HADMÉRNÖK
VÉDELEMINFORMATIKA
20. évfolyam (2025) 3. szám 49–70.
DOI: 10.32567/hm.2025.3.3

Bottyán Sándor[1]

Nagy nyelvi modellek az internetes nyílt forrású információgyűjtésben

Alkalmazások és védelmi stratégiák

Large Language Models in Online Open-Source Intelligence

Applications and Defensive Strategies

Absztrakt

A tanulmány az internetalapú nyílt forrású információgyűjtésben (open source intelligence, OSINT) megjelenő nagy nyelvi modelleket (large language models, LLMs) és az ezzel kapcsolatos kihívásokat vizsgálja. Bemutatja az internetalapú OSINT-tevékenység forráskörnyezetét és alkalmazási területeit, továbbá az LLM-ek szerepét ezen a területen. A kutatás fókuszában két olyan alkalmazás (LLM OSINT PoC és InfoHound) áll, amelyek kifejezetten nagy nyelvi modelleket integráltak internetes nyílt forrású információgyűjtésre tervezett keretrendszerekbe. A szerző összehasonlítás keretében vizsgálja a rendszerek architekturális jellemzőit és utasításaikat, elemzi továbbá azok hatékonyságát és fenyegetéseit. A tanulmány eredményei között javaslatokat találunk a keretrendszerek képességeivel szembeni olyan védelmi stratégiákra, amelyek mérsékelhetik az LLM-ek segítségével végzett OSINT eredményességét.

Kulcsszavak: OSINT, LLM, információgyűjtés, mesterséges intelligencia

[1] Doktori hallgató, Nemzeti Közszolgálati Egyetem Hadtudományi és Honvédtisztképző Kar Katonai Műszaki Doktori Iskola.

Abstract

This study examines the emergence of large language models (LLMs) in internet-based open source intelligence (OSINT) and the associated challenges. It presents the source environment and application domains of internet-based OSINT, as well as the role of LLMs in this area. The research focuses on two applications (LLM OSINT PoC and InfoHound) that specifically integrate large language models into frameworks designed for internet-based open source intelligence. The author analyses and compares the architectural characteristics and prompts of these projects, while also assessing their efficiencies and threats. The study concludes with proposals for defensive strategies against the capabilities of such frameworks, aiming to mitigate the effectiveness of LLM-supported OSINT operations.

Keywords: OSINT, LLM, intelligence, artificial intelligence

Bevezetés

Napjainkban a nagy nyelvi modelleket *(large language models,* LLM*)* alkalmazó interaktív csevegőbotok (például *ChatGPT, Gemini, DeepSeek)* megjelenése a társadalmi és a gazdasági területeken egyaránt jelentős hatást váltott ki. A jelenség körüli társadalmi népszerűség oka valószínűsíthetően a széles témakörben emberszerű interakciót prezentálni képes alkalmazások megjelenése. A gazdasági felkapottság pedig abban feltételezhető, hogy a transzformátoralapú modellek számos eljárást képesek potenciálisan gyorsítani, részben vagy egészben automatizálni, különösen az adatfeldolgozás, a kommunikáció és döntéstámogatás területén. E két perspektívából a nagy nyelvi modellek körüli diskurzus élénk, ugyanakkor polarizált is. Míg a társadalmi vonatkozásban számos szabályozási, etikai és filozófiai kérdés vetődik fel a témában, a gazdasági területen az LLM-ek alkalmazása új eredményeket és hatékonyabb folyamatokat ígér. Mindazonáltal az ilyen rendszerek széles körű elterjedése mindkét területre egyfajta veszélyt is jelent, így ezek megértése és előrejelzése akár tudományosan is értelmezhető feladat. Különösen figyelemre méltó kockázati szegmens a nagy nyelvi modellek alkalmazása a nyílt forrású információgyűjtésben (*open source intelligence,* OSINT). A nagy nyelvi modellek tanítóanyagát valós, időszerű, hiteles és nyers adatokból állítják elő, továbbá ugyanezek az adatok az OSINT-tevékenység forrásai is. Elgondolkodtató, hogy ez az unió hogyan gyakorol kölcsönhatást mindkét szereplőre, így hát ez az elméleti kérdés határozottan megfogalmazható néhány tudományos feltételezésben is.

Kutatási célok és hipotézisek

Jelen tanulmány alapvető célja, hogy áttekintse az LLM-ek alkalmazását az interneten végzett nyílt forrású információgyűjtésben, valamint speciális célkitűzése azon OSINT-LLM keretrendszerek részletes elemzése, amelyek hatékonyan képesek ötvözni

a hagyományos OSINT-eszközöket a nagy nyelvi modellek képességeivel. Ennek érdekében a kutatás során három fő hipotézist vizsgálunk:

H1: Feltételezhető, hogy a nagy nyelvi modelleket már aktívan alkalmazzák az internetalapú nyílt forrású információgyűjtés (OSINT) területén.

H2: Feltételezhető, hogy a nagy nyelvi modellek internetalapú OSINT-célú alkalmazása esetén azok új típusú, LLM-specifikus kiberbiztonsági fenyegetést jelentenek.

H3: Amennyiben a H1 és H2 hipotézisek igaznak bizonyulnak, feltételezhetően azonosíthatók olyan védelmi stratégiák, amelyek kifejezetten az LLM-ekkel végzett internetalapú OSINT-tevékenységekhez kapcsolódó fenyegetések megelőzésére és kezelésére alkalmasak.

A fenti célok elérése érdekében a hipotézisek mentén többrétű módszertant alkalmazunk. Az OSINT-témakör szükségszerű leíró feldolgozása mellett az internetes tartalomkutatásban (leginkább szürkeirodalomnak tekinthető esettanulmányok alapján) azonosított LLM-ágensek technológiai rendszerelemzése (architektúra és utasításkészlet) és ezek összehasonlításának leírása történik meg. A vizsgálathoz szükséges adatokat egy korábbi, a 2024. év második félévében végzett szekunder kutatás szolgáltatta. Ennek során kulcsszóalapú (kulcsszavak: OSINT, LLM, *open source intelligence,* OSINT-AI) kereséssel weboldalakról, dokumentációkból, GitHub oldalakról származó olyan projektek, szoftverek, keretrendszerek azonosítása ment végbe, amelyek az interneten történő nyílt forrású információgyűjtéshez nagy nyelvi modell képességet alkalmaznak architektúrájukban. Keresési követelményként lett megfogalmazva az OSINT-LLM funkcionalitás részletekbe menő megismerhetősége, a forráskódok elérhetőségén alapuló architektúra- és utasításvizsgálat. A tanulmány az összeállított információk szintézisével megállapítottak mellett a védelmi lehetőségek meghatározásával és predikciók alkotásával próbál hozzájárulni az esetleges új fenyegetésekből fakadó kockázatok kezeléséhez.

Az internetes nyílt forrású információgyűjtés

Az információgyűjtés konvencionálisan a különféle hírszerző szervek feladata volt,[2] így az OSINT is kizárólag nemzetbiztonsági és katonai hírszerzési tevékenységként jelent meg kezdetben. Ezt igazolja az Amerikai Egyesült Államok katonai fogalomtára is, amely alapján a teljes spektrumú hírszerzési tevékenység (*all-source intelligence,* ASI) három fő (úgymond hagyományos) kategóriából áll: emberi ügynökökön alapuló hírszerzés (*human intelligence,* HUMINT), technikai eszközökkel támogatott hírszerzés (*technological intelligence,* TECHINT), valamint a nyílt forrásokból történő információgyűjtés (OSINT).[3] Az elmúlt évtizedek során mindazonáltal jelentős változások következtek be a hírszerzés forráskörnyezetében és annak művelői körében egyaránt. Egyrészről a digitalizáció robbanásszerű fejlődésének köszönhetően a kibertérben található adatok mennyisége drámai módon megnövekedett, ami költséghatékonyabban feldolgozható, részben nyíltan elérhető információforrásként szolgál.

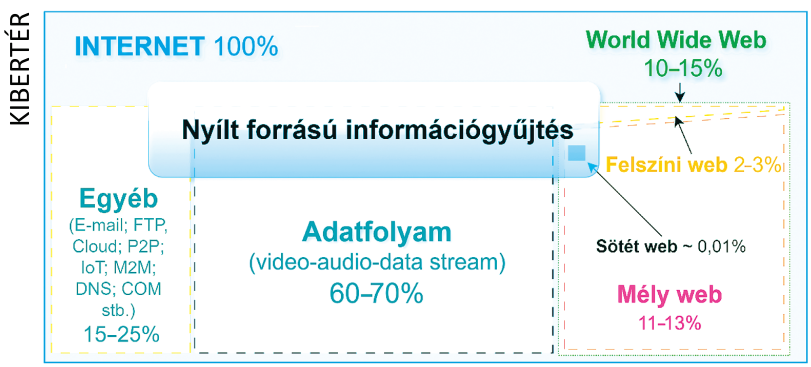
[2] Vadász 2018: 64.
[3] Department of Defense 2010.

Másrészt ez a jelenség létrehozta azt a változást, amely során a hírszerző tevékenység eredményei iránt érdeklődő entitások köre is bővült, ezáltal természetesen a korábbi hagyományosnak mondható cél és módszertan is megváltozott.

Az utóbbi időszakban az OSINT már bőven túllépett az állami alkalmazás privilégiumán és a versenyszférában is egyre jelentősebb szerepet tölt be. A gazdasági szervezetek partner-, konkurencia- és beszállítóelemzésre, az újságírók tényfeltáró munkára, a pénzügyi intézetek ügyfél-azonosításra és kockázatelemzésre használják, e mellett a kiberbiztonsági vizsgálatok egyik meghatározó biztonsági eljárásává is vált.[4] E többrétű folyamat céljának, feladatainak, eszközrendszerének és módszertanának pontos meghatározása nehéz feladat, kiváltképp hogy annak határa az adatvédelem szempontjából különösen elmosódott.[5] Bár az OSINT forrásai változatosak lehetnek (például nyomtatott média, nyílt rendezvények, közvetlen megfigyelés stb.), jelen tanulmány kizárólag az internetes nyílt forrású információgyűjtés területén tájékozódik. Ez a tevékenység lényegében az interneten nyíltan és bárki számára elérhető, transzparens és legitim módszertannal megszerzett adatok összegyűjtését, előre meghatározott cél szerinti elemzését, értékelését (együttesen feldolgozását) és továbbítását jelenti.

Az internetalapú OSINT forrásai

Az internetes információgyűjtés kapcsán lényeges megemlíteni, hogy az internet ismert információforrásait rétegekre tagolhatjuk, és megbecsülhetjük azok egymáshoz viszonyított terjedelmének arányát (1. ábra).



1. ábra: Az internet rétegeinek becsült összetétele és a nyílt forrású információgyűjtés viszonya
Forrás: a szerző szerkesztése

[4] Precognox 2022.

[5] Az itt említett konfliktus alatt azt értjük, hogy nyílt forrású adatnak minősül a legális módon, vagyis jogszabályba nem ütközően megszerezhető adat. Az internetalapú OSINT adatforrásait tekintve az azokkal szemben felállított legfőbb kritérium az, hogy nyíltak legyenek, tehát bárki számára elérhetők. Ugyanakkor az eltulajdonított, majd nyilvánosan közzétett személyes vagy védett adat is idetartozik, amelyet ugyanakkor továbbra is véd az adatvédelmi jog.

Teljes internetes információgyűjtés alatt az interneten fellelhető bármely adatforrás felhasználását értjük, amelybe természetesen beletartozik a teljes világháló (*world wide web,* www), valamint annak mélyebb rétegei is. A legszélesebb körben ismert felszíni web *(surface web)* a nyilvános és könnyen hozzáférhető weboldalak csoportja, amely technikailag a legkönnyebben elérhető OSINT-forrás. A mély web *(deep web)* tartalmai hagyományos keresőmotorokkal már nem érhetők el. Ez leginkább autentikációt követően elérhető weboldalak tartománya, például vállalati intranetek,[6] privát vagy előfizetéses tartalmak, jelszóval védett fórumok és a közösségi oldalak csoportjai sorolhatók ide. A közösségimédia-platformok látszólag ellentmondanak a nyílt forrású információgyűjtés alapdefiníciójának, de ki kell emelni, hogy a legtöbb ilyen mély webes weboldalhoz a hozzáférés nyilvánosan elérhető. Ilyenkor technikailag *nyíltan hozzáférhető mély webtartalomról* beszélünk, ami OSINT-szempontból forrásnak minősítendő, csak a módszertan eltérő némileg. Azonban vannak korlátozott hozzáférésű (például előfizetéses weboldalak vagy szervezeti intranet) mély weboldalak, amelyek már nem tartoznak az OSINT forrásai körébe. A *deep web* egy kisebb tartománya a sötét web *(dark web)*, amely bár méretét tekintve elhanyagolható, de funkciója (főként illegális kereskedelmi tevékenységeknek vagy az anonimitás megőrzésére törekvő egyéb tartalmaknak ad helyet) okán közismert terület.[7] Ez a világháló egy olyan speciális része, amely kizárólag anonim hálózatokon (például Tor, I2P, Freenet) érhető el. Ezzel összefüggésben itt is ki kell térni a meghívásos rendszer keretében vagy költség ellenében hozzáférhető weboldalakra, amelyek feldolgozása szintén nem felel meg a nyílt forrású információgyűjtés módszertani követelményének.

Emellett az internet jelentős adatforgalmat generáló része az adatfolyam *(streaming),* amely olyan platformoknak ad otthont, mint a videószolgáltatások (például YouTube, Netflix stb.), az audió- vagy zeneszolgáltatások (például Spotify, online rádiók stb.), illetve az élő közvetítések (például Twitch). Továbbá az internet szintén önálló részébe soroljuk a kisebb szolgáltatásokat, amelyek a teljesség igénye nélkül a fájlátviteli (például FTP, SFTP stb.), az e-mail (például SMTP, IMAP stb.), a kommunikációs (például IRC, WhatsApp, Messenger), a felhőtárhely (például Google Drive, Onedrive stb.), a *peer-to-peer* (P2P-) szolgáltatások, valamint az IoT *(internet of things)* és az M2M *(machine-to-machine)* kommunikációs és hálózati szolgáltatások (például DNS, DHCP). Tekintve, hogy ez az információforrás-csoport számtalan forrástípust tartalmaz, az OSINT-forrásként való értelmezésük nem homogén. A nyíltan hozzáférhető streamingből kinyerhető adatok (például videók tartalma, képi információk, helyszínek, személyek, időpontok, kommentek, reakciók, metaadatok) megfelelnek e tekintetben. Ugyanakkor a privát elektronikus levelezések természetesen nem, de a nyíltan hozzáférhető e-mailek és archívumok tartalmaikkal és metaadataikkal viszont már a legitim források közé sorolhatók. A többi említett szolgáltatásnál (például FTP, IoT, P2P stb.) is azonos elveket kell használni a forrásként való validálás során.

[6] Az intranet egy zárt, belső számítógépes hálózat, amelyet egy szervezet használ az információk és erőforrások megosztására a munkatársai között.

[7] Rosengren 2022: 24.

Az internetalapú OSINT eszközei

Az internetes források vonatkozásában értelmezett OSINT-eszközök olyan célszoftverek, platformok, vagy egyéb informatikai megoldások (valamint a használatukkal megvalósított technikák), amelyeket az interneten található nyilvánosan hozzáférhető információk hatékony megszerzésére és feldolgozására (strukturálásra és elemzésére) fejlesztettek. Az ilyen eszközök leltárának az információbiztonsági indíttatásból létrejött OSINT Framework[8] platform ad otthont, de azok színes skáláját mutatja be a svájci i-Intelligence vállalat által készített *Open Source Intelligence Tools and Resources Handbook,* vagyis a nyílt forrású információgyűjtési eszközök és források kézikönyve is, amelyet elsődlegesen nyomozók és információbiztonsági szakemberek támogatására hoztak létre.[9]

Az internetalapú OSINT-eszközöket alkalmazási területük és funkciójuk szempontjából megkülönböztethetjük az alábbi csoportok szerint:

- adatgyűjtő, adatfeldolgozó és adatvizualizációs eszközök (például RSS-olvasók, *scraping*[10] eszközök, hálózatelemző és infografika-készítő szoftverek és szolgáltatások stb.);
- általános keresőmotorok, dokumentumkutató és dokumentumelemző eszközök (például piacvezető keresők, metaadat-keresők stb.);
- infokommunikációs felderítő eszközök (például informatikai hálózati információk felderítésére, sebezhetőségazonosításra alkalmazott szoftverek stb.);
- közösségimédia-elemző eszközök (például főbb közösségimédia-platformok, blogok, fórumok és Q&A[11] oldalak kereső- és elemzőeszközei stb.);
- multimédiás eszközök (például képkereső, arcfelismerő, videó- és hangelemző eszközök stb.);
- személykereső és -ellenőrző eszközök (például nevek, felhasználónevek, telefonszámok és e-mailek keresésére, nyilvántartások forrásfelhasználására fejlesztett eszközök stb.);
- térinformatikai és közlekedésvizsgáló eszközök (például online térképek, műholdfelvételek, repülő-, hajó- és járműkövető szoftverek stb.);
- üzleti célú eszközök (például cégadatbázisok, állás- és béradatbázisok kezelésére, versenytársak feltérképezésére, üzleti partnerek megbízhatóságának vizsgálatára készült alkalmazások és szolgáltatások stb.).

Az eszközökkel szemben megfelelőségi követelményeket is meghatározhatunk, bár meg kell jegyezni, hogy ezek leginkább csak elvi követelmények, hiszen a többnyire ingyenesen elérhető OSINT-alkalmazásokat – a maximális hatékonyság elérése okán – nem készítik fel az elméleti alapvetések maradéktalan teljesítésére. Ezek a következők:

[8] Bielska 2020.
[9] Bielska 2020; Bottyán 2024:16.
[10] Az *internet scraping* (vagy *web scraping*) olyan automatizált folyamat, amellyel weboldalakról adatokat gyűjtenek le strukturált formában.
[11] *Questions and answers* (kérdések és válaszok).

- kizárólag nyíltan, külön engedély vagy jogosultság nélkül elérhető adatot használhat fel forrásként;
- az eszköz által begyűjtött információk eredete visszakövethető és ellenőrizhető;
- az információkhoz való hozzáférés nem lehet sértő szándékú, tehát nem tartalmazhat etikailag kifogásolható technikai tevékenységet (például *brute-force,* tiltott *scraping* stb.);
- az eszköz információgyűjtési mechanizmusa összhangban van a hatályos adatvédelmi jogszabályokkal (például GDPR,[12] Infotv.[13]), és működése transzparens (átlátható és dokumentált).

Az eszközöknek széles körű alkalmazási területe van, többek között:

- katonai hírszerzés: hadászati, hadműveleti és harcászati támogatás (például információs műveletek azonosítása, saját katonai műveletek támogatása);
- kiberbiztonság: saját infrastruktúra sérülékenységeinek felkutatása, fenyegetés-hírszerzés (*cyber threat intelligence,* CTI);
- nemzetbiztonsági hírszerzés: stratégiai (például geopolitikai) információgyűjtés, humanitárus válsághelyzetek követése, radikalizáció és terrorizmus-felderítés;
- nyomozó hatóságok és rendvédelmi szervek tevékenysége: gyanúsítottak, tanúk vagy eltűnt személyek azonosítása, kapcsolati hálók meghatározása, tartózkodási helyük megállapítása;
- oknyomozó újságírói kutatás: információk keresése, háttér-információk ellenőrzése;
- piackutatás, üzleti intelligencia: versenytársak elemzése, trendek felkutatása;
- politikai tevékenység: politikai ellenfelek elemzése, választói hangulat monitorozása, reputációmenedzsment, narratívatámogatás, civil szerveződések monitorozása.

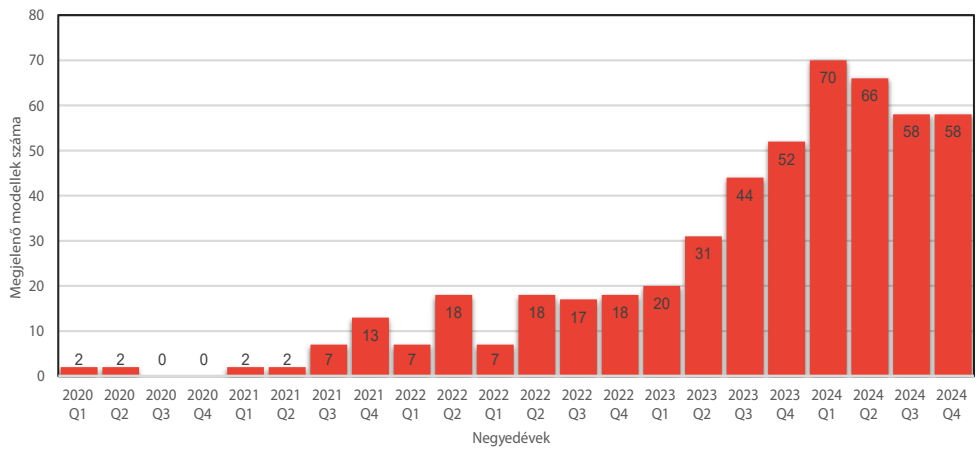
A nagy nyelvi modellek

A legtöbb ezzel a témával foglalkozó tudományos értekezés a nagy nyelvi modellek népszerűségének alapkövét egy 2017-ben történt eseményhez, vagyis a Google által fejlesztett transzformátormodell és annak figyelemmechanizmusának bemutatásához köti. Ez az esemény jelentős lendületet hozott a természetes nyelvi feldolgozás (*natural language processing,* NLP) tudományos területén, csakhogy a társadalmi és gazdasági rétegeket egyaránt megmozgató jelenség a ChatGPT megjelenése volt 2022 novemberében. A hétköznapi emberek nyilvánosan ekkor léptek kapcsolatba egy kifejezetten emberi csevegésre finomhangolt nagy nyelvi modell interaktív felületével, megismerhették annak megdöbbentő képességeit. A megjelenéssel járó popularitás

[12] Az Európai Parlament és a Tanács 2016. április 27-i (EU) 2016/679 rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK irányelv hatályon kívül helyezéséről (általános adatvédelmi rendelet).

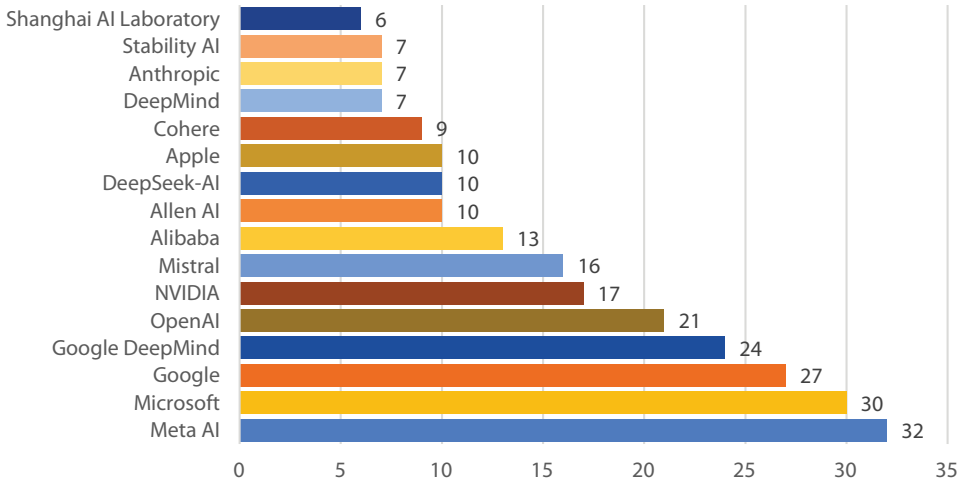
[13] 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról.

olyan globális gazdasági hatásokat eredményezett, amelyeknek köszönhetően az óriásvállalatok jelentős tőkét csoportosítottak át a mesterséges intelligencia, különösen az LLM fejlesztésébe. E mellett számos infokommunikációs ipari szervezet gondolta újra innovációs politikáját és dinamikusabban fordult az AI- (*artificial intelligence,* vagyis mesterséges intelligencia) fejlesztések irányába. Természetesen az olyan AI-terület, mint a gépi tanulás (*machine learning,* ML) továbbra is dinamikusan fejlődő megközelítés maradt, de a nagy nyelvi modellek egy negyedévre jutó megjelenéseiből egyértelműen megállapítható a közel négyszeres növekedés 2022 utolsó (Q4) és 2024 első negyedéve (Q1) között. Ez a magas számú LLM-eredménytermék-növekedés valószínűsíthetően a 2022 utolsó negyedévében (Q4) történt tőkeinjektálással meginduló, kifejezetten gyors ciklusú kutatási-fejlesztési (K+F-) folyamatok eredménye (2. ábra).



2. ábra: A nagy nyelvi modellek megjelenése 2020–2024 között
Forrás: a szerző szerkesztése a lifearchitect.ai 2025. 02. 10-ei adatai alapján

A növekedés azért is figyelemre méltó, mert a nagy nyelvi modellek fejlesztése kifejezetten összetett feladat. Szignifikáns mértékű interdiszciplináris humán szakértelmet igényel, hiszen több kutatói és mérnöki szerepkör (például adatmérnökök, nyelvészek, NLP-szakértők, szoftverfejlesztők, adatvédelmi és jogi szakértők stb.) együttes alkalmazása, valamint számos monumentális információs és kommunikációs technológiai (*information and communication technology,* ICT) erőforrás megléte indokolt. Ezeket a jelentős mennyiségű és szerteágazó szükségleteket csak a világviszonylatban legnagyobb szoftver- és technológiai ipari, kutatási fejlesztési képességekkel is rendelkező óriásvállalatok tudják biztosítani. Ilyen piaci szereplők az Amerikai Egyesült Államokban (Apple, NVIDIA, OpenAI, Google, Microsoft, Meta AI, Allen AI, Anthropic), Kínában (Shanghai AI Laboratory, Alibaba, DeepSeek-AI), az Egyesült Királyságban (DeepMind, Stability AI) és Kanadában (Cohere) vannak, az általuk 2020–2024 között létrehozott nagy nyelvi modellek számát a 3. ábra mutatja.



3. ábra: A legnagyobb technológiai nagyvállalatok által bemutatott LLM-megjelenések száma 2020–2024 között

Forrás: a szerző szerkesztése a lifearchitect.ai 2025. 02. 10-ei adatai alapján

Bár a nagy nyelvi modellek népszerűsége töretlen, valószínűsíthető, hogy a jövőben a modelltípusok száma csökkenni fog, és bár alacsonyabb számban, de eredményesebb architektúrák megjelenésére lehet számítani. A modellek az NLP területén elért eredményeikkel számottevő hatást gyakorolnak az élet számos területére, hiszen az emberszerű automatizáltság növelésére mindenhol igény van. Ezt a tényezőt használják fel egyesek a nyílt forrású információgyűjtés területén is, de mielőtt rátérnénk a gyakorlati alkalmazásra, röviden bemutatjuk az LLM-modellek jellemzőit.

A nagy nyelvi modellek csoportosítása

A nagy nyelvi modelleket elérhetőségük szerint két csoportba sorolhatjuk, vannak nyílt forráskódú *(open-source)* és zárt (esetleg részben zárt rendszerként működő, felhőalapú) úgynevezett *hostolt* modellek. Az előbbi csoportba tartozó LLM-ek esetében a modell forráskódját, a tanítási adatokat, súlyokat, dokumentációkat nyilvánosan – részben vagy egészben – bárki számára elérhetővé teszik, ennek köszönhetően a fejlesztők és kutatók szabadon módosíthatják azokat. A zárt vagy *hostolt* modellek esetében általában a fejlesztést végrehajtó nagyvállalat üzemelteti a modellt, és előfizetés (legtöbbször *pay as you go*[14] fizetési rendszer alkalmazásával) ellenében, általában API-n[15] vagy felhőszolgáltatáson keresztül, valamint interaktív webfelületen biztosítja a hozzáférést. Ez esetben a fejlesztő nem osztja meg a modell forráskódját

[14] A *pay as you go* (PAYG) egy fizetési modell, amelyben a felhasználó kizárólag a termék használatáért fizet. Leginkább technológiai szolgáltatásoknál (például felhőalapú szolgáltatások, API-k, mesterségesintelligencia-modellek használata stb.) alkalmazzák.

[15] *Application programming interface* (alkalmazásprogramozási interfész).

és a hozzá tartozó egyéb információkat, ellenben korlátozott lehetőséget biztosít a modell bizonyos képességeinek használatához. Általánosságban elmondhatjuk, hogy a nyíltan elérhető modellek képességei alulmúlják a *hostolt* modellekét, a különböző *benchmarkok*[16] listáját általában a legnagyobb óriásvállalatok kutatóközpontjai által létrehozott csúcsmodellek vezetik.

Az LLM-ek valós idejű információgyűjtési képessége

Az információkeresés az NLP egyik olyan területe, amelyben az LLM-ek elismerten kiváló teljesítményt nyújtanak. A kontextuális jellemzőkkel együttesen beágyazott vektorértékeknek köszönhetően hatékonyan képesek a szöveges tartalmak közötti kapcsolatok azonosítására, így képesek a keresések során megfelelő információt szolgáltatni. Ugyanakkor az LLM-ek tanítóadatai múltbéliek, így bizonyos kérdésekben nem képesek aktuális (időszerű) információkat szolgáltatni, ami ellenben lényegi jellemzője és követelménye is az információgyűjtésnek. Az utóbbi években, a nagy nyelvi modellek elterjedésével megjelent az igény arra, hogy annak kontextuális képességeit valamilyen módon internetes keresési funkcióként is alkalmazzák. A közismert ChatGPT-t üzemeltető OpenAI először 2023 nyarán egyfajta béta teszt jelleggel a Bing keresőmotorjának integrálásával lehetővé tette a modellen keresztüli internetes keresés lehetőségét. Az eseményt követően probléma merült fel, mert a Bing modullal végzett LLM-keresések képesek voltak hozzáférni bizonyos híroldalak alapvetően felhasználói előfizetéssel elérhető szöveges tartalmaihoz, majd az információt ingyen megosztani a ChatGPT felhasználóival.[17] Az OpenAI ekkor eltávolította a Bing modult a *chatbotból,* majd a hiba javítását követően újra visszahelyezte 2023 őszén.[18] A keresőfunkció néhány *ChatGPT release*[19] esetében a mai napig elérhető, de alkalmazása közben észlelhetjük a *chatbot* szigorú energiagazdálkodási konfigurálásának következményét, miszerint a keresési munkamenetet gyorsan lefuttatja, és minimális információt gyűjtve tér vissza a válasszal. A modellek által megszerzett információk mennyisége csekély, további feldolgozásra nem alkalmas, így önálló OSINT-eszközként valószínűleg nem tekinthetők hatékonynak.[20]

A kutatás

A kutatási cél érdekében végzett keresés során mindösszesen két olyan keretrendszert sikerült azonosítani, amelyek az előzetesen meghatározott követelményeknek evidenciákkal alátámaszthatóan eleget tesznek. A következőkben e projekteket mutatom be,

[16] A *benchmark* kifejezés általában egy standard teszt vagy referenciaérték, amelyet teljesítmény mérésére használnak.
[17] Okemwa 2023; Holupredictions 2023.
[18] Rogers 2023.
[19] A *release* kifejezés az adott modell új változatának kiadására, vagyis megjelenésére utal.
[20] Bottyán 2024: 54.

megemlítve architekturális jellemzőiket, kiemelten vizsgálva LLM-komponenseiket és -utasításaikat *(prompt)*.

LLM OSINT

Az ingyenesen használható LLM OSINT komplex keretrendszer, amelynek célja, hogy az interneten elérhető források felhasználásával minél többet tudjon meg egy adott személyről, szervezetről, helyről, dologról stb. Tekintve, hogy a projekt nem kapott sajátos elnevezést és csak *proof of concept* (PoC),[21] a következőkben a könnyebb azonosíthatóság okán LLM OSINT PoC-ként hivatkozunk rá. Ez a keretrendszer-architektúra több szálon, több típusú nagy nyelvi modellt vesz igénybe az információgyűjtés során. A koncepció már a kezdeti web ágensében (*initial web agent,* IWA) egy GPT-4.0-t kér fel az internetes nyílt információgyűjtésre, egy kifejezetten OSINT-ügynök persona[22] használatával (4. ábra).[23]

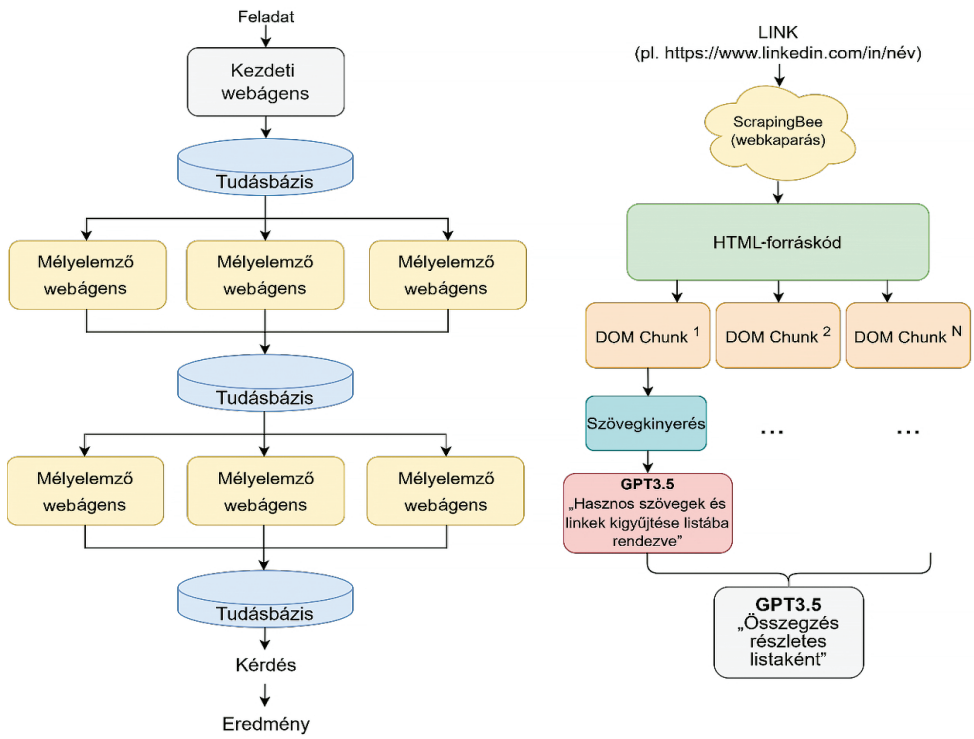
> PREFIX = Te egy automatizált OSINT-ügynök vagy, aki kiválóan képes információt felkutatni személyekről, vállalatokról, helyekről és más témákról. A kérdés/gondolat/tevékenység/megfigyelés formátumot használod a keresések során. Ha kimerítetted a rendelkezésre álló forrásokat, NE ÁLLJ MEG, végezz újabb keresést bármilyen olyan részletről, amelyről további információt szerezhetsz! Válaszold meg a felhasználó kérdéseit az internet átvizsgálásával és légy nagyon alapos!
>
> Tippek:
> Mindig olvasd el a hivatkozott weboldalakon szereplő tartalmakat, amelyek további információt nyújthatnak a kérdés megválaszolásához!
> A weboldalak tartalmát rekurzívan használd fel újabb információk felfedezéséhez!
> Kövesd a linkeket, amelyek további adatokkal szolgálhatnak!

4. ábra: Az LLM OSINT PoC IWA által futtatott angol nyelvű prompt magyar nyelvre fordítva
Forrás: a szerző szerkesztése és fordítása a https://github.com/sshh12/llm_osint oldalon található információk alapján

Az IWA által visszajuttatott információkból a mélyelemző webes ügynökök (*deep dive web agents,* DDWA) Serper[24] API felhasználásával olyan internetes hivatkozásokat kutatnak fel, amelyek tartalmazzák a megjelölt információkat. Ezt követően a weboldalak forráskódjait Scrapingbee API segítségével (ezt nevezik *web scrapingnek*[25])

[21] A *proof of concept* kifejezés jelentése: koncepció igazolása vagy próbaüzem, műszaki és üzleti szférában is használatos fogalom.
[22] A kifejezés a nagy nyelvi modellek esetében arra a mesterséges szerepre vagy személyiségre utal, amelyet a modell vesz fel a felhasználóval folytatott kommunikáció során. Ez a szerep meghatározhatja a beszédstílust, a szóhasználatot, az ismeretek körét, sőt azt is, hogy a rendszer milyen nézőpontból válaszoljon.
[23] Bottyán 2024: 62.
[24] A Serper egy gyors és költséghatékony Google-kereső, amely lehetővé teszi a felhasználók számára, hogy automatizált módon hozzáférjenek a Google keresési eredményeihez.
[25] A *web scraping* folyamán – emberi olvasásra alkalmas, de közvetlen gépi feldolgozásra alkalmatlan – adatot gyűjtünk weboldalakról.

további feldolgozás céljából letöltik. A keretrendszer a kinyert adatokat (HTML-forráskód) – azok méretbeli nagysága miatt – tovább tisztítja olyan formátumú DOM-*(document object model)*[26] *chunkokra,*[27] amelyek már alkalmasak a GPT-3.5 általi reprezentálásra.[28] A modell ebből a nagyszöveges *promptból*[29] kivonatolja a leghasznosabb adatokat egészen addig, amíg a GPT-3.5 tokenizálási korlátját[30] el nem éri, majd ez addig ismétlődik, amíg a legutolsó műveletnél elfogy a feldolgozandó tartalom. Az eredményként kinyert részletek újfent visszacsatolódnak a GPT-3.5-be, és az adatok összegzése történik meg végeredményként (5. ábra).[31]



5. ábra: Bal oldalt az OSINT-LLM PoC web ügynökeinek működése, jobb oldalt a web scraping adatfeldolgozás szemléltetése

Forrás: a szerző szerkesztése és fordítása a https://github.com/sshh12/llm_osint alapján

[26] Platformfüggetlen, fa struktúrájú programozási modell, amely lehetővé teszi a HTML- vagy XML-dokumentumok elemeinek dinamikus elérését és módosítását.

[27] A *chunk* egy adott hosszúságú szövegrészlet (pl. mondat, bekezdés vagy fix hosszúságú tokensorozat), amelyet a modell egyszerre képes feldolgozni.

[28] A nagy nyelvi modellek során a reprezentálás azt jelenti, hogy a nyelvi elemeket (szavakat, mondatokat) olyan matematikai struktúrákká, például többdimenziójú vektorokká alakítják, amelyeken a modell képes elvégezni a matematikai műveleteket.

[29] A *prompt* egy szöveges bemenet vagy utasítás, amelyet a felhasználó megad a modellnek annak érdekében, hogy a modell az adott szövegre reagálva választ, tartalmat vagy információt generáljon.

[30] *Token limit.* A nagy nyelvi modellek esetében azt a maximális mennyiségű tokenre (a szöveg egyik feldolgozási egysége) vonatkozó határt jelöli, amelyet a modell egy adott bemeneten belül kezelni tud.

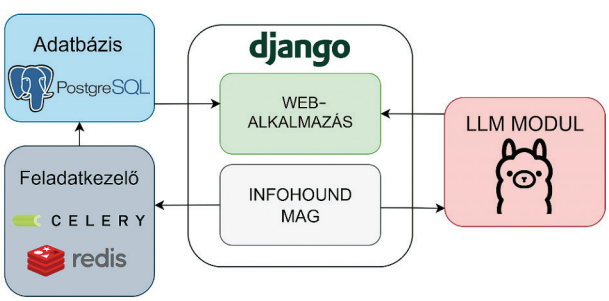
[31] Bottyán 2024: 63.

Az összegyűjtött információkat LLM-utasításban felhasználva a *hostolt* GPT – a korábbi tanítóanyagából kikövetkeztetve – képes akár a célszemély pszichológiai jellemzőit is (például személyiségvonások, motivációs tényezők, viselkedési mintázatok) feltárni.

InfoHound

Az InfoHound egy információgyűjtő és -feldolgozó OSINT-eszköz, amely képes jelentős mennyiségű adatot feltárni az előzetesen megcélzott domainről. Az i2CAT[32] által kifejlesztett eszköz leginkább személyekre vonatkozó információkat (például személyneveket, e-mail-címeket, felhasználóneveket), valamint aldomaineket, fájlokat, URL-[33] címeket gyűjt össze. Az adatvizualizáció mellett képes kimutatást készíteni a begyűjtött URL-ek, a felfedezett aldomainek, a felfedezett személyek, az észlelt fájlok, az észlelt e-mailek számáról.[34]

Architektúráját tekintve nyílt forráskódú modulokból áll (6. ábra). Egy Django keretrendszeren fejlesztett webes alkalmazás szolgáltatja a felhasználófelületet, a webalkalmazáson indított adatgyűjtési háttérfolyamatok futtatása Celery-n történik meg, amely a gyűjtés eredményeit a PostgreSQL adatbázisba továbbítja. A végleges adatokat a webalkalmazás jeleníti meg. A webalkalmazásból indított elemző feladatok során a feladatok munkafolyamatát szintén a Celery végzi, a Redis[35] szoftver a kommunikációért felel. Specifikuma, hogy több Python könyvtár alapú OSINT-eszköz (például Beautifulsoup, DNSpython, E-mail-validator, Holehe, Shodan stb.) van integrálva a keretrendszerbe. Az InfoHound önmagában is kifinomult OSINT-eszköz, de legújabb verziója már LLM-támogatással is rendelkezik, méghozzá az Ollama felhasználásával.[36]



6. ábra: az InfoHound keretrendszerének kialakítása
Forrás: a szerző szerkesztése és fordítása a GitHub Fundacio-i2CAT/InfoHound 2024 alapján

[32] Fundació Privada Internet i Innovació Digital a Catalunya.
[33] *Uniform resource locator* (egységes erőforrás-azonosító) rövidítése. Az interneten fellelhető erőforrások (pl. weboldalak, képek vagy videók) elérési útvonalát és címét határozza meg.
[34] CyberRaya 2023.
[35] A Redis nyílt forráskódú, memóriában futó kulcsértékalapú adatbázis, amelyet elsősorban gyors adatkezelésre és gyorsítótárazásra használnak.
[36] Bottyán 2024: 67.

Az Ollama egy nyílt forráskódú keretrendszer, amely biztosítja a szintén nyílt forráskódú LLM-ek futtatását felhő- és internetszolgáltatás nélkül saját számítógépen (vagy úgymond *on-premise*[37] megoldásként). Praktikus eszköz azok számára, akik a klasszikus operációs rendszereken (macOS, Linux, Windows) kívánnak nagy nyelvi modelleket (például Llama 2, Mistral, Dolphin Phi, Phi-2, Neural Chat, Starling, Code Llama stb.) futtatni. Az InfoHound legújabb verziója alapértelmezetten az Ollama által futatott Llama 2 nagy nyelvi modellt alkalmazza – az OSINT-eszközök által összegyűjtött adatok felhasználásával – a szervezeti struktúra meghatározásához vagy a személyek szervezeten belüli szerepének azonosításához.

```
summarize_prompt = "Foglalja össze a személy foglalkozását mindössze 150 szóban a következő adatok alapján:"
        raw_data = entry.raw_metadata
        print ("Az AI-alapú profilelemzés végrehajtása:" + entry.name)
        entry.ocupation_summary = ollama.ollama_flexible_prompt(summarize_prompt + raw_data)
        print ("Összegzés: " +entry.ocupation_summary)
                entry.save()
```

7. ábra: Az InfoHound azon forráskódrészlete, amely a Llama 2 részére kiadott utasítást tartalmazza (magyarra fordítva)

Forrás: a szerző szerkesztése és fordítása a https://github.com/ollama/ollama alapján

Annak ellenére, hogy az InfoHound OSINT szempontból egy igen komplex rendszer, az LLM-komponenst utasító prompt (7. ábra) egyszerűnek mondható, pedig a témában már igen részletes utasításábrázolás is létezik.[38] Eszerint a fejlesztő nem törekedett az LLM-képességek hatékonyabb kihasználására, bár természetesen e tekintetben a Llama 2 nyelvi funkciói is lényegesen korlátozottabbak *hostolt* társainál.

A két vizsgált megközelítés között egyértelműen azonosíthatók megegyező és diverzív architekturális jellemzők is, ami alapján klasszifikálhatók az LLM-et integráló OSINT-architektúrák.

Az OSINT-LLM-keretrendszerek azonosított típusai és jellemzői

A kutatás során feltárt OSINT-LLM-integrációk elemzése alapján megállapítható, hogy bár azok architektúrája, funkciója és jellemzője némileg különbözik, a vizsgált keretrendszerek azonos alapvető komponensekből épülnek fel.

Az OSINT-LLM-keretrendszer fő elemei:

1. Internetes nyílt forrású adatgyűjtő komponens (OSINT-modul): ez az egység felelős a nagy mennyiségű, feldolgozásra szánt nyílt forrású adatok felkutatásáért

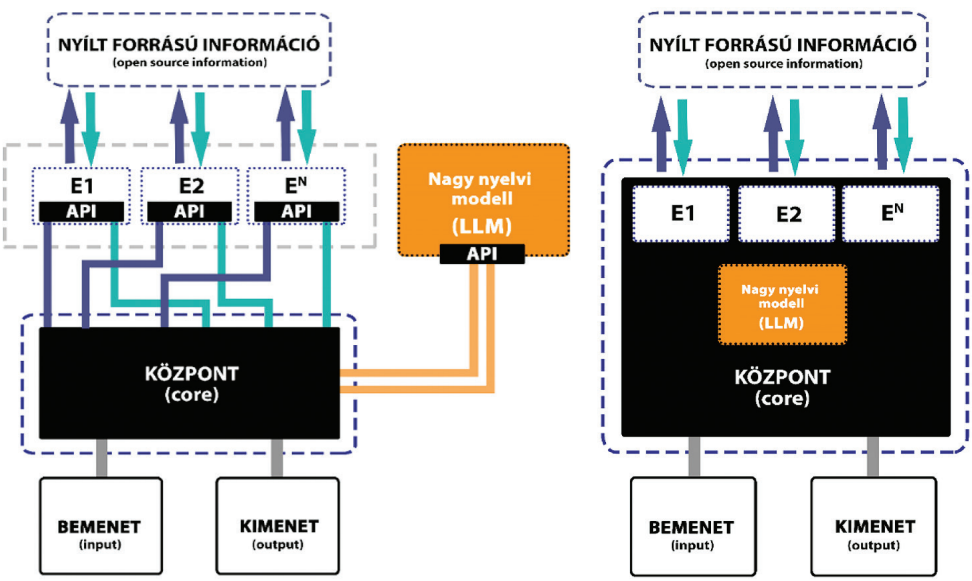
[37] A kifejezés az informatikában arra utal, hogy a szoftver, alkalmazás vagy adatbázis a vállalat/szervezet saját fizikai környezetében – például a cég saját szervertermében vagy adatközpontjában – fut és van telepítve, nem pedig felhőalapú szolgáltatásként külső szolgáltató infrastruktúráján.

[38] Černý 2024: 120.

és összegyűjtéséért. A modul külső szolgáltatók által biztosított megoldásokat, illetve szabadon elérhető, integrálható OSINT-eszközöket is tartalmazhat.

2. Nagy nyelvi modell (LLM-) modul: ez a komponens egy vagy több LLM-integrációt foglal magában, amely lehet külső szolgáltatón keresztül elérhető *hostolt* modell, vagy helyben telepített *(on-premise)* megoldás. Feladata az OSINT-modul által begyűjtött információk, elsősorban szöveges adatok feldolgozása és utasításalapú kivonatolása. Továbbá az API-n keresztül megszólított keresőfunkció alkalmazása esetén OSINT-eszközként is értelmezhető, valamint lehetőséget nyújt generatív feladatok elvégzésére is, például egy adott személy pszichológiai profiljának elkészítésére.
3. Központi vezérlő modul *(core):* ez az elem biztosítja az összeköttetést és az adatáramlást a rendszer különböző moduljai között, valamint irányítja azok működését. Kezeli a felhasználói bemeneteket (utasításokat), továbbítja és feldolgozza az adatokat, illetve biztosítja a felhasználó számára az eredmények megjelenítését. Ebbe a modulba integrálódnak azok az eszközök is, amelyek az OSINT- és LLM-modulok közötti adatkompatibilitást támogatják (például a webes adatgyűjtés során kinyert HTML-forráskód tisztításával).[39]

A keretrendszereken belül a modulok architekturális elhelyezkedése bizonyos mértékű aszimmetriát mutat, amely alapján két alapvetően különböző megközelítésű rendszerkialakítás állapítható meg: az egyik a külső szolgáltatás alapú *(third-party dependent),* míg a másik a lokálisan integrált modell (8. ábra).



8. ábra: Balról a külső szolgáltatás alapú, jobbról a helyi (on-premise) OSINT-LLM-architektúra-típusok
Forrás: a szerző szerkesztése

[39] Bottyán 2024: 74–75.

A szolgáltatásra épülő architektúra esetén a felhasználói bemenet – például egy adott személyre irányuló információs igény – először a központi egységhez érkezik. Ezt követően a rendszer a szükséges adatok lekérdezését különféle OSINT-eszközök API-interfészein keresztül hajtja végre (például Google Search, Shodan vagy ScrapingBee használatával).[40]

Az OSINT-eszközök visszajuttatják a nagyhalmazú adatot a központi egységbe, ahonnan viszont *LLM-prompt(ok)* kíséretében megküldik egy szintén külső szolgáltatású LLM-nek. Ezt követően a meghívott LLM-képesség (például elemzés, kivonatolás stb.) érvényesítése után az információ az utasítás szerinti tartalom formájában visszaérkezik kimenetként. A keretrendszer előnye, hogy skálázható, a kapcsolt szolgáltatások száma elméletben korlátlan.

Gyakorlati megközelítésben forrásigénye alacsony, hiszen a nagyobb számítási kapacitást igénylő feladatokat a kapcsolt szolgáltatások végzik. Ebből az okból kitettsége jelentős, alkalmazhatósága függ a szolgáltatások rendelkezésre állásától. Rejtve maradása végpontól végpontig nem értelmezhető, hiszen a behívott szolgáltatások megkapják a célpontra vonatkozó információkat. Erőssége a rendszernek, hogy az integrált szolgáltatások a piacon elérhető leghatékonyabbak, és kielégítő eredményt nyújtanak. Ugyanakkor problémásan felel meg az OSINT-tevékenységgel szembeni transzparenciakövetelménynek, hiszen a behívott szolgáltatók gazdasági érdekük alapján nem biztosítanak teljes és maradéktalan átláthatóságot az általuk szolgáltatott eszközökkel és eljárásaikkal összefüggésben.[41]

A másik csoport a helyi integrációban *(local integration)* létrehozott típus, amelyben nem találunk külső szolgáltatói kapcsolatot, és minden OSINT-LLM-funkció egy helyi keretrendszerbe építve jelenik meg (8. ábra jobb oldali kép). Az előző modellarchitektúrától eltérően itt az OSINT-eszköz és LLM-képesség önálló modulokként épülnek be, tehát nincs szolgáltató irányába megjelenő kitettség és költségvonzat, valamint biztosított a rejtve maradás, hiszen alapértelmezetten nincs harmadik félen keresztülmenő kommunikáció. Az integrált OSINT-modulok korlátozottabb hatékonysággal bírnak, továbbá a nyílt forráskódú LLM-ek sem a legeredményesebbek, ugyanakkor az üzemeltetési költségek is alacsonyabbak. Ez a keretrendszer szintén skálázható, hiszen az integrációk száma elméletben itt is korlátlan, azonban már megjelenik a rendelkezésre álló erőforrások okozta korlátozottság. Mivel a keretrendszer elemei integrálva vannak, így az OSINT-tevékenység transzparenciája is biztosítható. Annak okán, hogy megőrzi a használója anonimitását, sajnos a rosszindulatú tevékenységek végzésére is kiválóan alkalmas.[42]

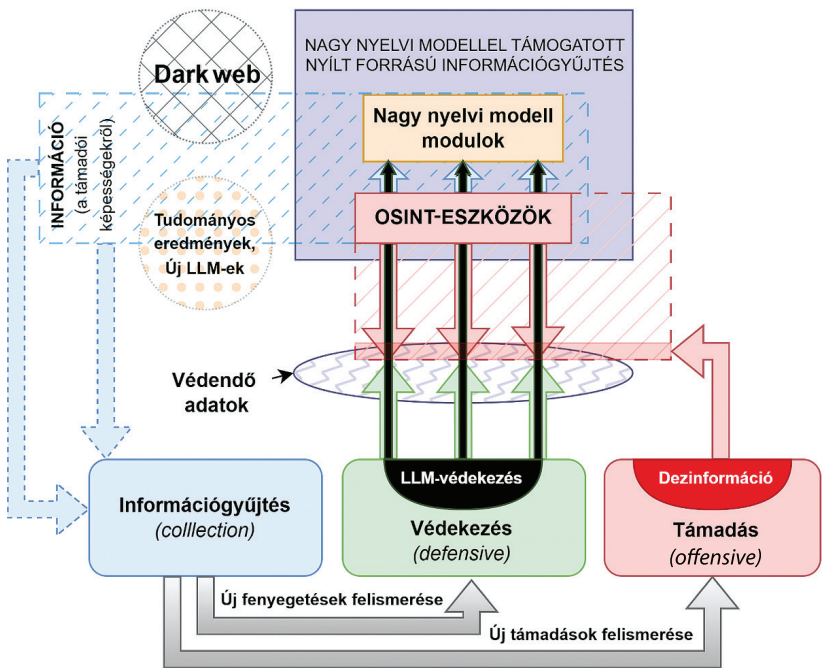
OSINT-LLM-tevékenység elhárítását támogató javaslatok

A korábban tett megállapításokra hivatkozva kijelenthetjük, hogy az OSINT-LLM-megoldások jelenlegi alkalmazási formái nem jelentenek kiemelkedően új fenyegetést

[40] Bottyán 2024: 75–76.
[41] Bottyán 2024: 74–76.
[42] Bottyán 2024: 76–78.

a nyílt forrású információszerzés elleni védekezés területén, ellenben az adatmennyiség kezelésében az eddigi korlátok többszörösére lesznek képesek. Az LLM-innovációnak köszönhetően az internetes nyílt forrású információgyűjtés valószínűbb kockázattá válik, így kiberbiztonsági szempontból felértékelődik annak kezelése. Egyetlen lehetőségünk az érzékeny adatok (például személyes adatok, üzletmenet-folytonosságot veszélyeztető adatkörök) eltávolítása vagy eltávolíttatása az aktív forrásokból, valamint a jövőbeni tudatosság gyakorlása mellett belenyugvás abba a felismerésbe, hogy az archívumokban továbbra is megtalálhatók korábban közzétett adataink.

Szervezeti szempontból jelentősebb a felelősség, hiszen munkavállalóink személyes és szervezetünk védett adataira egyaránt hat az LLM-OSINT-fenyegetés. Az eredményes védekezés legelső eleme az OSINT elleni módszerek és eljárások (például digitális lábnyom csökkentése) alkalmazása, amelyet ez a tanulmány részletekbe menően nem kíván tárgyalni. Helyette inkább azon speciális lehetőségeket kívánja azonosítani, amelyek felhasználása redukálhatja az LLM-mel támogatott OSINT hatékonyságát. Ezeket a klasszikus kémelhárítási *(counterintelligence)* kategóriák[43] szerint azonosítva tárgyaljuk és egyúttal a 9. ábrán szemléltetjük.



9. ábra: A nagy nyelvi modellel támogatott nyílt forrású információgyűjtés elleni védelmi stratégiák szemléltetése

Forrás: a szerző szerkesztése

[43] Lowenthal 2003: 263.

Információgyűjtés

Ez a típusú elhárítási forma információkat szerez az ellenfél hírszerzési képességeiről. Ebből következtetve az LLM-OSINT-megoldások megjelenésének rendszeres kutatása *(CTI feed)*[44] és elemzése hozzájárulhat a időszakos kockázat-felülvizsgálatok eredményességéhez. A témában releváns új tudományos publikációk, esettanulmányok, a nyílt forrású projektek vizsgálata és tesztelése, valamint a technológiai óriások által szolgáltatott LLM-képességek rendszeres elemzése megfelelő információt szolgáltat a fenyegetettségek jellemzőiről. A dark weben fellelhető információk, különösen az illegális piactereken áruba bocsátott LLM-technológiák (például FraudGPT vagy WormGPT[45]) elemzése rámutathat a legújabb fenyegetések indikátoraira, de ugyanez vonatkozik az OSINT-eszközökre is. Értékes információt lehet szerezni a felismert OSINT-tevékenységgel kapcsolatos esettanulmányok vizsgálatából is, kifejezetten az LLM-alkalmazás szemszögéből megvizsgálva a tapasztalatokat.

Defenzív elhárítás

A védekező jellegű elhárítás alapvető célja az ellenséges hírszerzési tevékenységek megakadályozása. Ebből a szemszögből célszerű az OSINT- és LLM-rendszerek gyengeségeinek és fenyegetéseinek együttes, védekezési szempontú elemzése és hasznosítása. Az alkalmazható eszközök között szerepelhet egy speciálisan kialakított *honeypot*[46] is, amely képes csökkenteni a nagy nyelvi modellek általi feldolgozás hatékonyságát. Ezen túlmenően a digitális lábnyomban (például weboldalak forráskódjaiban) elhelyezett *prompt injection*[47] sebezhetőséget kihasználó utasítások szintén károsan hathatnak az LLM-ek működésére, különösen, ha ezek az adatok az indexelés során létrejövő archívumokba (alias későbbi tanítóanyagba) jutva megzavarják az LLM-mel történő adatfeldolgozást. Hasonló eredményt lehet elérni a toxicitás módszeres alkalmazásával is, kihasználva a nagy nyelvi modellek azon sérülékenységét, hogy „megmérgezhetők". Ebben az esetben a digitális lábnyomba (például a *deep webre*) helyezünk el olyan mennyiségű és típusú toxikus adatot, amely az általános felhasználásban nem okoz fennakadást, ám az LLM-ek tanítását vagy működését jelentősen megzavarhatja. Vizsgálat tárgyát képezheti az is, hogy milyen védelmi hatást érhetünk el azzal, ha a digitális környezetet úgy alakítjuk át, hogy az ne legyen ideális az LLM-alapú feldolgozásra – például angol nyelvű modellek esetében a különösen nehezen kezelhető

[44] A kiberbiztonságban a *CTI feed (cyber threat intelligence feed)* egy folyamatosan frissülő adatforrás, amely ismert fenyegetésekről, támadási mintákról és indikátorokról (pl. IP-címek, *domainek, hash-ek*) szolgáltat információt a védekezés támogatására.

[45] A FraudGPT és a WormGPT két GPT-modellen alapuló, kifejezetten kiberbűnözők számára készült eszköz, amelyeket a *dark weben* és Telegram-csatornákon terjesztenek. Olyan illegális tevékenységek támogatására tervezték őket, mint a célzott adathalászat, rosszindulatú kódok generálása és sebezhetőségek kihasználása.

[46] A 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.

[47] A *prompt injection* (utasításbefecskendezés vagy injekció) a nagy nyelvi modellek egyik legjelentősebb sebezhetősége, amely során a modellutasításokat úgy módosítják, hogy reprezentálás során eltérítse a modellműködést az alapértelmezettől.

nyelvek (mint az orosz vagy a kínai) használatával. Ezenkívül érdemes a *honeypotok* szöveges tartalmait tokenizálni, mivel így a nagy nyelvi modellek könnyebben képesek azokat feldolgozni, ami fokozza a megtévesztő hatást.

Jogi eszközök is bevethetők, bár ezen a területen a lehetőségek korlátozottak. A jogalkotás hagyományosan késve követi a technológiai fejlődést, ennek ellenére indítványok megküldésével elérhető például, hogy a keresőmotorokból vagy archiváló szolgáltatók adatbázisaiból eltávolítsanak bizonyos tartalmakat, ami szintén befolyásolja az OSINT-LLM-rendszerek adathozzáférését. Védekezésünk során az OSINT-LLM-rendszerek kitettségéből (mint egyfajta ellátásilánc-sebezhetőség) is előnyt kovácsolhatunk azzal, hogy technikai intézkedéseket alkalmazunk a legelterjedtebben használt *web scraping* eszközök vagy LLM-alkalmazások ellen. Ezzel a lépéssel csökkenthetjük a feltérképezés kockázatát, és így fokozottabban védhetjük adatainkat.[48]

Offenzív elhárítás

Az offenzív típusú védekezés célja az ellenfél rendszererősségeinek feltérképezése, majd annak félrevezetése szándékosan torzított információkkal. Ez lényegében egy OSINT-LLM megtévesztési (dezinformációs) stratégia. A gyakorlatban ez megvalósulhat például *honeypot* rendszerek alkalmazásával, eredeti weboldalak forráskódjaiba ágyazott manipulált elemekkel, vagy akár *deep webes* felületeken közzétett tartalmakkal is. Hatékony módszer lehet olyan adatcsomagok elhelyezése, amelyek elavult vagy teljes mértékben valótlan információkat tartalmaznak. A megtévesztés eredményességét tovább fokozhatja, ha a szöveges anyagokat nagy nyelvi modellek segítségével állítjuk elő, mivel ezek tokenizált formában jobban illeszkednek az LLM-ek feldolgozási mechanizmusaihoz, ezáltal technikailag hatékonyabb a félrevezetés.[49]

Összefoglalás és eredmények

A vizsgált rendszerek vonatkozásában sikerült azonosítani és részletesen bemutatni az architektúrák legfőbb azonosságait (főbb komponenseit), azonban az eltérések részletes vizsgálatát egyfajta összefoglalásként az 1. táblázat szemlélteti.

Az alábbi táblázatból megállapítható, hogy bár csekély számú elemet érintett vizsgálat, ez a két projekt is bőven eltér a főbb LLM- és OSINT-funkciók tekintetében. Az eltérésekből azt is megállapíthatjuk, hogy az LLM OSINT PoC architektúrája az LLM-modul jellemzői tekintetében mutat fejlettebb szintű megvalósítást, míg az InfoHound vonatkozásában a csökkent LLM-funkciók mellett az OSINT-eszközök széles körű alkalmazásán van a hangsúly. Ez arra enged következtetni, hogy mindkét projekt esetében az adott modulok még bőven fejleszthetők, és a technológiai lehetőségek nincsenek maradéktalanul kihasználva.

[48] Bottyán 2024: 84.
[49] Bottyán 2024: 84.

1. táblázat: A vizsgált projektek további architekturális jellemzői

Architektúrajellemzők	**LLM OSINT PoC**	**Infohound**
LLM-verzió	GPT-4 és GPT-3.5	Llama 2 (Olama)
LLM-komponens	Többszintű (multi-agent)	Egyszintű (single-agent)
LLM-munkamenet	Ismétlődő (triggered process)	Egyszeri (one-time session)
LLM-funkció	Információkereső ügynök, szöveg összefoglalása és generatív szövegalkotás (abstractive summarisation and content generation)	Szöveg összefoglalása (only content generation)
LLM-utasítás (prompt)	Alapvető (structured)	Rövid, egyszerű (naive)
OSINT-eszközök	Egyszerűbb keresőmotorok, adatgyűjtő eszközök	Számos OSINT eszköztípus
OSINT-tevékenység	Webkaparás (web scraping)	Többféle tevékenység, de főként web scraping

Forrás: a szerző szerkesztése

Az eredmények tekintetében az H1 hipotézis bizonyítása sikeres, hiszen az LLM-alkalmazás már nem példa nélküli az internetes nyílt forrású információgyűjtésben, ugyanakkor a vizsgált elemek architektúrái tekintetében az LLM-komponensek kihasználása még nem kidolgozott. Ez a megállapítás bizonyítja azt a felvetést, hogy az LLM funkcionalitása harmonizálható az OSINT adatfeldolgozási folyamatában. Bizonyos az is, hogy az LLM-ek innovációs hatást gyakorolnak az OSINT-eszközök alkalmazására, és valószínűsíthetően a nagyobb, korábban leküzdhetetlennek hitt adatmennyiségek feldolgozásában fognak hatékonyan közreműködni.

A második hipotézis hamisnak bizonyult, amelynek oka, hogy e kutatás fókuszába olyan felhasználási módok kerültek, amelyekben az LLM-komponenseket főként az internetalapú OSINT adatfeldolgozási folyamatába integrálják azért, hogy a kívánt célok szerint kivonatolják a releváns információkat a nagy mennyiségű adatból. Önálló internetalapú OSINT-eszközként is alkalmazhatók, de valószínűleg nem hatékonyak. A nagy nyelvi modellek jellemzőiből kiindulva valószínűsíthető a hatékonyságnövelésük az internetalapú OSINT-folyamatokban, sőt képesek lehetnek redukálni vagy megoldani az OSINT néhány elemi nehézségét (például adatdömping, elemzési képességek). Egyértelmű azonban, hogy a nagy nyelvi modell komponensek alkalmazása a vizsgált projektek architektúrai jellemzői alapján nem nyitnak meg új kockázati területeket, az ezzel összefüggő támadási vektor azonos. Ugyanakkor növelik az internetalapú OSINT kiberbiztonságra irányuló fenyegetettségét, hiszen a kibertámadások időigényes felderítési fázisa gyorsabb és hatékonyabb lesz, időigénye valószínűsíthetően töredéke lesz a korábbinak.

A harmadik hipotézis bizonyított azzal, hogy az elhárítási stratégiai megközelítések (információgyűjtés, defenzív és offenzív elhárítás) tartalmaznak olyan speciális elemeket (például az LLM-sebezhetőségek és hatékonyságcsökkentő intézkedések

alkalmazása védelmi céllal), amelyeket kifejezetten az LLM-OSINT-tevékenységekkel szembeni kockázatok mérséklésére alkalmazhatunk. A továbbiakban egy prediktív kitekintést teszünk, és a feltárt információk figyelembevételével valószínűsíthető változásokat azonosítjuk a jövőre nézve.

Az OSINT és az LLM jövője

A korábban megfogalmazott megállapítás, miszerint az OSINT-LLM-keretrendszerek fejlődése még a vizsgált rendszerek alapján kezdetleges, felveti a kérdést, hogy a jövőben várható-e ebben változás, és ha igen, milyen irányokban. Valószínűsíthető, hogy az LLM-architektúrák tovább fejlődnek, kompaktabbá és ezáltal könnyebben testre szabhatóvá válnak, amely helyzet jól hasznosul az OSINT-ben. Manapság a technológiai nagyvállalatok egyre inkább integrálják az LLM-alapú megoldásokat saját szolgáltatásaikba (például a Microsoft CoPilot vagy a Google Assistant with Bard[50]) egyfajta asszisztensként. Hasonló fejlődési irány várható az OSINT területén is: az LLM-ek itt is támogató és hatékonyságnövelő szerepet fognak betölteni. Mindemellett várható, hogy a már azonosított keretrendszertípusokba fokozatosan egyre több OSINT-eszközt fognak beépíteni, mivel az LLM-ek adatfeldolgozó kapacitása még messze nincs maximálisan kihasználva, továbbá egyre hatékonyabb és erőforrás-takarékosabb LLM-verziók megjelenése is előrevetíthető.

A jövőben a kifejezetten hírszerzési célokra összeállított domainspecifikus korpuszok is megjelenhetnek, amelyekre a finomhangolt LLM még hatékonyabb komponenssé válhat az internetes adatgyűjtésben. Ezzel összefüggésben számos további olyan fejlődő tényező felmerülhet, amely szintén pozitívan hathat a nagy nyelvi modellel végzett nyílt forrású információgyűjtésre. Biztosra vehető a jövőben a mesterséges intelligencia képességeinek általános bővülése – különösen a nagy multimodális modellek (LMMs) fejlődése – is, ami további új megközelítések kidolgozásához járulhat hozzá, akár kihasználva az internet eddig kevésbé feltárt rétegeit is.

A kutatás eredményei rámutatnak arra is, hogy az LLM-ek és az OSINT közötti kapcsolat nem csupán technológiai, hanem stratégiai jelentőséggel is bír, különösen a kiberbiztonság területén. A vizsgált projektek ugyan csak pillanatképet nyújtanak az LLM-ek jelenlegi OSINT-alkalmazásairól, mégis jól érzékelhető belőlük a jövőbeni tendenciák iránya. Ennek megfelelően a témában elengedhetetlen a tudományos diskurzus fenntartása. Ehhez a megállapítások kellő alapot nyújthatnak, és hozzájárulhatnak új tudományos kérdések kidolgozásához olyan tudományterületeken, mint a jogtudomány, a rendészettudomány, a hadtudomány és a katonai műszaki tudomány. Emellett eredményei és ajánlásai érvényesíthetők olyan területeken, mint az adatvédelem, a katonai és nemzetbiztonsági hírszerzés és a védelmi ipar. Az sem kizárt, hogy a jövőben az OSINT-LLM-rendszerek fejlődése új kutatási irányokat nyithat meg, különös tekintettel a stratégiai kiberfenyegetés-értékelés és a dezinformációs kampányok elleni fellépés területén.

[50] Hsiao 2023.

Felhasznált irodalom

Bielska, Aleksandra (2020): *Open Source Intelligence Tools and Resources Handbook.* Online: https://i-intelligence.eu/uploads/public-documents/OSINT_Handbook_2020.pdf

Bottyán Sándor (2024): *Nagy nyelvi modellel támogatott nyílt forrású információgyűjtés a kibertérben.* Tudományos diákköri dolgozat. Budapest: Nemzeti Közszolgálati Egyetem Államtudományi és Nemzetközi Tanulmányok Kar.

CyberRaya (2023): *Infohound: Enumerate Emails, Usernames, Subdomains, Names and Other Information.* Online: https://osintteam.blog/infohound-enumerate-emails-usernames-subdomains-names-and-other-information-828bb9b8ac8e

Černý, Jan (2024): Implications of Large Language Models for OSINT: Assessing the Impact on Information Acquisition and Analyst Expertise in Prompt Engineering. *Proceedings of the 23rd European Conference on Cyber Warfare and Security,* 23(1). Online: https://doi.org/10.34190/eccws.23.1.2261

Department of Defense (2010): *Dictionary of Military and Associated Terms.* Online: https://irp.fas.org/doddir/dod/jp1_02.pdf?utm_source=chatgpt.com

Holupredictions (2023): *It Looks Like You Can Use ChatGPT to Bypass Paywalls.* Online: www.reddit.com/r/ChatGPT/comments/14j8q1u/it_looks_like_you_can_use_chatgpt_to_bypass/

Hsiao, Sissie (2023): *Assistant with Bard: A Step Toward a More Personal Assistant.* Online: https://blog.google/products/assistant/google-assistant-bard-generative-ai/

InfoHound – OSINT Tool for Domain Profiling (2024). Fundacio-i2CAT/InfoHound, Github. Online: https://github.com/Fundacio-i2CAT/InfoHound?tab=readme-ov-file

Lowenthal, Mark. M. (2003): *Intelligence: From Secrets to Policy.* Washington, DC: CQ Press. Online: https://archive.org/details/intelligencefrom0000lowe_g9z8

Mi az a proof-of-concept (PoC) exploit? (2023). Nemzetbiztonsági Szakszolgálat Nemzeti Kibervédelmi Intézet. Online: https://nki.gov.hu/it-biztonsag/tudastar/mi-az-a-proof-of-concept-poc-exploit/

Okemwa, Kevin (2023): ChatGPT Pauses Bing Integration to Stop People From Bypassing Paywalls. *Windows Central,* 2023. július 4.Online: www.windowscentral.com/software-apps/chatgpt-pauses-bing-integration-to-stop-people-from-bypassing-paywalls

Precognox (2022): *A hírek és információk megszerzése még nem hírszerzés.* Online: https://precognox.hu/blog/a-hirek-es-informaciok-megszerzese-meg-nem-hirszerzes/

Rogers, Reece (2023): How to Use ChatGPT's 'Browse With Bing' Tool – Plus 6 Starter Prompts. *Wired,* 2023. október 25. Online: www.wired.com/story/chatgpt-browse-with-bing-internet-connected-openai/

Rosengren, Kim (2022): *Contribution of Open-Source Intelligence to Social Engineering Cyberattacks.* Bachelor's Thesis. Turku University of Applied Sciences. Online: www.theseus.fi/bitstream/handle/10024/754826/Rosengren_Kim.pdf?sequence=2

Vadász Pál (2018): *A szemantikus keresés módszerei és alkalmazási lehetőségei a védelmi szférában, a közigazgatásban, illetve a gazdasági életben.* PhD-értekezés. Budapest: Nemzeti Közszolgálati Egyetem Katonai Műszaki Doktori Iskola. Online: www.uni-nke.hu/document/uni-nke-hu/Vadasz-Janos-Pal_Doktori-ertekezes_2018.pdf