

Mi az a nagy nyelvi modell? Hogyan állítja elő a szöveget – és miért hallucinál?

Szerző: Lengyel Zsolt, iTStudy Hungary Számítástechnikai Oktató- és Kutatóközpont

Lektor: Hartyányi Mária, iTStudy Hungary Számítástechnikai Oktató- és Kutatóközpont

Tanulási cél

Az olyan csevegőrobotok, mint a ChatGPT, a Claude vagy a Gemini, gördülékenyen és magabiztosan válaszolnak – akkor is, ha tévednek. A tananyag célja, hogy konkrét példákon keresztül elmagyarázza a nagy nyelvi modell (LLM) fogalmát, bemutassa, hogy hogyan állítja elő a szöveges válaszokat, és miért hozhat létre pontatlan, kitalált vagy félrevezető szövegeket. A tananyag végére a tanárok képesek lesznek meghatározni a token fogalmát, elmagyarázni a következő token előrejelzésének elvét, felismerni a hallucináció jellegzetes formáit és figyelmeztető jeleit, és megfelelő stratégiát és eszközöket alkalmazni az MI kimenet validálására.

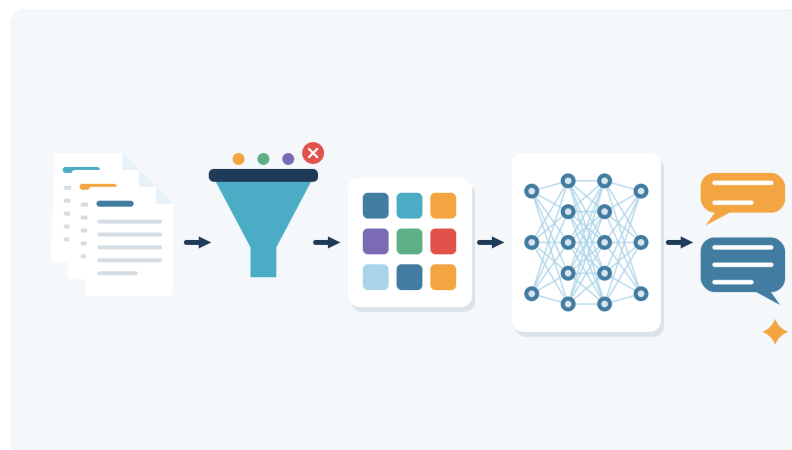
Bevezetés

2022. november 30-án jelent meg a ChatGPT első nyilvános verziója; öt napon belül egymillióan regisztráltak, és azóta a nagy nyelvi modellek az emberek mindennapi munkavégzésének eszközeivé váltak. Válaszaik jól szerkesztettek és magabiztosak. Éppen ezért veszélyesek a hibáik: a rossz válasz legtöbbször éppen olyan hihetőnek tűnik, mint egy jó válasz.

A tanárként nem kell ismerni a nagy nyelvi modellek mögött álló matematikai algoritmusokat, sok minden világossá válik, ha nagy vonalakban értjük működésüket: miért bánnak ilyen jól a nyelvvel, miért „találnak ki” időnként dolgokat, és miért szakmai kötelesség a kimeneteik ellenőrzése.

Hogyan készül egy nagy nyelvi modell?

A nagy nyelvi modell olyan MI-rendszer, amelyet hatalmas mennyiségű szövegen tanítottak, és fő képessége a szöveg értelmezése és előállítása. Azért „nagy”, mert több száz milliárd paramétert – a tanultakat tároló, hangolható számot – tartalmaz. Elkészítése négy fő lépésből áll:



A webes szöveg szűrésén, a „tokenizáláson” majd a tanításon át a csevegő asszisztensig: ezek a nagy nyelvi modellek elkészítésének fő lépései

Adatgyűjtés és szűrés

A fejlesztők letöltik a nyilvános web jelentős részét, és eltávolítják belőle a gyenge minőségű tartalmakat, az ismétlődéseket és a személyes adatokat; végül több tíz terrabájtnyi szöveg marad meg. A modell csak azt „tudhatja”, ami ebben a szűrt adathalmazban szerepelt – és csak egy bizonyos időpontig (ez a *tudáskorlát*, angolul knowledge cutoff).

Tokenizálás

A modell nem betűket vagy szavakat lát, hanem számokat, ún. *tokeneket*. A **tokenek** szótöredékekhez, szavakhoz vagy írásjelekhez, rendelt azonosító számok. Egy tipikus modell szótára mintegy százezer tokenből áll. Egy hosszú magyar szó gyakran több tokenre bomlik – ez az egyik oka annak, hogy a nyelvi modellek gyakran hibáznak, amikor meg kell számolniuk, hogy egy betű hányszor fordul elő egy szóban.

Előtanítás

A modell kap egy tokensorozatot, és megpróbálja megjósolni a következő tokenet, majd ellenőrzi, jól tippelt-e, és kicsit módosít a paraméterein. Ez billiószámra ismétlődik. Az eredmény egy rendkívül kifinomult „szövegkiegészítő” – de még nem asszisztens.

Utótanítás

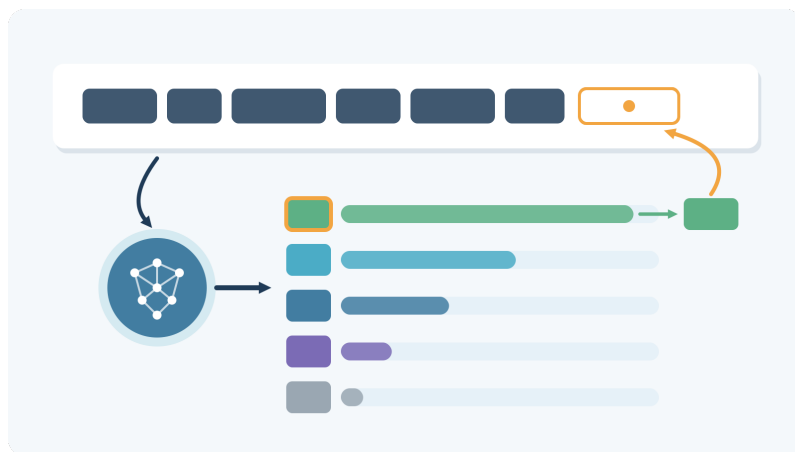
Ezután gondosan megírt kérdés–válasz példákön tanítják, így megtanul segítőkészen, udvariasan és használható formában válaszolni – és azt is, mikor kell nemet mondania. Végül emberek vagy értékelő modellek rangsorolják az alternatív válaszokat, így a modell megtanulja, melyeket részesítse előnyben.

A válasz születése: tokenről-tokenre

Amikor üzenetet küldünk, a modell minden lehetséges következő tokenre kiszámítja, **mennyire valószínű**, hogy az következik az addigi szöveg után. Kiválaszt egyet, hozzáfűzi a szöveghez, és kezdi előlről – tokenről tokenre. Amit a képernyőn „gépelésnek” látunk, az ezeknek a tokeneknek az egymás utáni érkezése. Vegyük egy műhelyi mondat elejét:

„Mielőtt sarokcsiszolóval elvágja az acélcsövet, mindig vegye fel a ...”

Az olyan folytatások, mint a *védőszemüveget*, a *védőkesztyűt* vagy a *hallásvédőt*, nagyon valószínűek, mert hasonló mondatok sokszor előfordultak a tanítóadatban; a *papucsot* szinte kizárt. A modell nem szabályt keres ki, hanem statisztikailag valószínű módon folytatja a szöveget.



A következő token előrejelzése: a modell pontozza a lehetséges folytatásokat, és egyiküket hozzáfűzi a szöveghez

Ebből három fontos következmény adódik:

A válaszok változhatnak.

A modell nem mindig a legvalószínűbb tokenet választja, ezért ugyanarra a kérdésre különböző válaszokat adhat.

A modellnek tokenekre van szüksége a „gondolkodáshoz”.

Amikor csak az eredményt kellett megadnia, egy modell azt válaszolta, hogy három gép, amely egyenként 12, 8, illetve 6 óra alatt végezne egy munkával, együtt 3,43 óra alatt készül el vele. Amikor azt kértük, hogy lépésről lépésre vezesse le a megoldást, eljutott a helyes 2 óra 40 percre. A „gondolkodó” modellek a válasz előtt ilyen gondolatmenetet generálnak.

A modell nem úgy emlékszik, mint az ember.

Minden választ abból a szövegből állít elő, amelyet az adott pillanatban megkap (lásd M1-T1-MLC5); friss információhoz csak eszközökkel – webes kereséssel, kód futtatással, csatolt fájlokkal – jut.

A **tokenizáció** az a folyamat, amely során a szöveget feldarabolják kis egységekre, majd minden egyes egységhez hozzárendelnek egy egyedi azonosító számot. A GPT-4 esetében például ez a „szótár” – vagyis az összes lehetséges token listája – 100 277 elemből áll. Az üzenetünk tehát a modellhez érve egy számsorozat lesz – például [15339, 1917, 264, 3918].

A válasz generálása is ugyanígy működik, csak fordított sorrendben: a modell tokeneket ad vissza egymás után, majd azokat egy egyszerű program visszaalakítja olvasható szöveggé. Mi folyamatos gépelésként látunk a képernyőn, de valójában tokenek érkeznek egymás után.

Próbáljuk ki!

A tokenizálást a Tiktokenizer (tiktokenizer.vercel.app) nevű interaktív oldalon lehet kipróbálni. Ha beírunk egy szöveget, látni fogjuk, hogy azt GPT 4o modell hogyan darabolja fel tokenekre, melyik szótöredék lesz önálló token és melyik nem.

Miért hallucinál az MI?

Gyakran elhangzik az a megtévesztő mondat, hogy „Az MI hazudik.” Valóban előfordul, hogy téves választ kapunk egy kérdésre, ez azonban a modell betanításának módjából fakad. Mivel a tanítópéldákban a kérdéseket magabiztos válaszok követik, a modell akkor is válasz formájú szöveget állít elő, ha nincs megbízható tudása.



Mi emberek gyakran hallucinálunk (Forrás: ChatGPT)

Szerencsésebb, hogy a jelenségre a szakirodalomban a **hallucináció** elnevezés terjedt el, még akkor is, ha a mögöttes folyamat alapvetően más, mint az emberi hallucináció: az MI-ből ugyanis hiányzik a **képzelet** képessége, nem „lát” vagy „képzeli” olyasmit, ami nincs jelen, hanem a tanult mintázatok alapján állít elő valószerű, de időnként téves információt.

A nagy nyelvi modellek (LLM-ek) esetében hallucinációnak nevezzük az olyan, MI által generált tartalmat, amely gördülékenynek, összefüggőnek és hihetőnek tűnik, valójában azonban téves, vagy nem támasztható alá megbízható forrásokkal. Ennek oka, hogy a modell nem a valósággal összevetve ellenőrzi állításait, hanem a tanult statisztikai mintázatok alapján a legvalószínűbb szövegfeldolgozást generálja.

Az OpenAI kutatói azt is kimutatták, hogy a szokásos tanítási és értékelési módszerek jobban jutalmazták a találgatást, mint a bizonytalanság beismerését (Kalai és mtsai, 2025). Az újabb, fejlettebb modellek azonban egyre gyakrabban mondják, hogy „*nincs erről információ*”, de a probléma nem szűnt meg.

A hallucinált válasz statisztikailag ugyanolyan, mint a valódi – ugyanaz a hangneme, ugyanaz a formátum –, csak éppen nem igaz.

Jellegzetes példák

Kitalált tények és személyek

Amikor egy nem közismert személyről kérdezünk, előfordul, hogy a csevegőrobot részletes életrajzot ad, foglalkozással, művekkel, évszámokkal, amely teljes egészében véletlenszerűen állított össze.

Kitalált hivatkozások és jogszabályok

Egy munkavédelmi szabály jogalapjára rákérdezve a modell tökéletes formában idézhet egy rendeletet számmal, dátummal és paragrafussal – amely nem létezik, vagy egészen mást szabályoz. 2023-ban egy New York-i bíróság bírságot szabott ki azokra az ügyvédekre, akik a ChatGPT által kitalált bírósági döntéseket idéző beadványt nyújtottak be (Mata kontra Avianca ügy).

Hihető, de hibás műszaki értékek

Egy csavar meghúzási nyomatéka, egy vezeték-keresztmetszete, a maghőmérséklet vagy a tűrésérték helyesnek tűnhet, mert a szokásos tartományba esik – mégis eltérhet a gyártó vagy a szabvány által előírt értéktől. A számítások ugyanígy könnyen félre csúszhatnak!

Csendben „kijavított” szakszavak.

A ritka szakkifejezéseket a modell elírásnak tekintheti. A szerző egyik próbájában a *bokrászás* vadászati szakszót a modell szó nélkül *bokszolásra* cserélte, és a válasz teljesen rossz irányba ment.

Kihagyások és elfogultság.

Arra a kérdésre, hogyan vágjunk el egy fémcsövet, az MI helyesen javasolhat sarokcsiszolót vágókoronggal – a szem-, kéz- és hallásvédelemről viszont egy szót sem szól. A modellek a tanítóadatban lévő sztereotípiákat is újratermelhetik.



A hallucinált válasz hivatalosnak és teljesnek tűnhet – egészen addig, amíg valaki meg nem nézi a mögötte álló forrást

A hallucináció valószínűbb, ha a téma ritka, helyi vagy friss (hazai jogszabályok, céges eljárások, a múlt havi változások), ha pontos számokra, nevekre vagy hivatkozásokra van szükség, és ha a kérdés téves feltevést tartalmaz.

Tedd fel ezt a kérdést a ChatGPT-nek:

Csoportosítsd a főbb szakterületeket és feladattípusokat aszerint, hogy mennyire megbízhatóak a ChatGPT válaszai. Készíts táblázatot a becsült megbízhatósági tartománnyal, a fő hibakockázattal és az ellenőrzés szükséges mértékével. A százalékokat csak akkor kezeld mérési eredményként, ha megbízható kutatás támasztja alá; egyébként jelöld őket becslésként.

A válaszként kapott táblázat fontos információt ad a tudatos MI használathoz: például azt, hogy programozási feladatoknál a hallucináció valószínűsége alacsony (5-15%), de aktuális információkkal kapcsolatban webes forrás nélkül lehet elérheti a 60%-ot!

A hibák felismerése és az válasz validálása

A gördülékenység és a magabiztosság nem jelenti azt, hogy a válasz helyes is. A jó kérdés nem az, hogy „jól hangzik-e”, hanem az, hogy „honnan tudom, hogy helyes?” Mi a teendő?

Ellenőrizzük a forrásokat!

Kérjünk forrást, majd nyissuk is meg: létezik-e a dokumentum, és valóban azt írja-e, amit a válasz állít? Vessük össze a választ hiteles forrásokkal: a gyártói kézikönyvvel, a szabvánnyal, a tankönyvvel, a tantervvel.

Kérdezzünk vissza!

„Válaszod mely részeiben vagy bizonytalan?”, „Honnan származik ez az érték?”, „Mi változna, ha...?” Tegyük fel ugyanazt a kérdést új beszélgetésben vagy egy másik modellnek is: az egymásnak ellentmondó válaszok figyelmeztető jelek.

Használtassunk eszközöket!

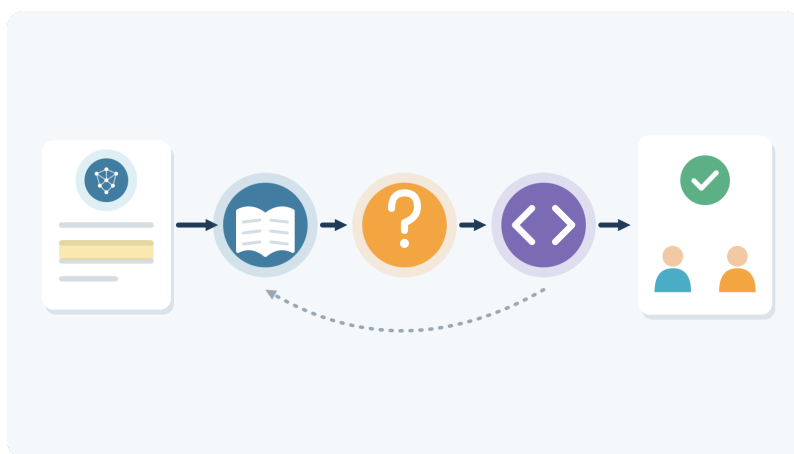
Friss tényekhez kérjünk webes keresést, matematikai, műszaki számításokhoz kérjük, hogy „használgon kódot”. Ekkor az MI programot fog írni a feladat megoldásához, és azt lefuttatva adja meg a választ, és ezzel jelentősen csökken a hibázás lehetősége.

Adjunk neki saját forrást!

Csatoljuk a vonatkozó tankönyvfejezetet vagy adatlapot, és adjuk utasításba: „Kizárólag a csatolt dokumentum alapján válaszolj; ha a válasz nincs benne, jelezd.”

Validáljuk, mielőtt az órára bevisszük!

Saját, vagy egy kolléga szakmai tudására támaszkodva ellenőrizzünk minden tényt, számot és biztonsági utasítást, amelyre a tanulók támaszkodni fognak.



Ellenőrizzük a forrást, kérdezzünk vissza, használjunk eszközöket – és csak ezután vigyük a tartalmat a tanórára

Öt ellenőrző kérdés az órai felhasználás előtt

1. Mely állítások tények, számok, nevek vagy hivatkozások – és mindegyiket ellenőriztem?
2. Van benne biztonsági szempontból kritikus elem? Egyezik a gyártói adatokkal és a hatályos előírásokkal?
3. Mi hiányzik belőle (védőfelszerelés, előfeltételek, kivételek)?
4. Megfelelő a szint, a szakszókincs és a kontextus a tanulóim számára?
5. Aláírnám-e ezt a tananyagot a saját nevemmel?

Pedagógiai és etikai kérdés

A hallucináció nem csupán technikai kellemetlenség. Ha a tanár kiosztja az MI által generált feladatlapot, a diákok a tanári tekintély miatt hibátlanak tekintik, és mivel kezdők a szakterületen, egyébként sem fogják észrevenni a hibákat. A tanműhelyben egy hibás érték vagy egy kimaradt biztonsági lépés a berendezés károsodásához vagy balesethez vezethet. A tanulók által bevésett hibákat nehéz később kijavítani.

A felelősség mindig a tanárnál marad: az MI elkészítheti a vázlatot, de a tanár validálja. Ugyanilyen fontos, hogy a tanulókat is megtanítsuk erre: kérdezzenek rá, honnan származik az információ, vessék össze megbízható forrásokkal, és kezeljék az MI-t kiindulópontként, ne végső tekintélyként. Ez a DigComp 3.0-ban leírt MI- és információs műveltség része.

Egy kis történelem

Az írás megszületése óta az emberiség állandó törekvése, hogy az írásban felhalmozott tudást megőrizze, továbbadja és hozzáférhetővé tegye a következő nemzedékek számára.

Az írás megjelenése, nagyjából i. e. 3200 körül Mezopotámiában, először tette lehetővé, hogy a tudás tartósan fennmaradjon az emberi emlékezeten túl. Az ékírásos agyagtáblák azonban még csak egy szűk, képzett írnokréteg számára voltak hozzáférhetőek. A papirusz- és pergamentekercsek hordozhatóbbá tették az írást, létrejöhettek az ókor nagy könyvtárai, de a felhalmozott tudáshoz továbbra is kevesen fértek hozzá.

A középkorban a tekercset fokozatosan felváltotta a lapozható kódex. A kolostorokban kézzel másolt könyvek elkészítése rendkívül időigényes és költséges volt, ezért a könyvek és az általuk őrzött tudás ritka értéknek számítottak. A 15. század közepén a könyvnyomtatás hozott alapvető fordulatot: a szövegek nagy példányszámban sokszorosíthatóvá és egyre olcsóbbá váltak. A könyvek terjedése, a nyilvános könyvtárak, az általános oktatás és az írástudás növekedése révén, egyre többen tudtak közvetlenül hozzáférni az emberiség által felhalmozott tudáshoz.

A 20. század végén a számítógép és az internet újabb ugrást jelentett. A digitalizálással és a hálózatok elterjedésével hatalmas mennyiségű információ vált szinte azonnal elérhetővé. A korábbi fizikai korlátok jelentős része eltűnt: ma egyetlen telefonról olyan tudásforrások érhetők el, amelyekhez korábban könyvtárakra, utazásra vagy intézményi hozzáférésre volt szükség. Ezzel azonban új probléma jelent meg: már nem elsősorban az információ hiánya,

hanem annak bősége, megtalálása, értelmezése és megbízhatóságának megítélése vált kihívássá.

A nagy nyelvi modellek ebben a folyamatban a hozzáférés demokratizálásának újabb lépcsőfokát jelentik. Az internet hozzáférést adott az információhoz; az LLM-ek ezen túl természetes nyelvű hozzáférést kínálnak annak feldolgozásához is. Nem szükséges pontosan tudnunk, melyik könyvben, adatbázisban vagy milyen szakkifejezéssel keressünk: hétköznapi nyelven kérdezhetünk, visszakérdezhetünk, kérhetünk rövidebb vagy részletesebb magyarázatot, példát, összefoglalást vagy fordítást. Ugyanaz a téma elmagyarázható kezdőnek vagy szakembernek, egyszerűbb nyelven vagy más nyelven is.



Az emberi tudás útja: az agyagtáblától a nyelvi modellekig (Forrás: ChaGPT)

Ebben rejlik az LLM-ek egyik legfontosabb demokratizáló hatása: *nemcsak a tudásforrásokhoz vezető utat rövidítik le, hanem csökkentik a megértéshez szükséges nyelvi, szakmai és keresési küszöböt is.* A korábbi technológiák egyre több ember elé tették oda a könyvet, majd az információt; a nyelvi modellek arra tesznek kísérletet, hogy az információt az ember kérdéséhez és tudásszintjéhez igazítva tegyék hozzáférhetővé.

Ez lényeges különbséget is jelent a korábbi tudáshordozókhoz képest. A könyv rögzített szöveget őriz, az LLM viszont a tanulás során kialakított mintázatok alapján minden kérdésre új választ állít elő. Rendkívül rugalmas, de nem tévedhetetlen: pontatlan vagy akár teljesen hibás választ is adhat, és a válasz eredete sem mindig követhető vissza egyértelmű forráshoz. *Minél könnyebbé válik tehát a tudáshoz való hozzáférés, annál fontosabbá válik a kapott válasz kritikus értékelése és szükség esetén ellenőrzése.*

Így az agyagtáblától a könyvön és az interneten át az LLM-ekig ugyanaz a hosszú folyamat rajzolódik ki: *a tudás egyre több ember számára, egyre kisebb akadályokkal válik hozzáférhetővé.* Az LLM-ek újdonsága, hogy a demokratizálódás következő lépése már nem csupán az információ eléréséről, hanem annak megértéséhez való hozzáférésről is szól.

Tanulságok tanároknak

- A nagy nyelvi modell tokenről tokenre állítja elő a szöveget, a tanítóadat mintázatai alapján hihető folytatásokat választva.
- Nem keresi ki a tényeket: a magabiztos válasz is lehet hibás.

- Különösen óvatosaknak kell lenni a számokkal, hivatkozásokkal, jogszabályokkal, ritka szakszavakkal, helyi vagy friss témákkal és a biztonsági utasításokkal.
- Ellenőrizzük a forrásokat, kérdezzünk vissza, kérjük, hogy használjon böngészőt és/vagy programkódot, adjunk saját forrást.
- Az órai felhasználás előtt validáljunk! Tanítsuk meg a tanulókat is az MI-kimenetek ellenőrzésére.

Hivatkozások

- Kalai, A. T., Nachum, O., Vempala, S. S. és Zhang, E. (2025): Why Language Models Hallucinate. arXiv:2509.04664.
- Huang, L. és mtsai (2024): A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. ACM Transactions on Information Systems. DOI: 10.1145/3703155.
- Mata v. Avianca, Inc., 678 F. Supp. 3d 443 (S.D.N.Y. 2023).
- Cosgrove, J. és Cachia, R. (2025): DigComp 3.0 – European Digital Competence Framework. 5. kiadás, Az Európai Unió Kiadóhivatala.
- Lengyel Zs. (2026): Bevezetés a mesterséges intelligencia fogalomrendszerébe; Konkrét alkalmazások tartalomgenerálásra. VITA képzési prezentációk.

Linkek

- Tiktokenizer – így bomlik tokenekre a szöveg: tiktokenizer.vercel.app
- Egy nagy nyelvi modell interaktív 3D-s szemléltetése: bbycroft.net/llm
- A. Karpathy: Deep Dive into LLMs like ChatGPT (videó, angol nyelvű): youtube.com/watch?v=7xTGNNLPyMI

Szerzői jog © 2026 VITA Konzorcium

Ez a mű a Creative Commons Nevezd meg! – Így add tovább! 4.0 Nemzetközi Licenc (CC BY-SA 4.0) feltételei szerint újra felhasználható és átdolgozható.



Az Európai Unió finanszírozásával. A kifejtett nézetek és vélemények kizárólag a szerző(k)éi, és nem feltétlenül tükrözik az Európai Unió vagy a Nemzeti Ügynökség álláspontját.