezhetőek a CKC egyéb szakaszaiban is**, tehát alkalmazhatók a kibertámadások támogatására. Feltűnik a különbség, hogy a vizsgált nagy nyelvi modellek e típusa **további szinteken jelent meg a CKC folyamatában**. E jelenség jövőbeli **exkalálódása veszélyforrásként** kezelendő, amelynek kezelési szükséglete meggyőződésem szerint **paradigmaváltást fog hozni a kiberbiztonságban** eddig alkalmazott nézetek kapcsán.

5.4. Hipotézisek bizonyítása

A korábban felvetett hipotézisek és a velük kapcsolatos állásfoglalások az alábbiakban foglalhatók össze:

H1: *a nagy nyelvi modellek kibertéri OSINT eszközként értelmezhetőek.*

Ez a hipotézis **igaznak bizonyult**. A nagy nyelvi modellek értelmezhetőek önálló OSINT eszközként tekintettel arra, hogy a különböző integrációk (pl.: a ChatGPT esetében a Bing, a Gemini vonatkozásában a Google Search) lehetőséget ad a valós idejű nyílt forrású információkeresésre (3.5.4. fejezet). Az LLM-OSINT PoC néven vizsgált esettanulmány első körös OSINT eszközként alkalmazza a GPT-4-et (4.1. fejezet). Azonban ki kell egészíteni azzal a

megállapítással, hogy a nagy nyelvi modellek keresési funkciójában jelentős az erőforráskorlátozás. Gyorsan visszatérnek a „válasszal” és csekély mértékű adatot szállítanak, nem szolgálják ki teljeskörűen az OSINT elvárásokat. Továbbá azt tapasztaltuk, hogy az LLM-ek leginkább OSINT célú keretrendszerekbe integrálódnak, vagy ilyen keretrendszerekbe szolgáltatnak LLM képességeket, hatékonyságukra tekintettel nem jellemző önálló OSINT eszközként való alkalmazásuk (5.2. fejezet).

H2: *A nagy nyelvi modellek bővíthetik a kibertéri OSINT tevékenységek információfeldolgozási határait.*

Álláspontom szerint a **hipotézis igaz**. A 4. fejezetben feldolgozott esettanulmányokban azonosított keretrendszerek (5.2. fejezet) eljárásai bizonyítják, hogy a 2.5. fejezetben említett OSINT gyengeségi jellemzők (adatdömping és elemzési képességek hiánya) redukálására az LLM természeténél fogva képes. Az LLM erősségek (3.7. fejezet) OSINT jellegű kihasználását is bizonyítják a gyakorlati példák, amelyekkel az OSINT-LLM keretrendszerek képessé válnak nagyobb mennyiségű információ hatékonyabb feldolgozására.

H3: *a nagy nyelvi modellek hatékonyabbá teszik a kibertámadásokkal kapcsolatos OSINT tevékenységet.*

Ez a hipotézis is **igaznak bizonyult**. A kutatás a kérdés elméleti alapjait körbejárta, a gyakorlatban megjelenő esettanulmányok (4. fejezet) elemzése meg bizonyította azokat. Az OSINT célú keretrendszerekbe OSINT eszközök sora kerül integrálásra (vagy kapcsolt szolgáltatásként jelennek meg), majd a nagyhalmazú adatok hatékonyabb elemzésére LLM képességeket vesznek igénybe (pl.: Ollama), vagy LLM asszisztenciát alkalmaznak (pl.: ORA). Az így létrejött és a dolgozatban bemutatott OSINT-LLM eszközök **értelmezhetőek a Cyber Kill Chain szakaszaiban**, amely igazolja a kibertámadásban való hatékonyabb alkalmazhatóságot.

H4: *Léteznek olyan új generációs elhárítási technikák, amelyek csökkenthetik a nagy nyelvi modellekkel végzett nyílt forrású információgyűjtés kiberbiztonsági kockázatait.*

Az LLM technológia elemzése során megállapításra került, hogy azokkal kapcsolatban a fenyegetések és gyengeségek száma jelentős (3.7. fejezet). Ezeket a jellemzőket elhárítási célú

technikai vagy eljárási elemekként is értelmezhetjük, **tehát ez a hipotézis igaz**. A hipotézisre adott válasz részletesebb kifejtése a dolgozat egy későbbi, önálló szakaszában (5.7. fejezet) történik meg.

5.5. Eredmények

A dolgozat sikerrel tárt fel olyan tudományos eredményeket, amelyek új ismeretként hozzájárulnak a kiberbiztonság, mint interdiszciplináris tudományterület fejlesztéséhez, előrehaladásához. Ezek a megállapítások álláspontom szerint jelentőséggel bírnak, továbbgondolásra érettek, és egyéb tudományos problémák megfogalmazásához segíthetik hozzá a tudományos közösséget.

- Részeredményként tekinthetünk az OSINT tevékenységgel, valamint a nagy nyelvi modellekkel kapcsolatos **stratégiai szempontú SWOT analízisekre**, amely ismeretanyagok további kutatások alapjául szolgálhatnak.
- Tudományos eredményként tekinthetünk a **hipotézisek igazolására**, miszerint a nyílt forrású információszerzés és nagy nyelvi modellek kölcsönhatása kulcsfontosságú a kiberbiztonsági fenyegetések megértésében és kezelésében.
- Tudományos eredményként fogalmazódott meg az OSINT-LLM keretrendszerek **alapvető komponenseinek, típusainak és jellemzőinek azonosítása**. Ezek az eredmények olyan megállapításokon nyugszanak, amelyeket a dolgozat leíró részében átláthatóan érvényesítettem.
- Tudományos eredménynek tekinthető az OSINT-LLM keretrendszerek Cyber Kill Chain modellben történt értelmezése, amely hozzájárul a **LLM-el támogatott kiberbehatolási tevékenységek azonosításához és megelőzéséhez**. Felhívja a kiberbiztonsággal kapcsolatos tudományos közeg, valamint a szakemberek figyelmét **az új és eltérő mértékű kockázatok jelenlétére és jövőbeli kiértékelésére**.

5.6. Az LLM további várható megjelenése az OSINT tevékenységben

Az általam tett megállapítás, miszerint az OSINT-LLM keretrendszerek fejlődése még **kezdeti fázisban van**, megfogalmazza azt a kérdést, hogy a jövőben milyen irányban várható előrelepésük. Napjainkban megfigyelhetjük azt a folyamatot, hogy a LLM megoldásokat számos technológiai

vállalat integrálja saját szolgáltatásába (pl.: Microsoft CoPilot²⁵³, Google Assistant with Bard²⁵⁴) egyfajta **asszisztenciaként**. Úgy vélem, hogy a közeli jövőben az OSINT területén továbbra is egyfajta **támogató, hatékonyság segítő funkciót fog ellátni** az LLM.

Az OSINT-LLM keretrendszerekbe egyre több OSINT eszköz integrálására lehet számítani, hiszen az **LLM feldolgozási kapacitásai még koránt sincsenek kihasználva**. Ez arra enged következtetni, hogy az OSINT-LLM tevékenységek érvényt fognak szerezni a CKC további szakaszaiban is. Vélekedésem szerint a jövőben számíthatunk speciális, **OSINT céllal összeállított korpuszok megjelenésére is**, és az ezen tanított nyílt forrású nagy nyelvi modellek még hatékonyabbá válhatnak a nyílt forrású információgyűjtésben.

Azt gondolom, hogy az OSINT képességek terén, a bemutatott keretrendszer-típusok közül **a helyi integrációt alkalmazó variáns megjelenésére számíthatunk hosszabb távon**, hiszen képességeiben, jellemzőiben, hatásaiban és a járulékos költségeiben **jobban illeszkedik az OSINT tevékenységekkel szembeni elvárásokhoz**. Míg a szolgáltatás alapú keretrendszert jelentős kitettsége **hátrébb fogja szorítani**.

Az OSINT tevékenységek során a megszerzett információk **hitelessége** továbbra is kulcsfontosságú lesz, így véleményem szerint, ha sikerül az LLM technológia **hallucináció-jelenségének megszüntetése**, vagy nem számottevő mértékre történő **visszaszorítása**, az LLM technológia térnyerésével paradigmaváltás fog bekövetkezni a kibertéri OSINT területén is.

Ettől függetlenül, valószínűsíthető, hogy **nem az OSINT tevékenységben való LLM alkalmazás hordozza a legmagasabb kockázatokat** a kiberbiztonság állapotára. Meglátásom szerint a különböző CKC szakaszokban való megjelenésre egyidejűleg fejlesztett támadó célú LLM keretrendszerek **számottevő zavarokat okozhatnak a jövőben**, hiszen a mélytanulási képességek rendkívüli automatizált támadási láncolatok végrehajtására lehetnek képesek.

A kiberbiztonság állapotát fenyegető tényezők rendszeresen indukálódnak, és bizonyos, hogy ezen tényezők közé fog tartozni az **LLM modellek okozta fenyegetés is**. Tekintettel azokra a megállapításokra, hogy az LLM rugalmasan felhasználható és számos cél kiszolgálására képes,

²⁵³ Copilot. Elérhető: <https://copilot.microsoft.com/> (letöltve: 2024.02.15.)

²⁵⁴ Sissie Hsiao (2023.10.04.). Assistant with Bard: A step toward a more personal assistant. Elérhető: <https://blog.google/products/assistant/google-assistant-bard-generative-ai/>

valamint arra, hogy a kevesebb paraméterekkel rendelkező modellek már otthoni környezetben is futtathatók, a jövőben megjelenő LLM fenyegetések formája egyelőre csak körvonalazódhat, de **joggal gondolhatunk arra a kedvezőtlen jövőképre, hogy a kibertámadások közel teljes palettáján teret fog nyerni.**

5.7. OSINT-LLM tevékenység elhárítását támogató javaslatok

Megítélésem szerint, és a korábban tett megállapításokra hivatkozva kijelenthetjük, hogy az OSINT-LLM megoldások jelenlegi alkalmazási formái **nem nyitnak meg kiemelkedően új megközelítéseket a nyílt forrású információszerzés területén.** Az adatforrásokból származó adatmennyiség kezelésében azonban várhatóan az eddigi **képességek többszörösét fogják elvégezni.** Az OSINT-LLM keretrendszerben megjelenő OSINT eszköz kapcsolatok száma nőni fog, amellyel a támadások komplexitása egyenes arányban növekszik. Egy célpontra, **egy műveleti eljárásban,** számos OSINT fenyegetés fog **megjelenni azonos időben,** így az ezzel szembeni védelmet **is fel kell felzárkóztatni.** Ennek elsődleges eleme a OSINT elleni védekezési módszerek, eljárások alkalmazása, amelyet ez a **tanulmány 3.4 fejezetben tárgyalt.** A digitális lábnyom csökkentése redukálja a felhasználható információt, amely egyenes arányban van az LLM képességek csökkentésével. Azonban a kutatás feltárt pár olyan speciális lehetőséget, amelyek felhasználása még tovább csökkentheti az LLM technológiákkal támogatott OSINT sikerességét. Ezek felsorolását a Mark. M. Lowenthal (2003) által meghatározott kémelhárítási kategóriák (counterintelligence)²⁵⁵ szerinti csoportosításban végzem el, az alábbiak szerint:

5.7.1. Kollektív elhárítás

A kollektív típusú elhárítás információkat szerez az ellenfél hírszerzési képességeiről. Ennek eleget téve, ebben a dolgozatban is alkalmazott módszertanok alapján az LLM-OSINT megoldások időszakos, **rendszeres felülvizsgálata és kiértékelése hozzájárulhat a kockázatok kezeléséhez.** A témában releváns új tudományos publikációk, esettanulmányok, a nyílt forrású projektek vizsgálata és tesztelése, valamint a technológiai óriások által szolgáltatott LLM képességek folyamatos elemzése megfelelő információt szolgáltat a fenyegetettségek jellemzőiről. A **dark weben fellelt információk,** különösen az illegális piactereken áruba bocsátott LLM technológiák

²⁵⁵ Mark. M. Lowenthal (2003). Intelligence: From secrets to policy. Washington, DC: CQ Press. p.263.

elemzése rámutathat a legújabb fenyegetések megjelenési formájára, de ugyanez vonatkozik az OSINT eszközökre is. A kifejezetten önálló, vagy API hozzáféréssel rendelkező OSINT eszközök integrálhatók az LLM-OSINT architektúrába, így az **ilyen eszközök megjelenésére fókuszáló hírszerzés is** jól hasznosulhat a normál OSINT és az OSINT-LLM alkalmazással szemben egyaránt. Értékes információt lehet szerezni a kibertámadások során felismert OSINT tevékenységgel **kapcsolatos esettanulmányok vizsgálatából**, kifejezetten az LLM alkalmazás szemszögéből vizsgálva a tapasztalatokat.

5.7.2. Defenzív elhárítás

A védekező **elhárítás meghiúsítja az ellenséges hírszerzés erőfeszítéseit**. E perspektívából alkalmazandók az OSINT-re és LLM-re vonatkozó SWOT **analízisekben feltárt gyengeségek és fenyegetések együttes, defenzív célú értelmezése**. Az LLM sebezhetőségek a végponttól-végpontig tartó LLM-el végzett valós idejű nyílt forrású információgyűjtés, valamint az OSINT-LLM megoldások ellen egyaránt jó eséllyel érvényesíthetők. Egy **honeypotba**²⁵⁶ több típusú védelmi elem elhelyezhető. A kibertéri **web scraping** vizsgálat alá veszi a surface és a deep webet is. A deep weben létrehozott **honeypotként elhelyezett tömeges adathalmaz** csökkentheti az LLM-ek hatékonyságát. Továbbá a digitális lábnyomba (pl.: weboldalak forráskódjaiba) elhelyezett **prompt injection utasítás** „káros kódként” érvényesülhet, de **az indexelés során létrejövő archívumokba jutva is megakadályozhatja az LLM-el történő adatfeldolgozást**.

Hasonlóan működhet a **toxicitás** alkalmazása, tekintve a nagy nyelvi modellek azon képességét, hogy „**megmérgezhetőek**”. Vagyis, olyan típusú és mértékű toxikus tartalmat rejtünk el a digitális lábnyomunkban, amely az **általános felhasználást nem érinti**, de az LLM tanítását, vagy az LLM adatfeldolgozását képes **akadályozni, befolyásolni**. Vizsgálható az a **felvetés** az is, hogy milyen **védelmi hatást indukál**, ha a digitális lábnyom környezetét aszerint alakítjuk, amely nem optimális az LLM feldolgozáshoz (pl.: a csak angol nyelven edzett modellek nem boldogulnak specifikus karakterű, például orosz, kínai nyelvek feldolgozásával). Ugyanakkor az említett **honeypotok szöveges tartalmát** érdemes **LLM-el generáltatni**, hiszen a **tokenizált szöveg**

²⁵⁶A kiberbiztonságban a "honeypot" egy csapda, amely úgy van kialakítva, mintha értékes célpont lenne. Így szándékosan vonzza a hackereket és más rosszindulatú felhasználókat, azonban valójában izolált és ellenőrzött környezetben működnek, így nem jelentenek valódi kockázatot a szervezet számára.

feldolgozását előtérbe helyezi és könnyebben is reprezentálja a modell, tehát a beszivárgás egyértelműen hatékonyabb lehet.

Jogi eszközök is igénybe vehetők, bár e terület nem bővelkedik lehetőségben. A jog mindig is utókövető módon volt/van a technológiához képest, de a különböző keresőmotorokból, archiválást végző **szolgáltatók indexeléséből való eltávolíttatás** néhány indítvány megküldésével végrehajtható, amely egyértelműen kihatással lehet az OSINT-LLM feldolgozásra.

Figyelembe vehetjük az OSINT-LLM kitettséget, hogy az ellátási lánc sebezhetőségeit is kihasználhatjuk. A kollektív elhárítás során feltárt információk alapján technikai tiltásokat alkalmazhatunk a legnépszerűbb – OSINT-LLM keretrendszerek által felhasznált – web scraping, vagy LLM eszközökre. Ezzel az intézkedéssel eshetőlegesen kivonjuk magunkat a feltérképezésből, így védve adatainkat.

5.7.3. Offenzív elhárítás

Az **offenzív elhárítás** azonosítja az ellenfél rendszerének erősségeit, és megpróbálja **hamis információkkal** ellátni azt. Tulajdonképpen ez egy potenciális **OSINT-LLM dezinformáció**. A dezinformáció történhet **honeypotot** alkalmazva, vagy eredeti forráskódba rejtve, de akár deep vagy dark weben elhelyezve is. Megoldás lehet, ha **elavult vagy teljesen hamis információkat** tartalmazó adathalmazokat helyezünk el ezeken a felületeken. A dezinformáció hatékonyságát növelheti, ha **a szöveges tartalmakat nagy nyelvi modellekkel generáltatjuk**, így egy már **tokenizált szöveget teszünk elérhetővé**, amelyet a feldolgozás során előnyben részesítenek a nagy nyelvi modellek.

6. A KUTATÁS ALKALMAZÁSÁNAK TERÜLETEI ÉS TOVÁBBI IRÁNYAI

A kutatott téma alkalmazási területe leginkább a **kiberbiztonság** azon ágazata, amely a **nyílt forrású információgyűjtés elleni védelmi lehetőségeket vizsgálja**. E mellett az eredmények több egyéb területen hasznosíthatóak. Egyrészt a **gazdasági szervezetek és az állami szervek védelmi megoldásaiban** és az ezzel kapcsolatos **stratégiaalkotásban**, másrészt a

kiberbiztonsági szakemberek képzési és oktatási programjai során a témával kapcsolatos kockázatok, valamint az elhárítási megoldások megismertetésére.

Ez a kutatás a kvalitatív eszközökkel végzett, elméleti megközelítést helyezte előtérbe. Ugyanakkor egy gyakorlati vizsgálat – vagyis a **kvantitatív eszközökkel való górcső alá vétel** – nyilvánvalóan lehetőséget biztosít további megállapítások megtételére. Akár az érintett esettanulmányokkal kapcsolatos, **mélyebb összefüggések feltárása** is ígéretes megközelítés lehet. Azoknak, akik a gyakorlati alkalmazás kihívásait keresik, **kiemelten ajánlom figyelmükbe egy OSINT-re finomhangolt nyílt forrású LLM rendszer létrehozását**, majd az elért **eredmények elemzését**, hiszen a kutatás látóterébe ilyen rendszer nem került, ugyanakkor a tapasztalatok alapján (elméletben) kiemelkedő eredményekre lehet képes.

Nem lehet figyelmen kívül hagyni azt a tényt, hogy az LLM fejlődése olyan iramot diktál, hogy vizsgált terület **időszakos újra kutatása** bizonyosan frissebb eredményekkel fog szolgálni. A személyes adatokat, az adatbiztonságot, az etikai és jogi megközelítéseket a központba helyezve **újabb perspektívákból** közelíthetjük meg akár mindkét tématerületet egyszerre.

A közelgő kockázatalapú **szabályozási szándék** (mesterséges intelligenciára vonatkozóan harmonizált szabályok) érvényesülése és alkalmazhatósága az OSINT-LLM területén érdekes felvetésnek bizonyul. Jogi megközelítéssel adott esetben „*de lege ferenda*” javaslatok is tehetők.

Bizonyos, hogy a jövőben megjelenő új LLM képességek – **különösen a multimodális rendszerek** – a lehetséges megközelítések tárházát fogják biztosítani. Biztos vagyok abban is, hogy a nagy nyelvi modellek – több tudományterületen egyaránt – kiváló vizsgálati alapot jelentenek majd a jövő kutatói számára az elkövetkezendő években.

„A tudományt művelő egyén szerepe, mint az ötvösmesteré, ki különféle matériákat egyesít egy nemesebb célért. Hasonlóképpen kell nekünk is a nagy nyelvi modellekben rejlő lehetőségeket harmonikus és eredményes eljárásokká formálnunk saját tudományunk üstjében. A mesterséges intelligencia, mint innováció, jelenkorunkat nemesíti. Nekünk csak jól kell tudni ötvözni...”

A szerző gondolatai

7. RÖVIDÍTÉSEK JEGYZÉKE

Rövidítés	Angol	Magyar
AI	Artificial intelligence	Mesterséges intelligencia (MI)
AFPET	Anti-Fingerprinting Privacy Enhancing Technologies	Digitális Ujjlenyomat-elleni Adatvédelmi Technológiák
API	Application programming interface	Alkalmazásprogramozási felület
ASI	All-source intelligence	Teljes spektrumú hírszerzési tevékenység
BEC	Business email compromise	Vállalati e-mail kompromittálás
BERT	Bidirectional Encoder Representations from Transformers	Kétirányú kódoló reprezentációk transzformátorokból
CIA	Central Intelligence Agency	Központi Hírszerző Ügynökség
CKC	Cyber Kill Chain	Kibertámadási láncolat
DL	Deep learning	Mély tanulás
DNS	Domain name system	Tartománynévrendszer
GPT	Generative pre-trained transformer	Generatív előre tanított transzformátor
HUMINT	Human intelligence	Emberi hírszerzés
IDE	Integrated development environment	Integrált fejlesztői környezet
IE	Information extraction	Információ kinyerés
IMINT	Imagery intelligence	Képi hírszerzés
IR	Information retrieval	Információ visszakeresés
LLM	Large language model	Nagy nyelvi modell
LM	Language model	Nyelvi modell
LMM	Large multimodal model	Nagy multimodális modell
ML	Machine learning	Gépi tanulás
MT-NLG	Megatron-turing natural language generation	Megatron-turing természetes nyelv generálás
NATO	North Atlantic Treaty Organization	Észak-atlanti Szerződés Szervezete
NLI	Natural language inference	Természetes nyelvű következtetés
NLP	Natural Language Processing	Természetes nyelv feldolgozás
OSINT	Open source intelligence	Nyílt forrású hírszerzés
PET	Privacy Enhancing Technologies	Magánélet védelmét elősegítő technológiák
PLM	Pretrained language model	Előre tanított nyelvi modell
QA	Question and answering	Kérdés-felelet
RL	Reinforcement learning	Megerősítéses tanulás
RoBERTa	Robustly optimized BERT approach	Robusztus optimalizált BERT megközelítés
SIGINT	Signals intelligence	Rádióforgalom figyelés
SL	Supervised learning	Felügyelt tanulás
SQL	Structured Query Language	Strukturált lekérdezési nyelv
SSL	Semi-supervised learning	Félig felügyelt tanulás
TECHINT	Technological intelligence	Technológiai hírszerzés
UL	Unsupervised learning	Felügyelet nélküli tanulás
URL	Uniform resource locator	Egységes erőforrás-helymeghatározó

8. IRODALOMJEGYZÉK

Monográfiák

K.R. Chowdhary (2020). Fundamentals of Artificial Intelligence. Springer India 1st ed. ISBN: 81-322-3972-5. DOI: <https://doi.org/10.1007/978-81-322-3972-7>.

Kim Rosengren (2022). Contribution of Open-Source Intelligence to Social Engineering Cyberattacks. Bachelor's Thesis. Turku University of Applied Sciences. Finland.

Kovács Zoltán (2015): A titkosszolgálatok és az információgyűjtés jövőbeli lehetséges irányai, új technológiák megjelenése és hatásai a titkosszolgálati tevékenységekre in: Boda József. A nemzetbiztonság technikai kihívásai a 21. században. ISBN: 9786155527746.

Mark. M. Lowenthal (2003). Intelligence: From secrets to policy. Washington, DC: CQ Press.

Lévai Gábor (2006). OSINT (Open Source Intelligence) – Nyílt információs hírszerzés. Egyetemi jegyzet, Budapest, Zrínyi Miklós Nemzetvédelmi Egyetem, 2006.

Vadász Pál (2018). A szemantikus keresés módszerei és alkalmazási lehetőségei a védelmi szférában, a közigazgatásban, illetve a gazdasági életben. Doktori (PhD) értekezés. Nemzeti Közszerológati Egyetem Katonai Műszaki Doktori Iskola.

Vida Csaba (2013). Nyílt forrású adatszerzés (OSINT) In: Kobolka István (szerk.): Nemzetbiztonsági alapismeretek, Nemzeti Közszerológati és Tankönyv Kiadó Zrt. ISBN 978-915-5344-32-9.

Folyóiratok és konferencia-kiadványok

A. Chernyavskiy, D. Ilvovsky, and P. Nakov (2021). Transformers: “the end of history” for natural language processing?” in Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain. September 13–17. 2021, Proceedings, Part III 21. Springer.

Bányász Péter (2015). A közösségi média, mint a nyílt forrású információszerzés fontos területe. Nemzetbiztonsági Szemle, 3. sz. (2). Elérés: <http://real.mtak.hu/id/eprint/72506>.

Blaise Agüera y Arcas (2022). Do large language models understand us. Daedalus. vol. 151, no. 2, p. 183–197. DOI: https://doi.org/10.1162/daed_a_01909.

Bonnyai Tünde, Kiss Adrienn, Tóth András (2023). Nagyvállalati biztonságtudatosság és nyílt forrású információszerzés: Kiber- és információbiztonsági tanulmányok, Budapest, Nemzeti Közszerológati Egyetem Ludovika Egyetemi Kiadó. ISBN: 978615659869. Elérés: <https://webshop.ludovika.hu/termek/konyvek/gazdasagtudomany/nagyvallalati-biztonsagtudatossag-es-nyilt-forrasu-informacioszerzes/>.

C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu (2020): Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, vol. 21, no. 1, p. 5485–5551.

Csizner Zoltán (2019). Az OSINT határai. Nemzetbiztonsági Szemle 7. évfolyam 2. szám, Budapest. DOI: <https://doi.org/10.32561/nsz.2019.2.2>.

Deák Veronika (2018). A nyílt forrású információszerzés szerepe a kibertámadások végrehajtása során. Hadmérnök, 13(3), 391–402. Elérés: <https://folyoirat.ludovika.hu/index.php/hadmernok/article/view/3859>.

Dobák Imre (2019). OSINT – Gondolatok a kérdéskörhöz. Nemzetbiztonsági Szemle 7. évfolyam 2. szám. Budapest. DOI: <http://doi.org/10.32561/nsz.2019.2.7>.

Ghioni, R., Taddeo, M. & Floridi, L. Open source intelligence and AI: a systematic review of the GELSI literature. AI & Soc (2023). DOI: <https://doi.org/10.1007/s00146-023-01628-x>.

João Rafael Gonçalves Evangelista, Renato José Sassi, Márcio Romero & Domingos Napolitano (2021). Systematic Literature Review to Investigate the Application of Open Source Intelligence (OSINT) with Artificial Intelligence. Journal of Applied Security Research, 16. DOI: <https://doi.org/10.1080/19361610.2020.1761737>.

John R. Searle (1980). Minds, brains, and programs. Behavioral and Brain Sciences. 1980;3(3):417-424. DOI: <https://doi.org/10.1017/S0140525X00005756>.

Pariwat. Ongsulee (2017). Artificial intelligence, Machine learning and Deep learning, 15th International Conference on ICT and Knowledge Engineering (ICT&KE), Bangkok, Thailand. DOI: <https://doi.org/10.1109/ICTKE.2017.8259629>.

Qusef Abdallah & Alkilani Hamzeh (2022). The effect of ISO/IEC 27001 standard over open-source intelligence. PeerJ Comput Sci. DOI: <https://doi.org/10.7717/peerj-cs.810>.

Solti István (2019). Az OSINT információgyűjtő eszközeiről. Nemzetbiztonsági Szemle, 7. évfolyam 2. szám. Budapest. DOI: <http://doi.org/10.32561/nsz.2019.2.1>.

Tóth Tivadar (2001). Katonai szervezetek és a nyílt forrású információgyűjtés. Repüléstudományi közlemények XIII. évfolyam 33. szám. Elérhető: https://real-j.mtak.hu/20046/2/rtk_2001_13_2_.pdf#page=66.

Yamin, Muhammad Mudassar, Mohib Ullah, Habib Ullah, Basel Katt, Mohammad Hijji, and Khan Muhammad. 2022. "Mapping Tools for Open Source Intelligence with Cyber Kill Chain for Adversarial Aware Security" Mathematics 10, no. 12: 2054. DOI: <https://doi.org/10.3390/math10122054>.

Tudományos dokumentációk

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, Parker Schuh, Kensen Shi, Sasha Tsvyashchenko, Joshua Maynez, Abhishek Rao, Parker Barnes, Yi Tay, Noam Shazeer, Vinodkumar Prabhakaran, Emily Reif, Nan Du, Ben Hutchinson, Reiner Pope, James Bradbury, Jacob Austin, Michael Isard, Guy Gur-Ari, Pengcheng Yin, Toju Duke, Anselm Levskaya, Sanjay Ghemawat, Sunipa Dev, Henryk Michalewski, Xavier Garcia, Vedant Misra, Kevin Robinson, Liam Fedus, Denny Zhou, Daphne Ippolito, David Luan, Hyeontaek Lim, Barret Zoph, Alexander Spiridonov, Ryan Sepassi, David Dohan, Shivani Agrawal, Mark Omernick,

Andrew M. Dai, Thanumalayan Sankaranarayanan Pillai, Marie Pellat, Aitor Lewkowycz, Erica Moreira, Rewon Child, Oleksandr Polozov, Katherine Lee, Zongwei Zhou, Xuezhi Wang, Brennan Saeta, Mark Diaz, Orhan Firat, Michele Catasta, Jason Wei, Kathy Meier-Hellstern, Douglas Eck, Jeff Dean, Slav Petrov & Noah Fiedel (2022). **PaLM: Scaling Language Modeling with Pathways**. Google Research. DOI: <https://doi.org/10.48550/arXiv.2204.02311>.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser & Illia Polosukhin (2017): **Attention Is All You Need**. DOI: <https://doi.org/10.48550/arXiv.1706.03762>.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov & Thomas Scialom (2023). **Llama 2: Open Foundation and Fine-Tuned Chat Models. GenAI, Meta**. DOI: <https://doi.org/10.48550/arXiv.2307.09288>.

Jacob Devlin, Ming-Wei Chang, Kenton Lee & Kristina Toutanova (2018). **BERT: Pre-training of deep bidirectional transformers for language understanding**. DOI: <https://doi.org/10.48550/arXiv.1810.04805>.

Richard Fang, Rohan Bindu, Akul Gupta, Qiusi Zhan, Daniel Kang (2024). **LLM Agents can Autonomously Hack Websites** DOI: <https://doi.org/10.48550/arXiv.2402.06664>.

Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Muniyikwa, Suraj Nair, Avaniika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu,

Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou & Percy Liang (2021). **On the Opportunities and Risks of Foundation Models**. Center for Research on Foundation Models, Stanford Institute for Human-Centered Artificial Intelligence, Stanford University. DOI: <https://doi.org/10.48550/arXiv.2108.07258>.

Sofia Eleni Spatharioti, David M. Rothschild, Daniel G. Goldstein, Jake M. Hofman (2023). **Comparing Traditional and LLM-based Search for Consumer Choice: A Randomized Experiment**. Microsoft Research, New York City. DOI: <https://doi.org/10.48550/arXiv.2307.03744> p.15.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever & Dario Amodei (2020). **Language Models are Few-Shot Learners**. DOI: <https://doi.org/10.48550/arXiv.2005.14165>.

Youngjin Jin, Eugene Jang, Jian Cui, Jin-Woo Chung, Yongjae Lee & Seungwon Shin (2023): **DarkBERT: A Language Model for the Dark Side of the Internet**. DOI: <https://doi.org/10.48550/arXiv.2305.08596>.

Yupeng Chang, Xu Wang, Jindong Wang, Yuan Wu, Linyi Yang, Kaijie Zhu, Hao Chen, Xiaoyuan Yi, Cunxiang Wang, Yidong Wang, Wei Ye, Yue Zhang, Yi Chang, Philip S. Yu, Qiang Yang & Xing Xie (2023). **A Survey on Evaluation of Large Language Models**. DOI: <https://doi.org/10.48550/arXiv.2307.03109>.

Jogszabályok

2013. évi L. törvény az állami és önkormányzati szervek elektronikus információbiztonságáról.
15 U.S. Code § 9401.

Az európai parlament és a tanács rendelete a mesterséges intelligenciára vonatkozó harmonizált szabályok (a mesterséges intelligenciáról szóló jogszabály) megállapításáról és egyes uniós jogalkotási aktusok módosításáról. Elérhető: <https://eur-lex.europa.eu/legal-content/HU/TXT/?uri=CELEX:52021PC0206>.

Egyéb kiadványok

Andres Munoz (2012). Machine Learning and Optimization. Elérhető: https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf.

Dictionary of United States Military Terms for Joint Usage, Departments of the Army, Navy, and Air Force, Washington, D.C., May 1995.

Mesterséges Intelligencia Elektronikus Almanach. Elérhető: http://project.mit.bme.hu/mi_almanach/books/aima/ch01s03.

NATO Open Source Intelligence Handbook (2001). Elérhető: <https://github.com/lawsecnet/OPSEC/blob/master/NATO%20OSINT%20Handbook%20v1.2%20-%20Jan%202002.pdf> .

Részletes leírás a CSIRT csoportok létrehozásáról. ENISA. Elérhető: <https://www.enisa.europa.eu/publications/csirt-setting-up-guide-in-hungarian/@@download/fullReport>.

Online hivatkozások

0day.today - Biggest Exploits Database and 0day market. Elérhető: <https://en.0day.today/faq> (letöltve: 2024.01.06.).

Ali Alvi (2021.10.21): Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model. Elérhető: <https://www.microsoft.com/en-us/research/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model> (letöltve: 2024.02.12.).

AndY your desktop mobile. Elérhető: <https://www.andyroid.net/> (letöltve: 2024.01.09.).

Beautifulsoup. Elérhető: <https://pypi.org/project/beautifulsoup4/> (letöltve: 2024.02.17.).

Bitbucket. Elérhető: <https://bitbucket.org/> (letöltve: 2024.03.02.).

BlueStacks by now.gg. Elérhető: <https://www.bluestacks.com/> (letöltve: 2024.01.09.).

Censys. Elérhető: <https://censys.com> (letöltve: 2024.01.05.).

Charlie Osborne (2023.10.24.). WormGPT: What to know about ChatGPT's malicious cousin. Elérhető: https://www.zdnet.com/article/wormgpt-what-to-know-about-chatgpts-malicious-cousin/#google_vignette (letöltve: 2024.03.12.).

Chip Huyen (2023.10.10). Multimodality and Large Multimodal Models (LMMs). Elérhető: https://huyenchip.com/2023/10/10/multimodal.html#part_1_understanding_multimodal (letöltve: 2024.02.18.).

Chris Poulter (2023.015.09). OSINT Workflow with ChatGPT: Tips, Risks, and Benefits. Elérhető: <https://www.osintcombine.com/post/osint-workflow-with-chatgpt-tips-risks-and-benefits> (Letöltés: 2024.03.16).

Common Crawl. Common Crawl maintains a free, open repository of web crawl data that can be used by anyone. Elérhető: <https://www.commoncrawl.org/> (letöltve: 2024.02.15.).

Copilot. Elérhető: <https://copilot.microsoft.com/> (letöltve: 2024.02.15.).

Corby Rosset (2020.02.13.). Turing-NLG: A 17-billion-parameter language model by Microsoft. Elérhető: <https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/> (letöltve: 2024.02.12.).

Cryptome. Elérhető: <https://cryptome.org> (letöltve: 2024.01.06.).

Cyber Kill Chain®. Elérhető: <https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html> (letöltve: 2024.01.05.).

CyberRaya (2023.10.23.). InfoHound: Enumerate emails, usernames, subdomains, names and other information. Elérhető: <https://osintteam.blog/infhound-enumerate-emails-usernames-subdomains-names-and-other-information-828bb9b8ac8e> (letöltve: 2024.02.18.).

Daniel Kelley (2023.07.13). WormGPT – The Generative AI Tool Cybercriminals Are Using to Launch Business Email Compromise Attacks. Elérhető: <https://slashnext.com/blog/wormgpt-the>

[generative-ai-tool-cybercriminals-are-using-to-launch-business-email-compromise-attacks](#) (letöltve: 2024.03.12.).

DarkBERT ダークウェブで訓練された最初のAI. (japán – magyar fordításban: „DarkBERT az első sötét weben edzett mesterséges intelligencia) Elérhető: <https://s2wjapan.com/> (letöltve: 2024.02.12.).

Default Password Lookup. Elérhető: https://www.fortypoundhead.com/tools_dpw.asp (letöltve: 2024.01.07.).

Default Passwords List. Elérhető: <https://default-password.info> (letöltve: 2024.01.07.).

Default Passwords|CIRT.net. Elérhető: <https://cirt.net/passwords> (letöltve: 2024.01.07.).

Django. Elérhető: <https://www.djangoproject.com/start/overview/> (letöltve: 2024.02.17.).

DNS Recon and Research, Find and Lookup DNS Records. Elérhető: <https://dnsdumpster.com> (letöltve: 2024.01.05.).

DNSpython. Elérhető: <https://www.dnspython.org/> (letöltve: 2024.02.17.)

E-mail validator Elérhető: <https://pypi.org/project/email-validator/> (letöltve: 2024.02.17.).

Ether. Elérhető: https://ether.gtisc.gatech.edu/web_unpack/ (Letöltve:2024.01.08.).

Exploit Database-Exploits for Penetration Testers, Researchers, and Ethical Hackers. Elérhető: <https://www.exploit-db.com> (letöltve: 2024.01.05.).

Find People for Free|Get Their Contact Info|ThatsThem. Elérhető: <https://thatsthem.com/> (letöltve: 2024.01.07.).

Free WHOIS – Domain Name Lookup. Elérhető: <https://www.ip2whois.com> (letöltve: 2024.01.05.).

Genymotion Android Emulator. Elérhető: <https://www.genymotion.com> (letöltve: 2024.01.09.).

GitHub — 0x09AL/raven. Elérhető: <https://github.com/0x09AL/raven> (letöltve: 2024.01.06.).

GitHub — achillean/shodan-Python Elérhető: <https://github.com/achillean/shodan-python> (letöltve: 2024.02.17.).

GitHub — Fundacio-i2CAT/InfoHound Elérhető: <https://github.com/Fundacio-i2CAT/InfoHound?tab=readme-ov-file> (letöltve: 2024.02.17.).

GitHub — lanmaster53/recon-ng. Elérhető: <https://github.com/lanmaster53/recon-ng> (letöltve: 2024.01.06.).

GitHub — maravento/blackweb. Elérhető: <https://github.com/maravento/blackweb> (letöltve: 2024.01.09.).

GitHub — mboot-github/python-whois. Elérhető: <https://github.com/mboot-github/python-whois> (letöltve: 2024.02.17.).

GitHub — Microsoft/DeepSpeed. Elérhető: <https://github.com/microsoft/DeepSpeed> (letöltve: 2024.03.02.).

GitHub — NVIDIA/Megatron-LM. Elérhető: <https://github.com/NVIDIA/Megatron-LM> (letöltve: 2024.02.12.).

GitHub — ollama/ollama. Elérhető: <https://github.com/ollama/ollama> (letöltve: 2024.02.18.).

GitHub — soxoj/maigret. Elérhető: <https://github.com/soxoj/maigret> (letöltve: 2024.02.17.).

GitHub — sshh12/llm_osint. Elérhető: https://github.com/sshh12/llm_osint?tab=readme-ov-file (letöltve: 2024.02.16.).

GitHub. Elérhető: <https://github.com/> (letöltve: 2024.03.02.).

GitLab. Elérhető: <https://gitlab.com/> (letöltve: 2024.03.02.).

GPT-4. Elérhető: <https://openai.com/research/gpt-4> (letöltve: 2024.02.12.).

GPT-J. Transformers documentation. Elérhető: https://huggingface.co/docs/transformers/model_doc/gptj (letöltve: 2024.03.12.).

Harangi László (2020.08.17.): Egy hónap alatt összeszerelte az NVIDIA a világ hetedik leggyorsabb szuperszámítógépét. Elérhető: <https://www.pcwplus.hu/pcwpro/egy-honap-alatt-osszeszerelte-az-nvidia-a-vilag-hetedik-leggyorsabb-szuperszamitogepet-283131.html> (letöltve: 2024.02.12.).

Have I Been Pwned. Elérhető: <https://haveibeenpwned.com> (letöltve: 2024.01.07.).

Hibrid Analysis. Elérhető: <https://www.hybrid-analysis.com/> (letöltve: 2024.01.08.).

Holehe v1.61. Elérhető: <https://snyk.io/advisor/python/holehe> (letöltve: 2024.02.17.).

Holupredictions (2023): It looks like you can use ChatGPT to bypass paywalls. Elérhető: https://www.reddit.com/r/ChatGPT/comments/14j8q1u/it_looks_like_you_can_use_chatgpt_to_bypass/ (letöltve: 2024.03.13.).

Homokozó (biztonság). Wikipédia. Elérhető: [https://hu.wikipedia.org/wiki/Homokoz%C3%B3_\(biztons%C3%A1g\)](https://hu.wikipedia.org/wiki/Homokoz%C3%B3_(biztons%C3%A1g)) (letöltve: 2024.03.17.).

Hunter (Email Hunter). Elérhető: <https://hunter.io> (letöltve: 2024.01.07.).

Internet Archive: Wayback Machine. Elérhető: <https://archive.org/web/> (letöltve: 2024.01.06.).

Isabelle Nguyen (2023.06.15.). The Beginner's Guide to LLM Prompting. Elérhető: <https://haystack.deepset.ai/blog/beginners-guide-to-llm-prompting> (letöltve: 2024.02.15.).

Jeff Dearn (2021.10.28.). Introducing Pathways: A next-generation AI architecture. Elérhető: <https://blog.google/technology/ai/introducing-pathways-next-generation-ai-architecture/>.

Kevin Okemwa (2023.06.04.). ChatGPT pauses Bing integration to stop people from bypassing paywalls. Elérhető: <https://www.windowscentral.com/software-apps/chatgpt-pauses-bing-integration-to-stop-people-from-bypassing-paywalls> (letöltve: 2024.03.13.).

Kyle Wiggers (2020.07.01.). OpenAI's massive GPT-3 model is impressive, but size isn't everything. Elérhető: <https://venturebeat.com/ai/ai-machine-learning-openai-gpt-3-size-isnt-everything/> (letöltve: 2024.02.08.).

Large Language Models. Elérhető: <https://www.nvidia.com/en-us/glossary/large-language-models/> (letöltve: 2024.02.07.).

Legközelebbi szomszéd analízis. Wikipédia, a szabad enciklopédia. Elérhető: https://hu.wikipedia.org/wiki/Legk%C3%B6zelebbi_szomsz%C3%A9d_anal%C3%ADzis

Malware Patrol. Elérhető: <https://www.malwarepatrol.net/> (letöltve: 2024.01.09.).

Map reduce. Elérhető: https://js.langchain.com/docs/modules/chains/document/map_reduce (letöltve: 2024.02.16.).

Meta (2023.08.23): Introducing Code Llama, a state-of-the-art large language model for coding. Elérhető: <https://ai.meta.com/blog/code-llama-large-language-model-coding/> (letöltve: 2024.02.12.).

Mi a vírusvédelem? – Biztonságosinternet.hu Elérhető: <https://www.biztonsagosinternet.hu/hu/tippek/60> (letöltve: 2024.01.08).

Mi a web scraping , és mire használható? Elérhető: <https://web-scraping.hu/mi-az-a-web-scraping/> (letöltve: 2024.02.16.).

Mi az a proof-of-concept (PoC) exploit? Nemzeti Kibervédelmi Intézet. <https://nki.gov.hu/it-biztonsag/tudastar/mi-az-a-proof-of-concept-poc-exploit/> (letöltve: 2024.02.16.).

Mi az önszerveződő térkép? Knowledge Exploration. Forrás: <https://houses-apartments.com/hu/design/parametric-and-computational-design/what-is-a-self-organizing-map/>

Michael Humor (2023.09.10.). Understanding “tokens” and tokenization in large language models. Elérhető: <https://blog.devgenius.io/understanding-tokens-and-tokenization-in-large-language-models-1058cd24b944> (letöltve: 2024.02.15.).

Michael McDonough (2024.01.11.). Large Language Model. Elérhető: <https://www.britannica.com/topic/large-language-model> (letöltve: 2024.01.31.).

Mikhail Khrushchev (2022.06.23). Yandex Publishes YaLM 100B. It’s the Largest GPT-Like Neural Network in Open Source. Elérhető: <https://medium.com/yandex/yandex-publishes-yalm-100b-its-the-largest-gpt-like-neural-network-in-open-source-d1df53d0e9a6> (letöltve: 2024.02.12.).

Mitre Att&ck. Elérhető: <https://attack.mitre.org/> (letöltve: 2024.01.14.).

Nagesh Singh Chauhan (2022.05.04.). OpenAI GPT-3: Understanding the Architecture. Elérhető: <https://www.theaidream.com/post/openai-gpt-3-understanding-the-architecture> (letöltve: 2024.02.12.).

Namechk|Username, Domain, and Trademark Search|Username Registration. Elérhető: <https://namechk.com> (letöltve: 2024.01.07.).

National Vulnerability Database. Elérhető: <https://nvd.nist.gov> (letöltve: 2024.01.05.).

Networkx. Elérhető: <https://pypi.org/project/networkx/> (letöltve: 2024.02.17.).

Nihad Hassan (2024.03.06). Explore mitigation strategies for 10 LLM vulnerabilities. Elérhető: <https://www.techtarget.com/searchEnterpriseAI/tip/Explore-mitigation-strategies-for-LLM-vulnerabilities>.

OpenAI Assistant. Elérhető: <https://platform.openai.com/docs/assistants/overview?context=with-streaming>.

OpenAI ChatGPT 4.0 lekérdezés. Elérhető: <https://chat.openai.com/share/e28593b3-4295-4d39-9181-5e05e374827f> (lekérdezve: 2023.12.17.).

OSINT Framework. Elérhető: <https://osintframework.com> (letöltve: 2024.01.12.).

Paresh Kharya and Ali Alvi (2021.10.11). Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World’s Largest and Most Powerful Generative Language Model. Elérhető: <https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/> (letöltve: 2024.02.12.).

Playwrite. Playwright enables reliable end-to-end testing for modern web apps. Elérhető: <https://playwright.dev/> (letöltés: 2024.03.24).

PostgreSQL: The World's Most Advanced Open Source Relational Database. Elérhető: <https://www.postgresql.org/> (letöltve: 2024.02.17.).

Precognox (2022.03.08.). A hírek és információk megszerzése még nem hírszerzés. Elérhető: <https://precognox.hu/blog/a-hirek-es-informaciok-megszerzese-meg-nem-hirszerzes/> (letöltve: 2023.12.22.).

Prominent Data Collection Methods and Tools for LLMs. Elérhető: <https://www.turing.com/resources/data-collection-methods-and-tools-for-llms> (letöltve: 2024.02.14.).

Prompthub. Elérhető: <https://www.prompthub.us> (letöltve: 2024.02.17.).

Redis. Elérhető: <https://redis.io/> (letöltve: 2024.02.17.).

Reece Rogers (2023.10.15.). How to Use ChatGPT's 'Browse With Bing' Tool—Plus 6 Starter Prompts. Elérhető: <https://www.wired.com/story/chatgpt-browse-with-bing-internet-connected-openai/> (letöltve: 2024.03.13.).

Registered Domain Names Search — Find Registered Domain Names. Elérhető: <https://instantdomainsearch.com> (Utolsó leoltés: 2024.01.06.).

ScrapingBee. Tired of getting blocked while scraping the web? Elérhető: https://www.scrapingbee.com/?fpr=huynh12&gad_source=1&gclid=CjwKCAiArLyBhA7EiwA-qo80GhBbmfvz-QWJsbjJE6fqWo8Cm1v4E6uoeTjrkW0B3xjht-D2sgRoC2QYQAvD_BwE (letöltve: 2024.02.16.).

Scrapy. An open source an collaborative framework for extracting the data you need from websites. Elérhető: <https://scrapy.org/> (letöltve: 2024.02.15.).

Sean Michael Kerner (2023): Pathways Language Model (PaLM). Elérhető: <https://www.techtarget.com/whatis/definition/Pathways-Language-Model-PaLM> (letöltve: 2024.02.12.).

Shodan. Elérhető: <https://www.shodan.io> (letöltve: 2024.01.05.).

Sissie Hsiao (2023.10.04.). Assistant with Bard: A step toward a more personal assistant. Elérhető: <https://blog.google/products/assistant/google-assistant-bard-generative-ai/>.

Social Searcher – Free Social Media Search Engine. Elérhető: <https://www.social-searcher.com> (letöltve: 2024.01.06.).

SQL-injekció. Wikipédia. Elérhető: <https://hu.wikipedia.org/wiki/SQL-injekci%C3%B3> (letöltve: 2024.03.17.).

Suleman Kazi & Adel Elmahdy (2023.10.17). Top Large Language Models (LLMs): GPT-4, LLaMA 2, Mistral 7B, ChatGPT, and More (2023). Elérhető: <https://vectara.com/blog/top-large-language-models-llms-gpt-4-llama-gato-bloom-and-when-to-choose-one-over-the-other/#gpt4> (letöltve: 2024.02.12.).

Sunil Potti (2023.04.25.). Supercharging security with generative AI. Elérhető: <https://cloud.google.com/blog/products/identity-security/rsa-google-cloud-security-ai-workbench-generative-ai> (letöltve: 2024.03.05.).

Szavak jelentése. Nonprofitmedia.hu. Elérhető: <https://nonprofitmedia.hu/multimodalis> (letöltve: 2024.02.15.).

Texttract. Elérhető: <https://pypi.org/project/texttract/> (letöltve: 2024.02.17.).

Together.ai. Elérhető: <https://api.together.xyz/> (letöltve: 2024.03.25.).

Trellix. Elérhető: <https://www.trellix.com/about/newsroom/stories/research/> (letöltve: 2024.01.14.).

Virginia Adams, Zenodia Charpy and Yi Dong (2022.07.12.). Adapting P-Tuning to Solve Non-English Downstream Tasks. Elérhető: www.developer.nvidia.com/blog/adapting-p-tuning-to-solve-non-english-downstream-tasks/ (letöltve: 2024.02.07.).

VirtualBox. Elérhető: <https://www.virtualbox.org/> (letöltve: 2024.01.15).

Virustotal. Elérhető: <https://www.virustotal.com/> (Letöltve:2024.01.08.).

VMware Workstation 17 Player. Elérhető: <https://www.vmware.com/nordics/products/workstation-player/workstation-player-evaluation.html> (Utolsó letöltés:2024.01.14.).

Vulnerability and Exploit Database. Elérhető: <https://www.rapid7.com/db> (letöltve: 2024.01.05.).

Vulnerability Database (VulDB). Elérhető: <https://vuldb.com> (letöltve: 2024.01.05.).

We Leak Info — Data Breach Search Engine. Elérhető: <https://weleakinfo.io/> (letöltve: 2024.01.06.).

What is business email compromise (BEC)? <https://www.microsoft.com/en-us/security/business/security-101/what-is-business-email-compromise-bec> (letöltve: 2024.03.12.).

who.is Elérhető: <https://who.is/> (letöltve: 2024.02.17.).

Yandex YaLM-100B. Elérhető: <https://github.com/yandex/YaLM-100B> (letöltve: 2024.02.12.).

Zone-h unrestricted information. Elérhető: <https://www.zone-h.org/> (letöltve: 2024.01.09.).

国家信息安全漏洞平台. (kínai – magyar fordításban: Nemzeti Információbiztonsági Sebezhetőségi Platform) Elérhető: <https://www.cnnvd.org.cn> (letöltve: 2024.01.05.).