
**1. ábra. Világszerte a választási kampányok meghatározó jelenségévé vált a chatbotok alkalmazása az ellenjelöltekről szóló dezinformációs narratíva terjesztése érdekében**
(Forrás: Shutterstock)

*Benke Bálint Péter**

# A megtévesztés új korszaka

## A mesterséges intelligencia fejlődésének hatása a dezinformációs kampányokra

Elsőként a mesterséges intelligencia fogalmát kell tisztáznunk. A mesterséges intelligencia az emberi intelligencia szimulációját jelenti olyan gépekben, amelyeket oly módon programoztak, hogy úgy gondolkodjanak és tanuljanak, mint az emberek. [1] A gépi tanulás a mesterséges intelligencia meghatározó területe.

A gépi tanulási algoritmus egy olyan számítási folyamat, amely a bemeneti adatokat oly módon használja fel a tervezett feladat elvégzéséhez, hogy azt nem szükséges explicit módon előre beprogramozni. Más szavakkal: nem szükséges „kemény kódolással” előre meghatározni a működését. Ezek az algoritmusok automatikusan módosítják a terveiket, vagy ismétlés, illetve tapasztalás révén alkalmazkodnak azokhoz, vagyis „puhán kódoltak”. Ahogy haladnak előre, egyre jobban teljesítik a kitűzött célt. A tanítási folyamat során bemeneti mintákat adunk meg a kívánt eredményekkel együtt. Az algoritmus ezután a legjobb képességei szerint konfigurálja magát annak érdekében, hogy képes legyen általánosítani, vagyis az adatok mellett a friss, korábban nem betanult adatokból is elő tudja állítani a kívánt kimenetet. Ezt a folyamatot nevezzük „tanulásnak”. [2] *(2. ábra)*

**ÖSSZEFOGLALÁS:** Napjainkban a rohamos mértékben fejlődő mesterséges intelligencia és a gépi tanulás nemcsak kihasználható képességeket hordoznak magukban, hanem megannyi veszélyt is. Egy ilyen veszélyforrás a dezinformáció mértékének növekedése és a terjesztés megkönnyítése. A mesterséges intelligencia a dezinformáció számos fázisát is automatizálta, illetve jelentősen segítette a dezinformációs szereplők munkáját. A tanulmányon belül szó lesz a deepfake és a szintetikus hangok generálásról, továbbá a közösségi botok dezinformáló hatásáról valamint a szentiment analízis által történő szövegelemzésről. A technológiák bemutatásán keresztül konkrét példák szemléltetik gyakorlati felhasználásukat.

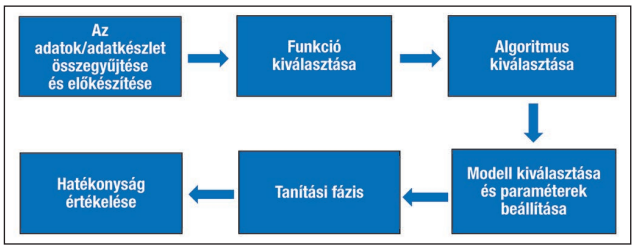
**KULCSSZAVAK:** mesterséges intelligencia, deepfake, álhír, generatív versengő hálózatok, természetes nyelvi generálás

**ABSTRACT:** Nowadays, rapidly advancing artificial intelligence and machine learning not only hold valuable potentials but also encompass numerous dangers. One such source of danger is the increase in the extent of disinformation and the facilitation of its spread. Artificial intelligence has automated various phases within disinformation and significantly aids the work of disinformation actors. Within this article, we will discuss the generation of deepfakes and synthetic voices, as well as the deceptive impact of chatbots and social bots, and the text analysis through sentiment analysis. Through the presentation of these technologies, the article will provide concrete examples to illustrate their practical applications.

**KEY WORDS:** artificial intelligence, deepfake, fakenews, generative adversarial networks, Natural Language Generation

* Nemzeti Közszolgálati Egyetem, Államtudományi és Nemzetközi Tanulmányok Kar, Kiberbiztonsági mesterképzési szak, hallgató.
ORCID: 0009-0001-2901-6319


DOI: 10.23713/HT.57.6.06

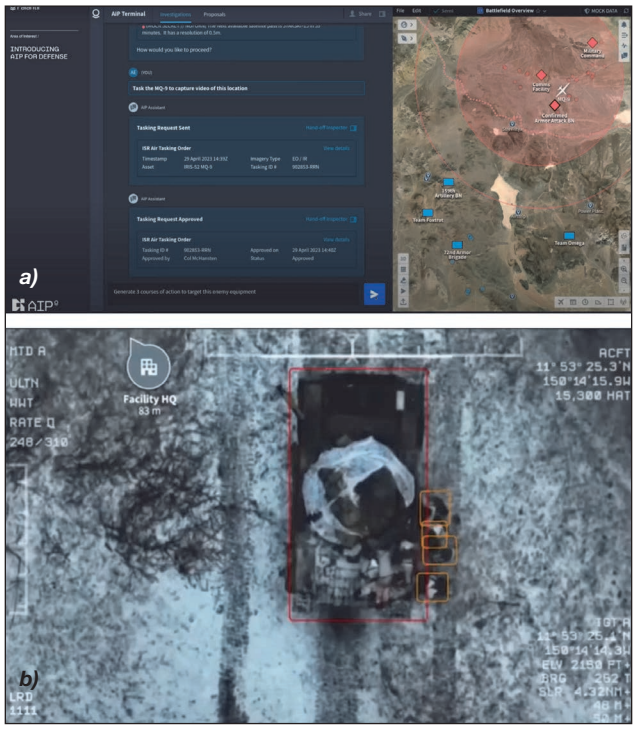



**2. ábra. Egy általános gépi tanulási modell elemei** (A szerző szerkesztése [3] alapján)

A gépi tanuláson belül háromfajta tanulási módszert különböztethetünk meg: a felügyelt tanulást, a nem felügyelt tanulást és a megerősítéses tanulást. A felügyelt tanulásnál a bemeneti és a kapcsolódó kimeneti változók (címkék) alkotják az algoritmus tanulási halmazát alkotó példapárokat. Az algoritmus ezek alapján létrehozza az adatfeldolgozási módszert/függvényt, majd a következő lépésekben felhasználja azt. A modell addig tanul, amíg el nem éri a szükséges pontosságot. Nem felügyelt tanulás esetében az algoritmus tanulási halmaza csak bemeneti változókat tartalmaz, címkéket nem. Az előre meghatározott kritériumok és a hasonlóság alapján osztályozza a bemeneti értékeket. Végül a megerősítéses tanulásnál az algoritmus minden elképzelhető kimenetet kipróbál, és a legjobbat választja ki, így biztosítva, hogy a döntés kimenetele a jó döntésekért járó jutalmak, és a rossz döntésekért járó büntetések alapján megerősítést kapjon. Mivel a mesterséges intelligencia a környezeti bemenetekre reagálva tanul, ez nem tanulóhalmazok és példapárok alapján történő tanulás, mint a korábbi módszerek esetén. [4]

Mielőtt részletesebben ismertetnénk a különféle módszereket, amelyekkel a mesterséges intelligencia segítségével hamis információk hozhatók létre és terjeszthetők, vizsgáljuk meg, hogy ezek az eszközök milyen hatást gyakorolhatnak a társadalomra. A könnyebb megértés érdekében néhány példát is említünk arra, hogyan működhet együtt a mesterséges intelligencia és a dezinformáció. Az egyik kiemelkedő eset a 2016-os amerikai elnökválasztás, amelyben közösségi botokat[1] használtak a választás menetének befolyásolására. [5] Brazíliában a 2018-as elnökválasztás során chatbotokat használtak a WhatsApp alkalmazáson a jelöltekről szóló dezinformáció terjesztésére. [6] Az Egyesült Királyságban, a brexitről szóló népszavazás alatt közösségi botokat alkalmaztak a Twitter nevű közösségi médiaplatformon. [7] A fenti példákból is jól látható, hogy a dezinformáció a 21. században is hatalmas szerepet játszik. A dezinformáció veszélyezteti a demokráciák integritását és aláássa a hírekbe vetett bizalmat. Az autokratikus államok médián keresztül történő szigorú ellenőrzése lehetővé teszi számukra az álhírek és dezinformációk hatékonyabb kezelését. A probléma súlyosságát jól szemlélteti, hogy egy Észak-atlanti Szerződés Szervezete (NATO) által támogatott kutatásban felmerült a kognitív hadviselés koncepciója, mint a hatodik hadviselési domén. [8]

A hadviselés területén a mesterséges intelligencia egyik lényeges új felhasználása a Palantir cég által fejlesztett PalantirAIP (All-Source Intelligence Platform) for Defense „targeting” funkciója. Az általános „targeting” funkció egy olyan folyamatot takar, amelyben a szoftver az adott területen lévő műholdképek alapján azonosítja és követi az ellenséges mozgást. Ez a képesség nem kötődik a dezinformáció területéhez, inkább az információgyűjtés és -elemzés hatékonyabbá tétele áll a középpontban, ugyanakkor fontosnak tartom a mesterséges intelligencia efféle újszerű




**3. ábra. A Palantir AIP kezelőfelülete *a)* és egy drón által készített felvétel *b)*** [9]

katonai alkalmazásának a megemlítését. Az eljárás rendszerint a következő lépésekből áll:

1. *Adatgyűjtés:* a PalantirAIP for Defense rendszeresen kap műholdképeket és más hírszerzői adatokat, amelyeket automatizált vagy szakember által felügyelt elemzésnek vet alá.
2. *Ellenséges mozgások azonosítása:* az AI-alapú szoftver az algoritmusok és a gépi tanulás segítségével elemezi a műholdképeket, hogy felfedezze a potenciálisan ellenséges erőket vagy tevékenységet jelző mozgó elemeket.
3. *Katonai értesítés:* miután az AI azonosította a potenciális célpontokat vagy ellenséges mozgásokat a műholdképeken, értesítést küld a hadműveletért felelős tiszteknek. Ez a tevékenység lehetővé teszi számukra a megfelelő reakció vagy további intézkedések megtételét, például felderítő járőrök küldését vagy más műveletek előkészítését. A katonák akár programozási képességek nélkül is további információkat kérhetnek a szoftvertől. Maga a szoftver egy nagy nyelvi modellen[2] (Large Language Model – LLM) alapul.
4. *Tervek kidolgozása:* a rendszer három műveleti tervet dolgoz ki, amelyek közül a felügyelő tiszt kiválasztja az általa legmegfelelőbbnek véltet. Az AIP részletezi, hogy hány ember, milyen felszereléssel mennyi idő alatt képes teljesíteni a megjelölt célt. [9]

A Palantir AIP-t a háborús területen, éles helyzetben is alkalmazzák a katonák, mint például az orosz–ukrán háborúban az ukrán fél. [10]

## A célcsoportok azonosítása gépi tanulási rendszerek segítségével szentiment analízissel és álláspontvizsgálattal

A szentiment analízis és az álláspontvizsgálat lehetővé teszi a dezinformáló szereplők számára, hogy megtalálják azokat a csoportokat vagy közösségeket, ahol az általuk

generált álhírek a legsikeresebben érnek célba. [11] Ezeknek a programoknak alapjául a természetes nyelvi feldolgozás és a természetes nyelvi generálás szolgál. Ez a két kutatási terület lehetővé teszi a gépi tanulási rendszerek számára a szövegek olvasását, írását, értelmezését, és lehetővé teszi ezáltal a szentiment analízist is. [12]

A szentiment analízis a felhasználó egy entitással kapcsolatos véleményét, attitűdjét és érzelmeit tanulmányozza számítógép segítségével. Az entitás lehet személy, esemény vagy különböző témák. A szentiment analízis érzelmeket és hangulatot – mint például az öröm (pozitív) vagy bosszúság (negatív) – próbál elemezni és kategorizálni a szövegben. Ezt a technológiát leggyakrabban markerting célokra használják, például egy termék megítélésének felméréséhez. Fontos kiemelni, hogy az elemzés eredményeként a vállalatok anyagi előnyhöz juthatnak. A cégek bevétele további fejlesztéseket tesz lehetővé a szentiment analízis technológiája számára. Ennek nyomán a technológia a jövőben még pontosabb adatokat szolgáltathat.

A fejlettebb gépi tanulási algoritmusok a közösségi médiában használt szlenget, hangulatjeleket, emojikat és rövidítéseket is képesek megérteni. Ezeket a kifejezéseket nyelvészek által kiválasztott adathalmazokon tanítják. E képességek megkönnyítik a dezinformációs szereplők munkáját, mert a szentiment analízis segíthet az emberek beállítottságának meghatározásában, például politikai nézeteik megállapítható az általuk követett emberek, és a közzétett posztjaik alapján. [13]

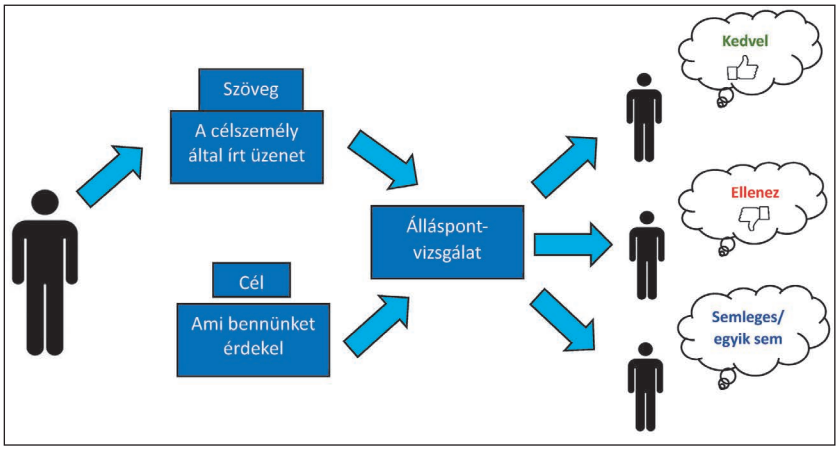
Az álláspontvizsgálat *(4. ábra)* meghatározza a szöveg szerzőjének hozzáállását egy olyan entitáshoz vagy véleményhez, amelyet a szöveg közvetlenül megszólít vagy sugall. A hozzáállás három kategóriába sorolható: ellenző, támogató vagy semleges. [14] Az álláspontvizsgálatra legtöbbször a felügyelt tanulási módszert (Support Vector Machine – SVM) alkalmazzák. A SVM egy olyan számítógépes algoritmus, amely példák alapján tanulja meg a kategória hozzárendelését az entitásokhoz. Ezzel a módszerrel az arcok felismerésétől a hitelkártyacsalásokig számos feladatra betaníthatók a programok. [15] Az álláspontvizsgálattal kapcsolatos kezdeti kutatások célja a résztvevők ideológiai, vagy vitás kérdésekkel kapcsolatos álláspontjának azonosítása online vitafórumokon. A rosszindulatú szereplők az elemzések ezen formáinak segítségével azonosíthatják célcsoportjaikat, hogy dezinformálásukkal csak ezeket a személyeket célozhassák meg. Ilyen vitatott kérdések lehetnek többek között a feminizmus, az ateizmus, az abortusz vagy éppen egy politikai párt, illetve személy. [14]

Fontos szót ejtenünk a pszichológiai műveletekről (Psychological Operations – PSYOP). A pszichológiai műveletek olyan tevékenységek, amelyeknek célja kiválasztott információk átadása a célcsoportnak annak érdekében, hogy azok befolyásolják kormányok, szervezetek, csoportok és vállalatok motivációit, érvelését, és végső soron a viselkedését. Befolyásolás információval, pedig a katonai információs támogató műveletek (Military Information Support Operations – MISO) küldetések során meghatározott információkat osztanak meg a külföldi közönséggel, hogy befolyásolják a külföldi kormányok és állampolgárok érzelmeit, motívumait, érvelését és viselkedését. Ez magában foglalhatja a kiberhadviselést és a fejlett kommunikációs technikákat a média minden formáján keresztül. Szándékos megtévesztésről akkor beszélünk, amikor harci helyzetekben, a katonai megtévesztési tevékenységek során a pszichológiai hadviselést használják az ellenséges erők szándékos félrevezetésére. Célja, hogy hatást gyakoroljanak az ellenség személyi állományának (a népesség különböző csoportjai) érzelmeire, s ezáltal befolyásolják viselkedését, szándékát, akaratát a politikai, és a katonai célokkal összhangban. Célja továbbá erősíteni a baráti, illetve a lojális népesség szimpátiáját a kitűzött célok elérését szolgáló műveleteket érintően; valamint elnyerni az el nem kötelezett népesség támogatását és fejleszteni együttműködési készségét. [16]

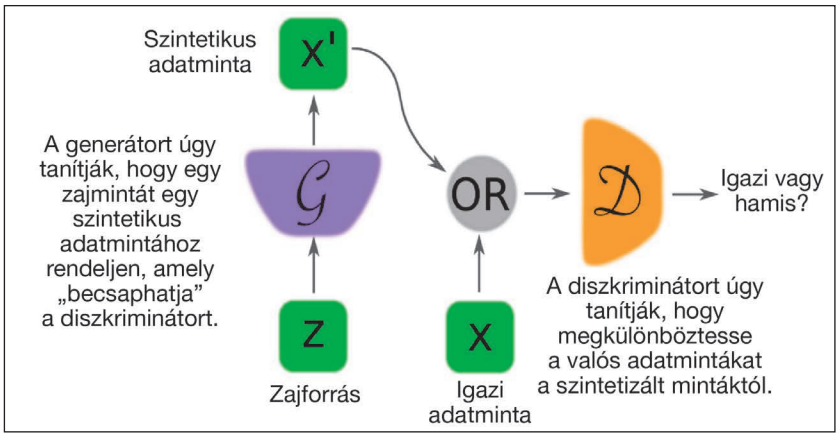
## A HAMIS TARTALOM ELŐÁLLÍTÁSA GENERATÍV VERSENGŐ HÁLÓZATOK SEGÍTSÉGÉVEL

A generatív versengő hálózatok (Generative Adversarial Networks – GANs) olyan típusú neurális hálózatok, amelyek két neurális hálózati modellből állnak, egy generátorból és egy diszkriminátorból, amelyeket a GAN arra használ, hogy megkülönböztesse a valódi mintákat a generált mintáktól. *(5. ábra)* Amint a neurális hálózatok versenyeznek, tanulnak egymástól, a versengés során egyre valósághűbb eredmények születnek, például kép-, videó- és hanginformációk esetében egyaránt. Fontos tudni, hogy a generátor csak a diszkriminátorral való interakció révén tanul, és nincs közvetlen hozzáférése a valódi forrásfájlokhoz. A szintetikus minták és a valódi képek halmazából vett minták egyaránt a diszkriminátor rendelkezésére állnak. Egy kép elemzése után a diszkriminátor a mérési alap (ground truth) segítségével megállapítja, hogy a kép valós, vagy generált. A diszkriminátor segítségével a generátor betanítható, hogy még valósághűbb és jobb minőségű hamisítványokat készítsen. [17]

**4. ábra. Álláspontvizsgálat (Stance detection) leegyszerűsített menete** (A szerző szerkesztése [14] alapján)



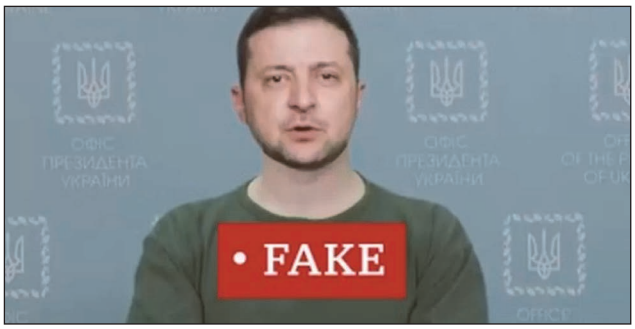

A neurális hálózat olyan gépi tanulási modell, amelyet az emberi agy neurális hálózatának szerkezete és működése ihlet. Ez egy számítási modell, amely rétegekbe szervezve, összekapcsolt csomópontokból vagy mesterséges neuronokból áll. A neurális hálózatokat úgy tervezik, hogy azok dolgozzák fel az adatokat és tanuljanak az információkból oly módon, hogy képesek legyenek mintákat felismerni, előrejelzéseket készíteni, valamint olyan feladatokat elvégezni, amelyeket nehéz lenne explicit módon programozni. A neurális hálózatban az információ az adatok kezdeti beviteli rétegén keresztül halad egy vagy több rejtett rétegen, majd végül a




**5. ábra. A GAN képzési folyamata során tanuló két modell a diszkriminátor „D” és a generátor „G”** (A szerző saját szerkesztése [18] alapján)

kimeneti réteghez jut, amely létrehozza a hálózat végső előrejelzését vagy eredményét. A csomópontok közötti minden kapcsolatnak van egy hozzárendelt súlya, és ezeket a súlyokat a tanulási folyamat során beállítják annak érdekében, hogy optimalizálják a hálózat teljesítményét egy adott feladat végrehajtása során.

A GAN-ok különösen jók deepfake videók és szintetizált hangok létrehozására. Ez a technika a „deepfake” nevet viseli, mivel az alapja egy mélytanulási modell. A többrétegű neurális hálózatok tanítása azon gépi tanulási részhez tartozik, amelyet mélytanulásnak nevezünk. A deepfak arra is felhasználható, hogy megváltoztassák egy prominens katonai vezető, politikai- vagy médiaszemélyiség hírnevét. A dezinformációs tartalmakat létrehozó személy a deepfake technológiát felhasználhatja arra, hogy hitelesnek tűnő videókat készítsen azáltal, hogy az arcot egy szintetikus hanggal és képpel párosítja, amely a tényleges szereplő hangját és mimikáját utánozza. Erre egy aktuális példa a Ukraine 24 nevű ukrán tévécsatornán megjelenő Volodimir Zelenszkij ukrán elnökről készült deepfake videó. *(5. ábra)* Ebben a videóban felszólítja az ukrán katonákat, hogy tegyék le a fegyvert és adják meg magukat az orosz erőknek. [19]




**6. ábra. A Volodimir Zelenszkijről készült deepfake videó részlete** [19]

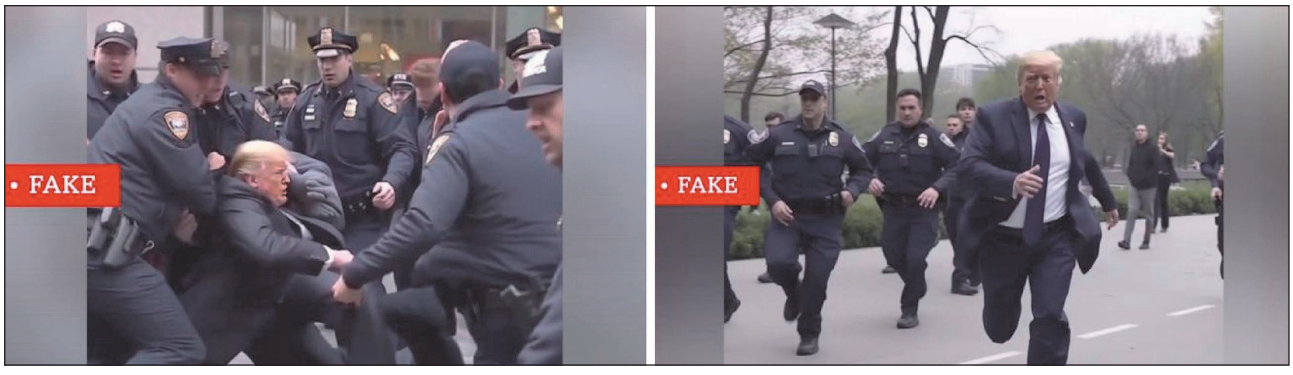
A deepfake és a szintetikus hang generálására alkalmas technológiák fejlesztésével együtt járhat az emberek növekvő szkepticizmusa a hírekkel kapcsolatban, és megnehezítheti annak eldöntését, hogy a látott vagy hallott hír, illetve videó valós-e vagy hamis. A GAN-ok nemcsak a tartalomkészítési fázisban használhatók, hanem a terjesztési fázisban is. A dezinformatív üzenetek hatékony terjesztése érdekében számos hamis profilt állítanak elő. Ezekhez a profilokhoz profilképek szükségesek, hogy még hitelesebbek legyenek. Ilyen hihető kép a gépi tanulás segítségével könnyedén előállítható. Erre jó példa a thispersondoesnotexist.com elnevezésű oldal. Minden egyes alkalommal, amikor a weboldalt frissítjük, egy új arckép jelenik meg előttünk, amelyet a GAN-technológia hoz létre. [13]

A Midjourney egy sokkal fejletteb weboldal, amely mesterséges intelligenciát használ a kép generáláshoz. A Midjourney kulcsszavak megadása alapján kreál valósághű képeket. Fontos hangsúlyozni, hogy ez a technológia is folyamatos fejlődésen megy keresztül. A hamis képek generálására alkalmas eszközök egyik példája a DALL-E, amelynek első változata 2021 januárjában jelent meg, amelyet a DALL-E 2 követett 2022 áprilisában. Ezalatt az idő alatt még kifinomultabban és gyorsabban képes képeket előállítani a program, ráadásul ezzel az eszközzel a felhasználók már módosíthatják, hozzáadhatják vagy törölhetik egy meglévő kép elemeit. *(7. ábra)* Egy kitűnő példa a hamis hangok generálására az ElevenLabs által nyújtott hangklónozási szolgáltatás. A weboldal ingyenesen elérhető felületén a felhasználó egészen 2500 karakterig megadhatja a saját szövegét, majd az mp3 hangfájlok feltöltése után a program a megadott hangon ismétli a betáplált szöveget. [15] Az itt felsorolt példák mind ingyenesen elérhetők.

## Szöveges tartalom előállítása természetes nyelvi generálással

A természetes nyelvi generálás (Natural Language Generation – NLG) olyan folyamatot jelent, amely során emberi kommunikációhoz hasonló szöveg vagy beszéd

**7. ábra. Donald Trump elnök letartóztatásáról készített hamis képek** [21]



**1. táblázat. A természetes nyelvi generálás 6 lépése** (A szerző szerkesztése [22] alapján)

| Tartalom-elemzés | Adatok megértése | Dokumentumok strukturálása | Mondatok összesítése | Nyelvtani strukturálás | Nyelv ismertetése |
|---|---|---|---|---|---|
| az adatok szűrése és a fontos tartalmak azonosítása | az adatok értelmezése, a minták azonosítása és kontextusba helyezése | dokumentumterv és narratív struktúra létrehozása | releváns mondatok kombinálása a téma összefoglalása érdekében | nyelvtani szabályok alkalmazása a természetes hangzású szöveg létrehozása érdekében | utolsó változtatások végrehajtása és bemutatása |

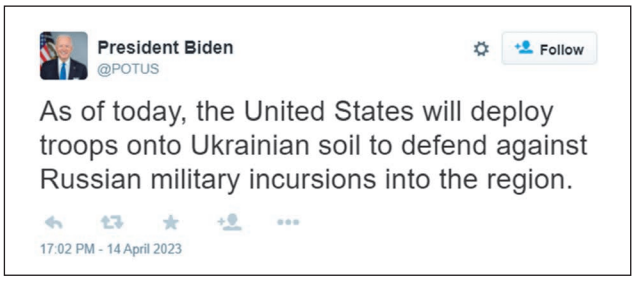
kerül generálásra. Ez a folyamat nem emberi nyelvi bemenet alapján, hanem az automatizált rendszerek segítségével hajtható végre. Ez a természetes nyelvfeldolgozás egyik részterülete, amely olyan koherens és folyékony szöveg létrehozására összpontosít, amely úgy tűnik, mintha ember írta vagy mondta volna. Az NLG a nyelvi bemenetből előállított szövegre vagy beszédre is vonatkozhat, mint például összegzés vagy szövegegyszerűsítés. A szöveg generálás módszerét 6 lépését az *1. táblázat* szemlélteti.

A tartalomkészítési fázison belül az NLG hozta a legnagyobb változást a dezinformációs kampányok számára. Ez a technológia meghatározott paraméterek alapján képes szövegeket létrehozni, amelyek hihetőnek és egyedinek hangzanak, ily módon, a kampányok számára megkönnyítik a dezinformáció írásának folyamatát. Kiemelendő, hogy ezek a szöveggeneráló rendszerek képesek célzott és precíz tartalmak létrehozására is, mivel a generatív nyelvi rendszerek példák alapján utánozzák a különböző írásmódokat. Kvalitásaik révén hitelesnek tűnő, valójában azonban hamis levelek vagy e-mailek létrehozására képesek, amelyek megkérdőjelezhetik egy adott közszereplő megbízhatóságát. Ezeket a hamisított üzeneteket akár a „hack-forge-leak" típusú támadásokban is felhasználhatják, ahol a hamisított e-mailt úgy teszik közzé, mintha az valós lenne. [13] A szöveg generálása napjainkban már a médiából és a hírekből is ismertté vált jelenség, és a ChatGPT révén meglepően könnyen hozzáférhető és alkalmazható technológiává vált. A ChatGPT ingyenes változata a GPT 3.5-ös (Generative Pre-trained Transformer) verziót alkalmazza, ami 175 milliárd paraméterből áll. Azonban a fizetős változat már a GPT 4-est használja, amely paramétereinek száma eléri a 170 billiót. A modell paramétereinek száma befolyásolja a modell összetettségét és kifejezőképességét, valamint azt, hogy mennyi adatot képes kezelni. [23]

A Journal of Experimental Political Science című folyóiratban megjelent tanulmányban a kutatók két különböző kísérleti megközelítéssel vizsgálták a nyelvi modellek társadalmi hatásait. A tanulmányban mesterséges intelligenciával írt hírcikkeket alkalmaztak, hogy megvizsgálják azok valóságtartalmát és azt a képességüket, hogy megkérdőjelezzék a külpolitikával kapcsolatos előítéleteket. A tanulmányban az OpenAI által létrehozott GPT-2 három különböző verziójának közepes-nagy 355 millió, nagy 724 millió és extra nagy 1,5 milliárd paraméterét használták. A kutatók az első kísérletben a mesterséges intelligencia által generált szövegek emberi hitelesség-érzékelésének felső határát vizsgálták, valós hírekkel összehasonlítva. A második kísérlet azt vizsgálta, hogy a párthovatartozás milyen hatást gyakorol a mesterséges intelligencia által generált hírek hitelességére. Ez utóbbi vitákat váltott ki arról, hogy a politikailag egyetlen párt mellett kiálló emberek nagyobb valószínűséggel hisznek-e a politikailag kedvező híreknek, és nagyobb valószínűséggel ragaszkodnak-e ezekhez a véleményekhez, még úgy is, ha feltüntették, hogy a szöveget szintetikusan generálták. A harmadik kísérletben három különböző méretű mesterséges intelligencia szöveggeneráló modell képességét vizsgálták, hogy azok mennyire képesek emberi segítség nélkül automatikusan híreket létrehozni, és a generált hírek mennyire hitelesek. Az első kísérlet azt az eredményt hozta, hogy az olvasók a mesterséges intelligencia által generált, de emberek által szerkesztett szöveget ugyanolyan hitelesnek vélték, mint egy ugyanarról az eseményről szóló, eredeti és emberi kézzel írott cikket. A második kísérletben arra jutottak a kutatók, hogy a pártállás egyértelműen jelentős hatást gyakorol arra, hogy az emberek számára mennyire tűnik hitelesnek egy adott hír. A kutatók végül arra a következtetésre jutottak, hogy nagy veszélyt jelenthet, ha ezeket a közegeket a dezinformáló szereplők specifikusan célozzák. Személyes álláspontom szerint a kutatók legjelentősebb eredménye a harmadik kísérletből származik. A szöveggeneráló modellek képesek voltak emberi beavatkozás nélkül is hitelesnek tűnő hírcikkeket alkotni, de ezek nyilván nem voltak annyira hihetők, mint az emberek által utólag szerkesztett változatok. [24]

## Közösségi botok és az álhírek automatizált terjesztése

A közösségi bot olyan számítógépes program, amely automatikusan készít tartalmakat és lép interakcióba a közösségi médián keresztül a felhasználókkal, hogy utánozza, és megváltoztassa a viselkedésüket. [25] Ezeknek a közösségi botoknak a célja egyes profilok, csoportok közönségének mesterséges növelése lehet, illetve egy vállalat vagy egy nyilvánosan ismert személyiség hírnevének rombolása. A múltban a botok főként egyfajta tevékenységet, a tartalmak automatikus közzétételét végezték. Ezek a korai botok még kezdetlegesek voltak, így a kiszűrésük sokkal kevésbé volt bonyolult feladat, mint napjainkban. Manapság a közösségi botok sokkal kifinomultabbak. A botok képesek olyan típusú információkat keresni a világhálón, amelyekkel feltölthetik a profiljukat. Emellett képesek úgy is megosztani tartalmakat, hogy ne keltsenek gyanút a közösségi botokat észlelő rendszerekben, például előre meghatározott időpontokban posztolnak, így utánozva a tartalom előállításának emberi időbeli jellegzetességeit. Ugyanakkor képesek beszivárogni a vitákba is, és természetes nyelvi algoritmusok segítségével a témához illő válaszokat produkálni. Ezekkel a válaszokkal a bot profilja új követőkre tehet szert, vagy tartalmait valódi emberek is megoszthatják. A botok a korábban említett szentiment analízissel elemezhetik a szöveget, ezzel segítve a válaszok létrehozását. [25] Egy kiváló minőségű, ingyen elérhető webes példa a Prank Me Not oldal. [26] Ezen a weboldalon hamis tweetek vagy facebookos csevegések állíthatók elő, amelyeket a közösségi botok felhasználhatnak hamis információk terjesztésére. A *8. ábrában* egy általam kreált példa olvasható, amely az Amerikai Egyesült Álla-




**8. ábra. Prank Me Not weboldal segítségével készített hamis tweet: „Mától az Egyesült Államok csapatokat telepít Ukrajna földjére, hogy megvédje magát a térségbe irányuló orosz katonai betörések ellen”** (A szerző saját szerkesztése)

mok elnökének tweetjét mutatja, az amerikai szárazföldi erők ukrajnai bevetéséről. A weboldal használatához minimális angol tudás szükséges, de más szakértelmet különösebben nem igényel. *(8. ábra)*

A clearneten és a darkneten[3] léteznek olyan piacok, ahol az internetfelhasználók előre meghatározott díj ellenében igénybe vehetik a közösségi botok üzemeltetői által nyújtott szolgáltatásokat. A nyíltan hozzáférhető szolgáltatásokat clearnetnek nevezzük. Ha valaki a darkneten használ egy bizonyos keresőmotort, akkor a személyazonossága rejtve marad. A kedveléseket, megosztásokat és kommenteket leggyakrabban ezek a szolgáltatások kínálják. Ezeket a közösségi botokat gyakran alkalmazzák adathalász és túlterheléses támadásokra. Végső soron a közösségi botok jelentős szerepet játszanak a dezinformációs kampányokban, mivel a már rendelkezésre álló építőelemeket, például a szövegelemzéshez szükséges hangulatelemzést és a GAN-rendszert felhasználhatják egy hihető profil kialakításához. [27]

A *9. ábra* a dezinformáció fázisait szemlélteti, amelyben az általam kiemelt módszerek segítséget nyújthatnak.

## 1. Felderítés

Ez a szakasz magában foglalja az információgyűjtést és -kutatást a potenciális célpontok, a sebezhetőségek és a leghatékonyabb terjesztési csatornák azonosítása érdekében. Legfontosabb az adatok gyűjtése. Ilyen technika például a „data scraping” és a profilozás, valamint a célközönség preferenciáinak és meggyőződéseinek megértése. A data scraping jelentése: nagy mennyiségű adat gyors kinyerése weboldalakról. Szoftvereszközök, vagy szkriptek segítségével a program weboldalakon megkeresi az egyes adatelemeket, majd összegyűjti és strukturált formátumba rendezi ezeket az adatokat. Ebben a fázisban az álláspontvizsgálat és a szentiment analízis óriási segítséget jelenthet a dezinformációs szereplők számára.

## 2. Tartalomkészítés és eltérítés

Itt a dezinformációs szereplők olyan tartalmat hoznak létre vagy manipulálnak, amely megfelel a saját céljaiknak. Előfordulhat, hogy teljesen hamis „híreket” hoznak létre, szelektíven szerkesztik az információkat, vagy meghamisítják a valós híreket, hogy azok megfeleljenek a saját narratívájuknak. Fontos kiemelni a természetes nyelvfeldolgozáson alapuló szöveggeneráló programokat és a GAN-okat, amelyek szintetikus hangokat és deepfaket állítanak elő. A tartalomeltérítésre példaként említhetjük, amikor a közösségi bot a követők száma hitelesebb hírcsatornának állítja be a dezinformáló oldalakat.

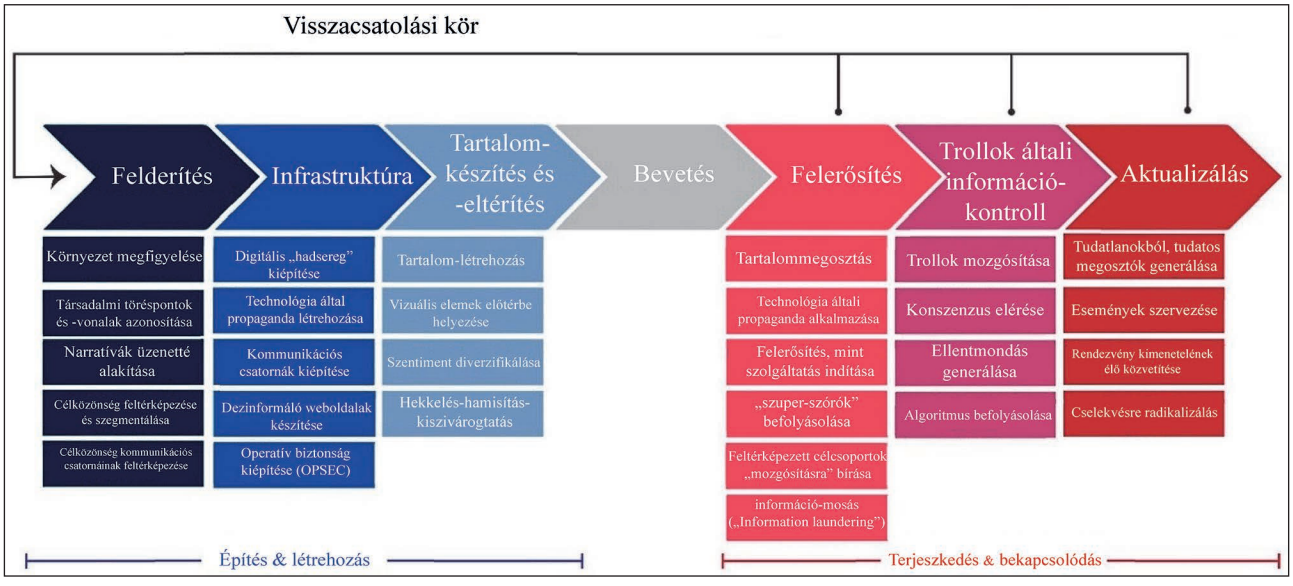
## 3. Felerősítés

Ebben a fázisban hatalmas szerepet játszanak a közösségi botok, amelyek megosztják a tartalmakat, és figyelmet generálnak. Ebben az esetben az úgynevezett visszhangkamra (echo chamber[4]) effektust is alkalmazhatják, hogy minél több elkötelezett követőre tegyen szert a dezinformáló szereplő. [10]

## Összegzés

A tanulmány bemutatja, hogy a gépi tanulás és a mesterséges intelligencia számos fejlesztése és eszköze dezinformációs kampányokban is felhasználható. Ezek a technikák nem képesek a dezinformáció terjesztésének teljes körű

**9. ábra. A dezinformáció fázisai** (A szerző szerkesztése [13] alapján)



automatizálására, de a 8. ábrán feltüntetett fázisok mindegyikében segítő hatást érhet el a mesterséges intelligencia és a gépi tanulás. Sok esetben az is megállapítható, hogy ezekhez a képességekhez nem szükséges hatalmas tudással rendelkezni, mivel gyakran a már megalkotott eszközök nyílt forrásból beszerezhetők, és csupán a használatukat kell elsajátítani. Mivel a mesterséges intelligencia fejlesztése és kutatása folyamatosan fejlődik, ezért nem tudható, hogy ez a fejlődés a jövőben hová vezet. Emiatt az eszközök szabályozása szükséges, és megfelelő ellenőrzést kell kidolgozni az ilyen technológiák által előállított tartalmak felismerése érdekében. Fontos, hogy létezzen olyan alternatíva, amely ellensúlyozza a mesterséges intelligencia által létrehozott fenyegetéseket, mivel ezen eszközök rosszindulatú használata jelentős következményekkel járhat. Ilyen ellensúlyt jelenthetnek maguk a mesterséges intelligencia alapú ellenőrző rendszerek is. Katonai szemszögből mind a hátországban élő civilek, mind a katonák figyelmét lényeges felhívni a kognitív műveletek befolyásoló képességére. Kiemelten fontos, hogy a hátországban élő civilek és a katonák tisztában legyenek azzal, hogy az ellenség kognitív műveleteinek befolyásolási képessége jelentős hatást gyakorolhat a hadműveletek kimenetelére és a politikai vezetés támogatottságára. Ezért elengedhetetlen a dezinformáció lehetőségének hangsúlyozása. A többdimenziós műveletek során a kognitív tényezők egyszerre több műveleti színtéren is jelentős hatást gyakorolnak. Az egyidejűleg zajló katonai tevékenységek sokszínűsége és komplexitása miatt a kognitív képességek befolyásolják az események alakulását és a döntések eredményességét. A katonák kognitív teljesítménye kulcsfontosságú annak érdekében, hogy hatékonyan reagáljanak az adott helyzetekre, megfelelően értelmezzék az információkat, helyes válaszokat adjanak azokra, valamint koordinálják az egyidejűleg folyó műveleteket. A helyes döntések meghozatala és a gyors cselekvés az összes érintett műveleti színtéren befolyásolja az együttműködés sikerét vagy kudarcát. [28]

## Hivatkozott irodalom

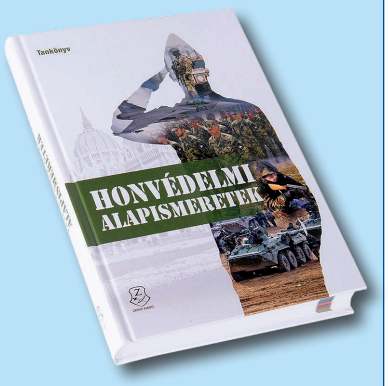

[1] Kaplan, J. Artificial Intelligence: What Everyone Needs to KnowR (Oxford University Press, 2016), 1.;
[2] Naqa, Issam El, Murphy, M. J. „What Is Machine Learning?”, in Machine Learning in Radiation Oncology: Theory and Applications, szerk. Issam El Naqa, Ruijiang Li, és Martin J. Murphy (Cham: Springer International Publishing, 2015), 4 https://doi.org/10.1007/978-3-319-18305-3_1;
[3] Alzubi, J., Nayyar, A., Kumar, A. „Machine Learning from Theory to Algorithms: An Overview”, Journal of Physics: Conference Series 1142, sz. 1 (2018): 012012, https://doi.org/10.1088/1742-6596/1142/1/012012;
[4] Dr. Németh András, Virágh Krisztián. Mesterséges intelligencia és a haderő – A mesterséges intelligencia területei III. rész Haditechnika LVI. évf. – 2022/3 pp. 2–7. https://doi.org/10.23713/HT.56.3.01;
[5] Badawy, A., Ferrara, E., Lerman, K. „Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign”, in 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, 258–65. https://doi.org/10.1109/ASONAM.2018.8508646;
[6] Ruediger, M. A. et al. „Bots and Brazil´s Electoral Legal System: 2018 Elections”, 2019. január, http://bibliotecadigital.fgv.br:80/dspace/handle/10438/26229 (Letöltve: 2023.7.20.);
[7] Howard, Ph. N., Kollanyi, B. „Bots. #StrongerIn, and #Brexit: Computational Propaganda during the UK-EU Referendum” (arXiv, 2016.6.20.), https://doi.org/10.48550/arXiv.1606.06356;
[8] Le Guyader, H. „Cognitive Domain: A Sixth Domain of Operations.” Bernard Claverie; Baptiste Prébot; Norbou Buchler; François du Cluzel. Cognitive Warfare: The Future of Cognitive Dominance, NATO Collaboration Support Office, pp.3, 1-5, 2022, 978-92-837-2392-9. (hal-03635898);
[9] Palantir AIP | Defense and Military, 2023, https://www.youtube.com/watch?v=XEM5qz__HOU;
[10] George Grylls. „Ukraine Is Outflanking Russia with Ammunition from Big Tech”, 2023. augusztus 6., szak. news, https://www.thetimes.co.uk/article/ukraine-is-outflanking-russia-with-ammunition-from-big-tech-lxp6sv3qz;
[11] Medhat, W., Hassan, A., Korashy, H. „Sentiment Analysis Algorithms and Applications: A Survey”, Ain Shams Engineering Journal 5, sz. 4 (2014. december): 1109, https://doi.org/10.1016/j.asej.2014.04.011;
[12] „What Is Natural Language Processing? IBM”, https://www.ibm.com/topics/natural-language-processing. (Letöltve: 2023.1.4.);
[13] Sedova, K., McNeill, Ch., Johnson, A., Joshi, A. Wulkan, I. „AI and the Future of Disinformation Campaigns”, é. n., 17, https://doi.org/10.51593/2021CA011;
[14] Küçük, D., Can, F. „Stance Detection: A Survey”, ACM Computing Surveys 53. sz. 1 (2021. január 31.): https://doi.org/10.1145/3369026;
[15] Noble, W. S. „What Is a Support Vector Machine?”, Nature Biotechnology 24, sz. 12 (2006. december): https://doi.org/10.1038/nbt1206-1565;
[16] Goldstein, F. L., Findley, B. F. „Psychological Operations: Principles and Case Studies”, Air University. Press Maxwell Air Force Base, Alabama 1996., 5–7.;
[17] Goodfellow, I. et al. „Generative adversarial networks”, Communications of the ACM 63, sz. 11 (2020. 0 22.): 2–3, https://doi.org/10.1145/3422622;
[18] Creswell, A. et al. „Generative Adversarial Networks: An Overview”, IEEE Signal Processing Magazine 35, sz. 1 (2018. január): 53–65. https://doi.org/10.1109/MSP.2017.2765202;
[19] „Deepfake Presidents Used in Russia-Ukraine War”, BBC News, 2022. március 18., szak. Technology, https://www.bbc.com/news/technology-60780142. (Letöltve: 2023.8.6.);
[20] „ElevenLabs - Prime AI Text to Speech | Voice Cloning” https://beta.elevenlabs.io/. (Letöltve: 2023.4.19.);
[21] „Fake Trump Arrest Photos: How to Spot an AI-Generated Image”, BBC News, 2023. március 24., szak. US & Canada, https://www.bbc.com/news/world-us-canada-65069316. (Letöltve: 2023.8.6.);
[22] „What Is Natural Language Generation?”, Enterprise AI, https://www.techtarget.com/searchenterpriseai/definition/natural-language-generation-NLG.) (Letöltve: 2023.4.14.);
[23] Koubaa, A. „GPT-4 vs. GPT-3.5: A Concise Showdown” (TechRxiv, 2023. április 7.), https://doi.org/10.36227/techrxiv.22312330.v2;

[24] Kreps, S., McCain, R. M., Brundage, M. „All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation", Journal of Experimental Political Science 9, sz. 1 (2022. ed): 3–4, https://doi.org/10.1017/XPS.2020.37;
[25] Ferrara, E. et al. „The Rise of Social Bots", Communications of the ACM 59, sz. 7 (2016. június 24.): 96, https://doi.org/10.1145/2818717;
[26] „Fake Tweet,Chat & Facebook Status Generator | Prank Me Not" https://www.prankmenot.com/. (Letöltve: 2023.4.19.);
[27] Assenmacher, D. et al. „Demystifying Social Bots: On the Intelligence of Automated Social Media Actors", Social Media + Society 6, sz. 3 (2020. július): 5, https://doi.org/10.1177/2056305120939264;
[28] Hegedűs E., Hennel S. „Többdimenziós (multidomain) hadműveletek", Hadtudomány 30, sz. 2 (2020): 3–27, https://doi.org/10.17047/HADTUD.2020.30.2.3.

**JEGYZETEK**

1 A bot a „robot" szó rövidítése, egy olyan szoftver (automatizált program vagy szkript), amelyet egy meghatározott, ismétlődő feladat elvégzésére programoztak. (Pl: weboldalak bejárása, információk gyűjtése) A botokat gyakran használják a keresőmotorok (pl: a Google) annak érdekében, hogy feltérképezzék és indexeljék, strukturált módon járják be a weboldalakat, követve a linkeket, analizálva a tartalmat és rögzítve az információkat a keresőmotor adatbázisában.
2 A nagy nyelvi modell egyfajta mesterséges intelligencia algoritmus, amely mély tanulási technikákat és hatalmas adatkészleteket alkalmaz az új tartalom megértéséhez, összegzéséhez, generálásához és előrejelzéséhez.
3 A Deep Web (Darknet) az olyan weboldalakat foglalja magába, amelyeket nem lehet a hagyományos keresőkkel elérni**.** Az ilyen oldalakra csak ismert URL- vagy IP-címmel lehet belépni, azokat többszörös tűzfalak védik az illetéktelen behatolás ellen. „Clear Web"-nek, vagy tiszta webnek nevezik az internet általánosan ismert és használt felületét, amely a teljes internetnek csak a kisebb részét teszi ki.
4 A visszhangkamra olyan környezet, ahol az ember csak azokkal az információkkal vagy véleményekkel találkozik, amelyek tükrözik és megerősítik a sajátját. A visszhangkamrák téves információkat hozhatnak létre, és torzíthatják a személy perspektíváját, így nehézséget okoz az ellentétes nézőpontok mérlegelése és a bonyolult témák megvitatása terén. Visszhangkamrák bárhol előfordulhatnak, ahol információcsere történik, akár online, akár a való életben, de az interneten bárki gyorsan találhat hasonló gondolkodású embereket a közösségi médián keresztül. Emiatt a visszhangkamrák száma az online térben sokkal nagyobb.

# Honvédelmi alapismeretek




A magyar nyelvben hajdan a kadét szó hadapródot jelentett, majd a katonatiszti iskola növendékeit illették ezzel a régi kifejezéssel. A francia eredetű cadetből (ifjabb fiútestvér) a késő középkorban azon hagyomány alapján alakult ki a *fiatal tisztjelölt nemes katona* jelentése, hogy a francia nemesi családokban egykor a legidősebb fiú örökölte a birtokot, míg az ifjabba(ka)t katonai pályákra adták.

A Honvéd Kadét Program a hazai közoktatásban több mint egy évtizede meghonosodott fogalom. A középiskolákban elindított képzésforma arra kínál lehetőséget a fiataloknak, hogy érettségi vizsgát tegyenek a honvédelmi ismeretek tantárgyból. A honvédség iránt érdeklődő, hazájuk védelmére elkötelezett diákok heti 1, illetve 2 tanóra keretében az iskolában, valamint szabadidős foglalkozásokon szereznek személyes tapasztalatokat a magyar katonák mindennapi tevékenységéről. A tárgy oktatásának célja Magyarország biztonság- és szövetségi politikájának, a honvédelemmel összefüggő szabályoknak, a Magyar Honvédség felépítésének, technikai eszközeinek, a katonai értékek és hagyományok alapjainak megismerése, ezen keresztül a honvédelem iránti elkötelezettség növelése. A közérdeklődésre számot tartó, kereskedelmi forgalomban is kapható, *Honvédelmi alapismeretek* című akkreditált tankönyv a kadét programban résztvevő fiatalok számára készült. A tankönyv egyes fejezeteit a különböző területek vezető főtisztjei, elismert szakemberei írták, a kötetet Demeter József, Isaszegi János ny. vezérőrnagy és Kiscelli Piroska szerkesztette. Az alaki felkészítés, a hadtörténelmi alapismeretek, a túlélési és a haditechnikai ismeretek, a katonai testnevelés, a biztonságpolitikai és válságreagáló, valamint a hadijogi alapismeretek éppen úgy a tananyag részét alkotják, mint a lőelmélet. A diákok mindezek mellett számos, a hétköznapok során is hasznosítható tudásra tesznek szert, ezek közé tartoznak például az egészségügyi ismeretek és elsősegélynyújtás, a térképtan és a terepen való tájékozódás vagy akár az önvédelem.

A kötet felhasználóbarát tipográfiája segíti a tananyagban történő eligazodást. Az egyes fejezetek saját színcsíkot kaptak, és a témakörökhöz jól elkülönített feladatsorokat, kérdéseket, és rövid összefoglalást is kapcsoltak a szerkesztők. Az informatív szakszöveget színes és fekete-fehér fotók, grafikák, ábrák, táblázatok és térképek oldják. A tankönyv mellékletében az 1:50 000 méretarányú topográfiai térképek oktatótérképét veheti kézbe az érdeklődő olvasó. A kadét programban részt vevő diákok mellett, a könyvet minden középiskolás fiatalnak ajánljuk.

**A Zrínyi Kiadónál 2023-ban megjelent keménytáblás kötet terjedelme 554 oldal. 3700 Ft-os áron kapható a könyvesboltokban, illetve közvetlenül a Zrínyi Kiadótól 25%-os helyszíni kedvezménnyel 2775 Ft-ért. Cím: 1024 Budapest, Fillér utca 14., (tel.: 06 1-459-5373, e-mail: ugyfelszolgalat@hmzrinyi.hu), továbbá megrendelhető a shop.hmzrinyi.hu weboldalon is.** (R.A.)