

# NYELV, ÉRTELEM, MESTERSÉGES INTELLIGENCIA AZ EMBERI GONDOLKODÁS TÜKRÉBEN

LANGUAGE, MEANING, ARTIFICIAL INTELLIGENCE  
IN THE MIRROR OF HUMAN THOUGHT  
JEZIK, SMISAO, VEŠTAČKA INTELIGENCIJA  
U OGLEDALU LJUDSKOG RAZMIŠLJANJA

CSUVIK VIKTOR

PhD, az informatikai tudományok doktora

Szegedi Tudományegyetem, Természettudományi és Informatikai Kar  
Informatika Doktori Iskola, Mesterséges Intelligencia

MALI Labs, technológiai vezető

viktor.csuvik@malilabs.com

ÖSSZEFOGLALÓ • A mesterséges intelligencia az elmúlt években a közbeszéd központi témájává vált, különösen a generatív modellek, mindenekelőtt a ChatGPT 2022-es megjelenése óta. A kezdeti *hype* és apokaliptikus jóslatok után 2025-re jóval tisztábban látjuk, mire képesek a nagy nyelvi modellek – és mire nem. Tanulmányom három fő üzenete: (1) a mai MI nem egyetlen „mesterséges elme”, hanem sok kicsi, szűk feladatra tanított program ökoszisztémája; (2) a generatív modellek nem varázslatosak, hanem statisztikai mintázatfelismerésre épülő, transzformeralapú nyelvi modellek, amelyek tokenek valószínűségét becslik; (3) ezek az eszközök nem helyettesítik, hanem kiegészítik az emberi munkát és gondolkodást. A cikk röviden áttekinti a gépi tanulás és mélytanulás alapfogalmait, a nyelvi modellek működését és fejlődését 2018–2025 között, gyakorlati felhasználási területeiket, valamint a *prompt engineering* alapjait. Külön fejezet foglalkozik a korlátokkal, a hallucináció jelenségével, a felelős használattal és a visszaélések kockázataival. A dolgozat fő állítása, hogy a generatív MI megértése ma alapvető digitális kompetencia, jövője pedig nem a gépekben, hanem az emberi döntésekben dől el.

**KULCSSZAVAK • mesterséges intelligencia, generatív MI, nagy nyelvi modellek (LLM), transzformer, gépi tanulás, mélytanulás, természetes nyelvfeldolgozás (NLP), prompt engineering, MI-ökoszisztéma, hallucináció, digitális kompetencia, oktatás és MI**

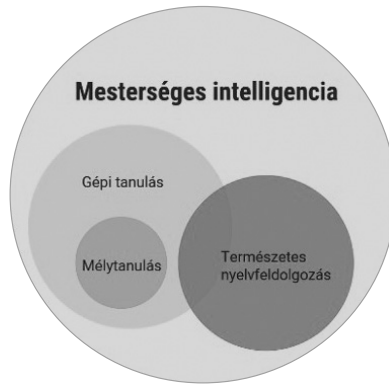
## BEVEZETÉS

A mesterséges intelligencia (MI) fogalma az elmúlt években a közbeszéd egyik meghatározó témájává vált. 2022 végétől, a generatív modellek – mindenekelőtt a ChatGPT – megjelenésével a nagyközönség számára is láthatóvá vált, milyen lehetőségek rejlenek a nagy nyelvi modellekben. Bár a kezdeti időszakot erős bizonytalanság és sokszor apokaliptikus hangvételű jóslatok kísérték, 2025-re jóval tisztábban látjuk, mire képesek valójában ezek a rendszerek – és mire nem. A félelem helyét egy józanabb, kritikus, de alapvetően pragmatikus megközelítés váltotta fel: a mesterséges intelligencia nem világmegváltó csodafegyver, de nem is fenyegető, kontrollálhatatlan „mesterséges elme”, hanem számos kisebb, jól körülhatárolt feladatot ellátó program együttese.

Az emberi gondolkodás és a gépi számítás kapcsolatának keresése nem új keletű törekvés: már a 17–18. századi filozófusok és matematikusok is felvetették a kérdést, vajon leírható-e az értelem valamilyen formális rendszerrel. A mesterséges intelligencia mint tudomány azonban csak a 20. század közepén – az 1940-es évek kutatási eredményeire építve –, 1956-ban vált önállóvá. Bár fiatal diszciplínának számít, mára rendkívül szerteágazó és dinamikusan fejlődő területté nőtte ki magát, amely a nyelvészettől az orvostudományon át az oktatásig számos iparágat és kutatási irányt érint.

A legújabb fejlesztések középpontjában a generatív mesterséges intelligencia áll. Ezek a rendszerek nem gondolkodnak, nem értenek, és semmiféle belső „tudatossággal” sem rendelkeznek: működésük statisztikai mintázatok felismerésén alapul. Pontosabban: nagy mennyiségű szöveges adatból tanulják meg, milyen szó vagy mondatrész következik a legnagyobb valószínűséggel. Ez az egyszerűnek tűnő mechanizmus azonban meglepően összetett, és a gyakorlatban rendkívül hatékony.

Jelen írás célja, hogy bemutassa röviden a legfontosabb szempontokat a mai mesterségesintelligencia-rendszerekkel kapcsolatban. Az előadás három alappillére – nem egy mesterséges elme, hanem sok kicsi program; nem varázslat, hanem statisztika; nem helyettesít, hanem kiegészít – iránytűként szolgál majd a téma értelmezéséhez. Ezek mentén igyekszünk megérteni, mit tekinthetünk MI-nek, hogyan működnek a nagy nyelvi modellek, és milyen szerepet tölthet be ez a technológia az oktatásban és a nyelvészeti gyakorlatban.



1. ábra. A mesterséges intelligencia jelen dokumentumban tárgyalt részterületei

A mesterséges intelligencia világa hatalmas és folyamatosan bővül, de a jelenlegi generatív modellek megértéséhez csak bizonyos részterületek relevánsak. Ezekre a következő fejezetekben részletesen kitérünk.

#### GÉPI TANULÁS (MACHINE LEARNING)

A gépi tanulás olyan algoritmusok fejlesztésével foglalkozik, amelyek képesek a tapasztalatból tanulni, és teljesítményüket idővel javítani anélkül, hogy a fejlesztők minden viselkedési szabályt előre meghatároznának. Míg egy hagyományos szoftver esetében a programozó pontosan előírja, hogy adott bemenetre milyen kimenetet kell előállítani (például egy számológép minden műveletének eredménye előre definiált), a gépi tanulási rendszerek esetében a fejlesztő inkább a képességet határozza meg, amelyet a modellnek el kell sajátítania.

A „szabályokat” tehát nem a programozó kódolja le, hanem maga a modell tanulja meg a bemeneti adatok alapján. Ez egyben azt is jelenti, hogy a gépi tanulás eredményeit a betanításhoz használt adatok minősége alapvetően meghatározza – ez a generatív modelleknél különösen fontos szempont.

#### MÉLYTANULÁS (DEEP LEARNING)

A mélytanulás a gépi tanulás azon területe, amely összetett problémák megoldásához többrétegű, úgynevezett mély neurális hálózatokat használ. A neurális hálózat egy olyan matematikai modell, amely az idegsejtek működéséből merít inspirációt, és rétegekbe szervezve képes egyre absztraktabb mintázatokat felismerni.

Minél több példán tanul egy ilyen háló, annál pontosabban képes végrehajtani a feladatát – legyen szó képfelismerésről, beszédfeldolgozásról vagy éppen szöveggenerálásról. A mély neurális hálózatok előnye, hogy a bemeneti adatokból maguktól képesek magasabb szintű mintázatokat és absztrakciókat kialakítani. Hátrányuk ugyanakkor, hogy betanításuk hatalmas mennyiségű adatot és számítási kapacitást igényel, valamint működésük belső logikája sokszor nehezen átlátható.

Mindezek ellenére napjainkban – különösen a nagy nyelvi modellek (LLM-ek) esetében – ez a megközelítés vált a mesterséges intelligencia kutatásának és alkalmazásának meghatározó módszerévé.

### TERMÉSZETES NYELVFELDOLGOZÁS (NATURAL LANGUAGE PROCESSING – NLP)

A természetes nyelvfeldolgozás az MI és a nyelvészet határterülete, amely az emberi nyelv számítógépes reprezentációjával, feldolgozásával és (korlátozott) értelmezésével foglalkozik. Interdiszciplináris tudományág, amely magában foglalja a nyelvészeti elméleteket, a számítógépes algoritmusokat és a statisztikai módszereket.

Az NLP célja, hogy a gépek képesek legyenek:

- szövegekből információt kinyerni,
- azonosítani a nyelvi jelenségeket,
- kategorizálni és rendszerezni dokumentumokat,
- valamint – a generatív modellek esetében – új szöveget létrehozni.

Fontos azonban hangsúlyozni, hogy a jelenlegi nyelvi modellek nem „értik” a szöveget emberi értelemben: működésük statisztikai mintázatokon alapul, nem pedig valódi jelentésreprezentációkon.

### FELHASZNÁLÁSI PÉLDÁK

Sokan úgy vélik, hogy a mesterséges intelligenciával először a ChatGPT-hez hasonló rendszerek megjelenésekor találkoztak, valójában azonban a gépi tanulás már több évtizede láthatatlanul jelen van a mindennapjainkban.

Az MI működik a keresőmotorok rangsorolásai mögött, ajánlórendszerekben, spamszűrőkben, útvonaltervezőkben, fordítóprogramokban és okostelefonos fotóalkalmazásokban. A generatív AI csak annyiban új, hogy most először találkozik a nagyközönség közvetlenül egy nyelvalapú MI-rendszerrel.

A nagy nyelvi modellek sokrétűen felhasználhatók – megfelelő körülményekkel és korlátaik ismeretében. Néhány jellemző példa:

- **E-mail válaszok, sablonok készítése:** különösen ismétlődő kérdések esetén.
- **Ügyfélszolgálati chatbotok:** jól meghatározott témakörben nagy pontossággal képesek válaszolni.
- **Ötletgenerálás:** projektötletek, kampányszlogenek, alternatív megfogalmazások.
- **Jegyzetelés és összefoglalás:** megbeszélések, előadások vagy hosszabb dokumentumok lényegi tartalmának összegzése.
- **Nyelvi ellenőrzés:** stilisztikai javaslatok, helyesírási hibák javítása, szövegátfogalmazás.

A generatív modellek működésének megértéséhez azonban kulcsfontosságú kérdés, hogy valójában micsoda egy nagy nyelvi modell, és hogyan állít elő szöveget. Ezzel a következő fejezetben foglalkozunk részletesen.

## NEM VARÁZSLAT – STATISZTIKA

A nagy nyelvi modellek (LLM-ek) működésének alapját a transzformer architektúra adja – ez a megközelítés mára a természetes nyelvfeldolgozás legmeghatározóbb technológiájává vált, és olyan rendszerekben működik, mint a ChatGPT vagy a 2023 után megjelent legtöbb generatív modell. A transzformer valójában egy speciális felépítésű neurális hálózat, amely képes egyszerre több bemeneti elemet is figyelembe venni, és ezáltal hatékonyan feldolgozni a szekvenciális adatokat, például a szöveget.

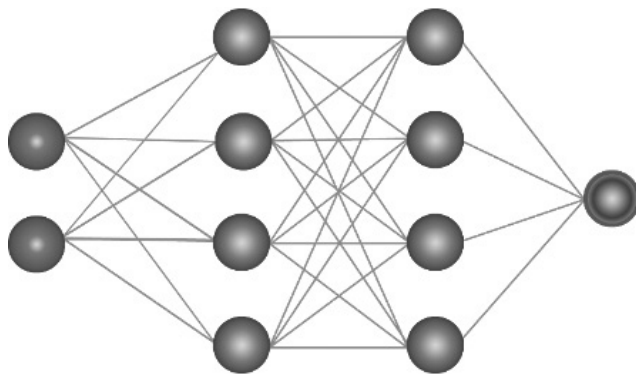
## NEURÁLIS HÁLÓZAT: HONNAN INDULTUNK, HOVÁ JUTOTTUNK

A ma „varázslatként” érzékelt technológiák alapjait már több mint fél évszázada lefektették. Az első jelentős mérföldkő az 1957-es **perceptron modell**, amely egyetlen mesterséges neuront próbált matematikailag leírni, az emberi idegsejtek működését imitálva. A perceptron gyakorlati alkalmazhatósága akkor még korlátozott volt, ezért a kutatás hosszú időre háttérbe szorult.

A valódi előrelépést a **többrétegű neurális hálózatok** hozták. Ezekben a neuronok különböző rétegekbe szerveződnek, és tanulás közben a hálózat finomhangolja az egyes kapcsolatok „súlyait”. A modern modellek paramétereinek száma – legyen az néhány száz millió, több milliárd vagy akár több száz milliárd – lényegében ezeknek a súlyoknak a mennyiségét jelzi.

A lényeg: minél több neuron és kapcsolat áll rendelkezésre, annál összetettebb mintázatokot képes a háló megtanulni, de ehhez arányosan több adat és számítási kapacitás is szükséges.

Az 1990-es és 2000-es években több specializált architektúra jelent meg (konvolúciós hálók, autoencoderek stb.), majd a 2017-ben publikált transzformer modell teljesen új irányt adott a gépi tanulásnak. Ez lett a nagy nyelvi modellek alapja.



2. ábra. Teljesen összekapcsolt neurális hálózat. Minden egyes „pötty” egy-egy neuront szimbolizál. Az adat a hálón balról jobbra halad, a két bal oldali neuron a bemenet, a jobb oldali a kimenet, a közbensőket pedig rejtett rétegeknek nevezzük

#### TRANSZFORMER MODELL: A NAGY NYELVI MODELLEK MOTORJA

A transzformer szakít a korábbi neurális hálózatok lineáris, rétegsorrendben előrehaladó felépítésével. Két fő komponensből áll – **kódoló** (encoder) és **dekódoló** (decoder) blokkokból –, és működésének kulcsa az úgynevezett **figyelemmechanizmus** (attention).

A figyelemmechanizmus teszi lehetővé, hogy a modell meghatározza, a bemeneti szöveg mely elemei „fontosak” egy adott pillanatban. Ez különösen nagy előny hosszú szövegek, összetett mondatok vagy távoli szavak közötti kapcsolatok esetén.

Vegyük példaként a következő két mondatot:

*Vizet öntött a kancsóból a pohárba, amíg az tele nem lett.*

*Vizet öntött a kancsóból a pohárba, amíg az ki nem ürült.*

Mindkét mondatban az *az* mutató névmás más-más tárgyra utal: az elsőben a pohárra, a másodikban a kancsóra. Emberek számára ez könnyen értelmezhető, ám a régebbi modellek számára komoly kihívást jelentett.

A transzformerek azonban az attention segítségével képesek felismerni az ilyen típusú, távoli elemek közti összefüggéseket, és ennek megfelelően súlyozni a szavakat.

A transzformer sikertörténete azonban nem egyetlen ötlet eredménye: a figyelemmechanizmus mellett **kulcsszerepe van az internetes szövegtörzsek óriási méretének, a modern hardverek megnövekedett számítási teljesítményének és a hatékonyabb tanítási algoritmusoknak is.** Ezek együtt tették lehetővé, hogy a 2020-as évekre megszülessenek a mindennapi használatra is alkalmas, emberközeli szövegalkotó rendszerek.

### SZÖVEGGENERÁLÁS = VALÓSZÍNŰSÉGEK BECSLÉSE

A nagy nyelvi modellek működése első ránézésre „intelligens” vagy akár kreatív folyamatnak tűnhet, valójában azonban **valószínűségi becslések sorozatán alapul.** A jelenleg (2025-ben) elérhető LLM-ek túlnyomó többsége úgy állít elő szöveget, hogy minden egyes lépésben kiszámítja: mi a következő token (szó, szótag, szó vagy írásjel), amely a legnagyobb valószínűséggel követi az eddigieket.

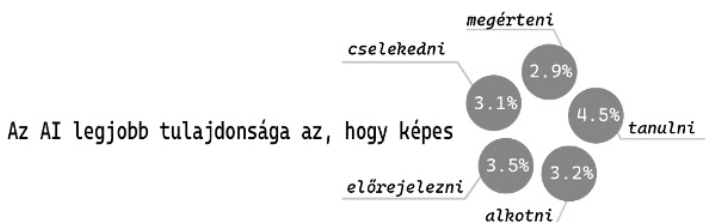
A modell tehát nem szabályokat alkalmaz, nem „érti” a mondatot, és nem tervezi meg előre a teljes szöveget – egyszerűen mindig a **következő elem legvalószínűbb jelöltjei közül választ.**

Például ha a bemenet így hangzik:

„Tegnap séta közben láttam egy ugató \_\_\_\_.”

A modell a korábbi szövegek statisztikai mintázatai alapján nagy valószínűséggel a „kutyát” választja kiegészítésként. A háttérben lényegében egy rangsor készül azokról a szavakról/tokenekről, amelyeket a modell valaha látott – ezek mindegyikéhez rendel egy valószínűségi értéket, és a generálás során ezekből választ.

Természetesen sokszor nem ennyire egyértelmű a helyzet: egy mondat folytatására több, egymástól nagyon különböző folytatás is valószínű lehet, és a modell ezek közül választhat – ez okozza a kreativitásnak tűnő jelenséget (VASWANI et al., 2017).



3. ábra. A következő szó becslése az előzőek alapján: sok esetben nem egyértelmű, melyik választás a legjobb

### HONNAN SZÁRMAZNAK EZEK A VALÓSZÍNŰSÉGEK?

A legegyszerűbb megközelítés az lenne, ha pusztán megszámolnánk, milyen gyakran fordulnak elő egyes szavak egy szövegben, de ez természetesen nem elég: így nem kapnánk értelmes mondatokat.

A gyakorlatban ezért nem önmagában a szavak előfordulási gyakoriságát figyeljük, hanem a szavak egymásutánosságát. Ezek a modell által megtanult mintázatok az úgynevezett: **n-gramok** (két, három, illetve több szóból álló sorozatok), illetve a modern modellekben a jóval összetettebb **szókapcsolati és szövegkörnyezeti** viszonyok.

A jelenlegi transzformeralapú modellek nem csupán néhány szót látnak előre, hanem a teljes kontextust figyelembe vehetik, és ennek alapján becsülik meg a valószínűségeket. A mondat folytatását tehát egy finomhangolt statisztikai rendszer egyetlen lépése határozza meg – és ezt ismétli meg újra és újra, amíg létre nem jön a teljes szöveg.

### MIÉRT NEM SZAVAKAT, HANEM TOKENEKET HASZNÁLUNK?

Az ember számára természetes, hogy a nyelvet szavakra bontja, a modellek azonban nem így működnek. A nagy nyelvi modellek a szöveget **tokenekre** bontják – ezek lehetnek:

- teljes szavak (ritka),
- szóelemek (gyakoribb),
- ragok, toldalékok,
- vagy akár gyakori karaktersorozatok.

A tokenek használata sok szempontból előnyös:

- Hatékonyabb: nem kell minden ragozott alakot külön megtanulni.
- Rugalmasabb: jól kezeli az összetett szavakat és új kifejezéseket.
- Nyelvfüggetlenebb: a módszer bármilyen írott nyelvre adaptálható.

A tokenek előállításának folyamata a tokenizálás, és modelleszaládonként eltérhet. A ChatGPT-ben alkalmazott tokenizáló például körülbelül 1000 token  $\approx$  750 szó arányban dolgozik.

### (NAGY) NYELVI MODELLEK

A nyelvi modell a természetes nyelv statisztikai leírása: egy olyan rendszer, amely nagy mennyiségű egy- vagy többnyelvű szövegtörzs alapján megtanulja megjósolni, milyen token (szó, szó vagy írásjel) következik legnagyobb valószínűséggel egy adott szószorozat után. A nagy nyelvi modellek (Large Language Models – LLM-ek) ennek a megközelítésnek a méretben és teljesítményben megnövelt változatai.

A szakirodalomban nincs éles határ arra, hogy mikortól „nagy” egy nyelvi modell – a megnevezés általában azt jelzi, hogy a modell **paramétereinek száma több száz milliótól a több tíz vagy száz milliárdig terjed**, és ennek megfelelően képes összetett nyelvi mintázatokat megtanulni. A méret ugyan nem minden, de a mai modellek teljesítményét nagymértékben meghatározza.

Fontos különbséget tenni a **nyelvi modellek** és a rájuk épülő **alkalmazások** között. A ChatGPT például egy **termék**, egy felhasználói felület és szolgáltatáscsomag, amely mögött különböző nyelvi modellek működtek és működnek (GPT-3.5, GPT-4, GPT-4o, majd 2025-től több specializált modellvariáns). Hasonló módon a Google Gemini, az Anthropic Claude, a Meta Llama vagy a Mistral modelljei is különböző platformokon és termékekben jelennek meg.

Ezek a modellek különböző célokra optimalizáltak, de közös bennük, hogy hatalmas adatmennyiségen tanítják őket, és tanításukhoz jelentős számítási erőforrásokra van szükség. A pontos költségek üzleti titoknak számítanak, de iparági becslések szerint egy korszerű, csúcsteljesítményű modell betanítása **egy bizonyos év technológiai szintjén akár több millió dolláros nagyságrendet is elérhet** – jellemzően 1–10 millió dollár, de a legnagyobb modellek esetében még ennél is nagyobb lehet.

A nagy nyelvi modellek jelentősége nem csupán méretükben rejlik, hanem abban, hogy általánosan használható eszközökké váltak: képesek szöveget generálni, kérdésekre válaszolni, hosszú dokumentumokat összefoglalni, kódot írni, vagy éppen nyelvi elemzést végezni. Ezek a képességek tették lehetővé, hogy a generatív MI 2022 óta ilyen látványos szerepet kapjon a mindennapi életben és az oktatásban is.

## A NAGY NYELVI MODELLEK FEJLŐDÉSE (2018–2025)

A nagy nyelvi modellek fejlődése 2018 óta rendkívül gyors ütemben zajlik. Bár az egyes modellek között vannak architekturális különbségek, a trend jól látható: **nagyobb paraméterszám, több adat, hosszabb kontextusablak, multimodalitás és egyre hatékonyabb következő-token-becslés.** Az alábbi lista a legfontosabb mérföldköveket foglalja össze (RADFORD et al., 2019).

- **GPT-1 (2018)** – 117M paraméter; az első transzformeralapú, előre tanított szövegeneráló modell. A következő token becslésének alapja itt született meg.
- **GPT-2 (2019)** – 1.5B paraméter; nagyobb skála, első valóban koherens hosszabb szövegek. 1024 tokenes kontextusablak.
- **GPT-3 (2020)** – 175B paraméter; megjelent a zero/one/few-shot tanulás paradigmája, amely lehetővé tette a sok feladatra történő finomhangolás nélküli használatot (BROWN et al., 2020; HOFFMANN et al., 2022).
- **GPT-3.5 (2022)** – tovább finomított modell, hosszabb kontextussal (4096 token); a ChatGPT első generációinak alapja.
- **ChatGPT (2022 →)** – nem modell, hanem termék: felügyelt tanúlással és RLHF-fel finomhangolt GPT-3.5 (majd később GPT-4).
- **GPT-4 (2023)** – multimodális, stabilabb, 8k/32k kontextusablak; számos benchmarkon emberközeli teljesítmény.
- **GPT-4 Turbo / 4o (2024)** – olcsóbb, gyorsabb, hosszabb kontextus (128k+), egységes multimodalitás.
- **GPT-4.1 (2024 vége)** – megbízhatóbb utasításkövetés, jobb hosszú távú koherencia, integrált hang/kép és videókezelés.
- **GPT-5 (2025)** – a legújabb generáció: nagyobb kontextusablak (akár több százezer token), javított következtetési képességek, stabilabb logikai feladatmegoldás, egységes multimodális feldolgozás és fejlettebb biztonsági rétegek. Nem általános intelligencia, de jelentős előrelépés komplex feladatokban.



4. ábra. A GPT nyelvi modelleszalád 2018-as megjelenése óta átforgalmazta a gépi modellekkel való interakciónkat

Más jelentős modellek ugyanebben az időszakban:

- **Google PaLM → Gemini (2023–2025)** – a Gemini modellek (Gemini Ultra / Gemini 1.5 / Gemini 2) nagy kontextust, multimodalitást és integrált keresési képességeket kínálnak.
- **Anthropic Claude (2023–2025)** – a Claude 2/3 modellek kiemelkedően jók hosszú dokumentumok elemzésében; a hangsúly a biztonság és a pontosságon van.
- **Meta Llama (2023–2025)** – Llama 2 → Llama 3 → Llama 3.1; nyílt forrású család, rendkívül gyors közösségi adaptációval és finomhangolt változatokkal.
- **Mistral (2023–2025)** – Mixtral és Mistral Large modellek, hatékony MoE (Mixture-of-Experts) architektúrával, gyors és költséghatékony működéssel.
- **Open-source modellek** – Falcon, Vicuna, Dolly, Zephyr, OpenAssistant; lokálisan is futtatható, közösségi fejlesztésű modellek.

2018 óta a nagy nyelvi modellek fejlődése meghatározóvá vált a mesterséges intelligencia területén. A modellek egyre nagyobbak, de még fontosabb trend az, hogy egyre jobban képesek hosszú kontextust kezelni, több modalitást integrálni, és stabilabb, emberközelibb interakciót biztosítani. Mindezek ellenére továbbra is szűk, statisztikai rendszerek, amelyek nem rendelkeznek valódi megértéssel vagy tudatossággal – viszont kiváló eszközök a szöveg-generálás, az összefoglalás, az oktatás és számos nyelvi feladat támogatására (OPENAI 2023; CHOWDHERY et al., 2022).

#### NEM EGY MESTERSÉGES ELME, SOK KICSI PROGRAM

A mesterséges intelligenciáról szóló közbeszédben máig erősen él a gondolat, hogy az AI valamiféle egységes, emberhez hasonló „digitális elme”, amely minden kérdésre tudja a választ, és bármilyen feladatot el tud látni. Ez a kép azonban félrevezető. A jelenlegi technológiák nem egy mindentudó mesterséges agyat alkotnak, hanem sok kicsi, egymástól különálló, pontosan meghatározott feladatokra tervezett rendszer együttműködéséből állnak.

Modern MI-rendszereink – legyen szó képfeldolgozásról, beszédfelismerésről, navigációról vagy szövegalkotásról – mind külön tanított, külön működési elveken alapuló modellek. Egy önvezető autó például egyszerre tucatnyi algoritmust használ: külön rendszert a gyalogosok felismerésére, más a sávkövetésre, más a tárgyak távolságának becslésére, és megint más a döntéshozatalra. Ezek nem egyetlen „tudat” részei, hanem specializált eszközök, amelyek csak együtt képesek komplex viselkedést létrehozni.

A generatív AI – köztük a legnépszerűbb nagy nyelvi modellek – különösen jól illusztrálja ezt a jelenséget. Bár sokszor úgy tűnik, mintha egy átfogó, általános intelligenciával rendelkeznének, valójában egyetlen, szűken meghatározott feladatot optimalizálnak: adott bemenetre a legvalószínűbb kimenetet kell előállítaniuk. A rendszerek „mindentudásának” hatását az emberi pszichológia is erősíti: hajlamosak vagyunk intelligenciát feltételezni minden olyan entitás mögött, amely folyékonyan kommunikál velünk.

A fejezet célja, hogy eloszlassa a „mesterséges elme” mítoszát, és bemutassa: a jelenlegi AI-technológia alapvetően **moduláris, feladatspecifikus és statisztikai működésű**. E megértés nélkülözhetetlen nemcsak a technológia helyes használatához, hanem a társadalmi, oktatási és szabályozási viták tisztánlátásához is.

### AZ MI LÉPCSŐI: SZŰK FELADATOKTÓL A HIPOTETIKUS SZUPERINTELLIGENCIÁIG

A mesterséges intelligenciáról folytatott beszélgetésekben gyakran összemósódnak azok a kategóriák, amelyek valójában jól elkülöníthető szinteket jelölnek. A jelenlegi rendszerek túlnyomó többsége az úgynevezett *szűk* vagy *gyenge MI* kategóriájába tartozik. Ezek a modellek egyetlen, pontosan meghatározott feladat megoldására készülnek: képfelismerésre, beszédértésre, fordításra, útvonaltervezésre, vagy éppenséggel arra, hogy egy szöveges kérdésre koherens választ adjanak. Egy ilyen rendszer adott feladatban akár felül is múlhatja az embert, de tudását nem képes átvinni más problémákra – egy sakkprogram például tökéletesen játszhat, de ettől még nem fog tudni autót vezetni vagy egy tárgyat megfogni.

A következő, jelenleg még elméleti szint az **általános MI (AGI)**. Ez olyan rendszert jelölne, amely különböző területek között is képes átvinni a tudását, új helyzetekben rugalmasan alkalmazkodik, és széles körű problémákra talál megoldást – vagyis bizonyos értelemben az emberi intelligencia működéséhez hasonló képességekkel rendelkezik. Bár a modern nagy nyelvi modellek néha úgy tűnnek, mintha közelednének ehhez, működésük alapvetően továbbra is szűk: egyetlen inputtípusból (szövegből) dolgoznak, és ott is statisztikai mintázatokat követnek.

A harmadik lépcsőfok a **szuperintelligencia**, amely nemcsak reprodukálná, hanem minden területen meghaladná az emberi kognitív képességeket. Ez a kategória jelenleg teljes egészében hipotetikus, és inkább filozófiai, etikai vagy sci-fi viták témája, semmint mérnöki valóság. Bár fontos elképzelni és vizsgálni az ilyen forgatókönyveket, a mai technológiától nagyon távol állnak.

E három lépcső tiszta megkülönböztetése segít abban, hogy a mesterséges intelligenciáról szóló vitákat és elvárásokat a megfelelő helyen kezeljük.

A jelen AI-ja nem egy univerzális „gondolkodó gép”, hanem sok szűk, egymást kiegészítő rendszer együttese – az általános és a szuperintelligencia pedig továbbra is a jövő lehetséges, de nem létező állomásai.

#### A MESTERSÉGES INTELLIGENCIA RENDSZEREINEK VALÓDI FELÉPÍTÉSE

A mai mesterségesintelligencia-rendszerek működését akkor értjük meg igazán, ha nem egységes „digitális elmeként”, hanem **moduláris gépezetként** tekintünk rájuk.

Egy korszerű chatbot például nem egyetlen modelltől áll. A háttérben külön fut a szövegértési modul, a biztonsági szűrés, a hosszú távú kontextuskezelés, a hangfelismerő rendszer, a beszéd szintézis és a nyelvi modell, amely a tényleges választ generálja. Ezek egymástól függetlenül tanultak más-más adatkészleten és más célra. Ugyanez a modularitás jellemző a sokkal bonyolultabb rendszerekre is: egy önvezető autó tipikusan **tucatnyi különálló MI-modult** használ. Az egyik csak a közlekedési táblákat azonosítja, a másik a gyalogosokat detektálja, a harmadik az autó környezetét térképezi fel, míg egy külön rendszer dönt a fékezésről, gyorsításról vagy sávváltásról. Bár a kívülről látható viselkedés egységesnek tűnik, valójában sok apró, specializált modell összehangolt működéséből jön létre.

Ennek oka magából a mélytanulás természetéből fakad. A neurális hálóak ugyanis nem általános problémamegoldók, hanem statisztikai mintázatokat tanuló rendszerek, amelyek teljesítménye erősen kötődik a tanítás során látott adatoktól és feladattól. Minden modell arra optimalizál, hogy egy konkrét problémára minél jobb választ adjon. A szövegfordításra tanított háló nem fog képet felismerni; egy orvosi diagnosztikai modell nem fog kódot generálni; egy képosztályozó nem fog beszédet érteni. Hiába tűnnek „okosnak”, képességeik valójában szűk csatornába rendeződnek. Ezért nem létezik olyan neurális háló, amely egyszerre képes lenne a teljes emberi képességek kálát reprodukálni – a rendszer mindig ahhoz ért, amire tanították, és csak ahhoz.

A generatív mesterséges intelligencia eszközei ezt a specializációt különösen jól példázzák. A nyelvi modellek, a képgenerátorok, a beszéd felismerők és a beszéd szintetizátorok mind külön technológiák, még ha egy szolgáltatásban össze is vannak csomagolva. Egy nyelvi modell szöveget állít elő, egy képgeneráló vizuális pixeleket hoz létre, egy hangmodell pedig akusztikus hullámformát. Bár egy alkalmazás – mint például a ChatGPT – ezt egységes élményként mutatja a felhasználó felé, a háttérben **három-négy különböző rendszer** dolgozik: más a szöveget értelmező algoritmus, más a képeket generáló modell, és megint más az, amelyik hangzó beszédre alakítja a kimenetet.

Ezek összehangolt működése teremti meg azt az érzetet, mintha egyetlen „általános intelligencia” válaszolna.

A valóság azonban ennek az ellenkezője: minél fejlettebb egy MI-rendszer, annál tagoltabbá és specializáltabbá válik. A mai technológia hatékonyságát éppen az biztosítja, hogy sok kicsi, jól definiált feladatra tanított modell együttműködésével jön létre a komplex működés, nem pedig egyetlen univerzális mesterséges elme révén.

#### AZ MI ÖKOSZISZTÉMA, NEM EGYETLEN TUDAT

A mesterséges intelligenciáról alkotott laikus képzetben erősen él az az elképzelés, hogy az AI egyetlen, mindent átfogó tudatként működik: egy digitális agy, amely „érti” a világot, és az emberhez hasonló módon gondolkodik. A technológiai valóság azonban egészen más. A modern MI-rendszerek inkább hasonlítanak egy összetett ökoszisztémára, amelyben számtalan, eltérő feladatokra tervezett komponens együttműködik.

Egyes részek képeket értelmeznek, mások szöveget elemeznek, megint mások hangot dolgoznak fel, és e modulok eredményeit egy magasabb szintű rendszer fűzi össze koherens kimenetté. Ez a működés nem egyetlen „tudat” jele, hanem mérnöki struktúra, amelyben minden modell azt teszi, amire tanították. A koherencia illúzióját az teremti meg, hogy mindez a felhasználó számára egy folyamatnak látszik.

A modell-ökoszisztéma azonban nem csak technikai értelemben sokszereplős. A jelenlegi MI-képességek mögött külön adatforrások, külön szabályrendszerek és külön felhasználási területek állnak. Egy nyelvi modell például nem rendelkezik vizuális vagy fizikai tapasztalattal; egy képfelismerő nem tud bölcs tanácsadó lenni; egy navigációs algoritmus pedig nem képes megírni egy prezentációt. Az MI-rendszerek csak akkor válnak hasznossá, amikor ezeket a specializált modulokat összehangoljuk, és egy nagyobb rendszer részeként működtetjük.

Ez a kép közelebb áll a valósághoz, mint a sci-fi által inspirált elképzelés, amely szerint az AI egyetlen, emberhez hasonló tudattal rendelkező entitás lehetne. A jelenlegi technológia nem hordoz egységes, belső reprezentációt a világról, és nem rendelkezik olyan általános megértéssel, amely egységes identitáshoz vagy tudathoz vezetne. Amit intelligenciaként látunk, az valójában mintázatok felismerése, statisztikai becslések, és sok apró komponens pontos együttműködése.

Az MI-ökoszisztéma megértése praktikus szempontból is fontos. Segít elhelyezni az elvárásokat: felismerni, mire alkalmas egy modell, és mire nem. Csökkenti a technológia körüli túlzott félelmeket és irreális reményeket. És

iránymutatást ad abban is, hogyan lehet felelős módon integrálni ezeket a rendszereket oktatásba, iparba vagy közigazgatásba.

Ha az AI-t nem misztikus, homogén elmeként, hanem sok együttműködő rendszer ökoszisztémájaként látjuk, pontosabban értjük meg erősségeit, korlátait és valódi potenciálját. Ez az a szemlélet, amelyre a következő évek technológiai döntései épülhetnek.

### NEM HELYETTESÍT, HANEM KIEGÉSZÍT

A generatív mesterséges intelligenciával a legtöbben webes felületeken keresztül találkozunk (ChatGPT, Gemini, BingChat, Llama-alapú chatbotok stb.). Ezek a felületek regisztráció után azonnal használhatók, és jól működnek egyszerűbb, egyedi kérések esetén. Ha azonban nagy mennyiségű, ismétlődő vagy automatizálható feladatot szeretnénk megoldani, akkor az API-alapú integráció, illetve a bővítmények (pluginok) jelentik a hatékony megközelítést. Fő használati módok:

- **Webes interfészek:** beszélgetős felület, alapvetően kézi használatra. Bizonyos rendszerek (BingChat, Gemini) képesek interneten keresni, mások zárt környezetben működnek.
- **Pluginok / bővítmények:** speciális feladatokhoz – pl. matematikai számítás (Wolfram), táblázatkezelő műveletek, foglалások, adatbázis-lekérdezés. A generatív modellek gyenge pontjait (pl. matek, logikai következtetés, valós idejű adatok) ellensúlyozzák.
- **API-k:** alkalmazásokba épített MI funkciókhoz, tömeges szöveggeneráláshoz, automatizált elemzéshez. Itt lehetőség van saját munkafolyamatok, rendszerek, oktatási vagy vállalati megoldások integrációjára.

**KÖLTSÉGEK.** A webes felületek gyakran ingyenesek, de a nagyobb teljesítményű modellek, hosszú kontextus vagy API-használat fizetős. Nagy nyelvi modelleket *nulláról* betanítani extrém költséges (millió dolláros nagyságrend), míg egy meglévő modellt *finombangolni* már lényegesen olcsóbb és gyakoribb megoldás. Speciális, érzékeny vagy iparági adatok esetén fontos szempont lehet a házon belüli modell futtatása is.

### MIRE ÉRDEMES HASZNÁLNI A GENERATÍV MODELLEKET?

A generatív modellek akkor a leghasznosabbak, amikor jól meghatározott, nyelvi jellegű feladatokat kell gyorsan, nagy mennyiségben vagy jó kiindulási pontként megoldani. Kifejezetten hasznos:

- **Vázlatkészítés, ötletgenerálás, brainstorming:** kreatív kiindulópontok, alternatív megfogalmazások.
- **Szövegalkotás és átfogalmazás:** e-mailek, bejelentések, sablonok, bemutatkozók.
- **Összefoglalás:** hosszabb szövegek, cikkek, megbeszélésjegyzetek gyors kivonata.
- **Nyelvi ellenőrzés:** helyesírás, stílus, koherencia javítása.
- **Oktatási támogatás:** példafeladatok generálása, magyarázatok készítése, tananyagvázlatok.
- **Egyszerű, rutinszerű ügyfélszolgálati feladatok:** gyakran ismétlődő kérdések megválaszolása.
- **Adatátalakítás, strukturált kimenet előállítás:** listák, táblázatok, JSON-formátum, jegyzetek.

Ezekben a modellek gyorsak, produktívak, és jelentősen növelik a hatékonyságot – kiegészítik, nem helyettesítik a szakmai döntéshozatalt.

#### MIRE NEM ÉRDEMES HASZNÁLNI A GENERATÍV MODELLEKET?

A generatív modellek nem általános mesterséges intelligenciák, ezért több terület kifejezetten kockázatos vagy alkalmatlan számukra. Nem ajánlott:

- **Előrejelzésre** (gazdasági, időjárási, politikai stb.): a modellek nem a valóságot, hanem a lehetséges folytatásokat jósolják.
- **Osztályozásra, diagnosztikára vagy döntés automatizálására:** hibáik gyakran „buták”, nincs valódi logikai garancia.
- **Magas kockázatú területeken:** jogi döntések, orvosi diagnózis, pénzügyi jóváhagyások.
- **Tényellenőrzésre:** a modellek nem igazságot mondanak, hanem valószínű mondatokat.
- **Adatbiztonságot érintő helyzetekben:** érzékeny, bizalmas információk feltöltése harmadik fél szolgáltatásába kockázatos.
- **Valós idejű vagy nagy pontosságú számításokra:** matematika, hálózati címképzés, kriptográfia – ezeknél több a tévedés veszélye.
- **Fizikai feladatok irányítására:** gépkezelés, járművezetés, robotos műveletek – nem erre valók.

A generatív modellek leginkább nyelvi és kreatív támogatásra alkalmasak. Kiváló eszközök a produktivitás növelésére, de nem alkalmasak kritikus döntéshozatalra vagy valóságos világmodellezésre. Nem helyettesítik az

embert, hanem kiegészítik – megfelelően, a megfelelő feladatokra használva rendkívül hatékonyak.

## PROMPT ENGINEERING

A *prompt engineering* lényege az, hogy úgy fogalmazzuk meg a bemenetet (promptot), hogy a modelltől a lehető legjobb, legbiztonságosabb és legjobban használható választ hozzuk ki. Ez igaz szöveges (ChatGPT, Claude, Gemini), képgeneráló (DALL-E, Midjourney) és más generatív modellekre is. A jelenlegi rendszerek még igénylik a tudatos kérdezést, de a trend az, hogy idővel egyre „megbocsátóbbak” lesznek a rosszul megfogalmazott promptokkal szemben – ettől függetlenül az alapszintű prompt-műveltség hosszú távon is fontos digitális készség marad.

A prompt tipikusan három elemből áll:

- **Instrukció:** mit csináljon a modell („Fogalmazd át...”, „Magyarázd el...”, „Írj vázlatot...”).
- **Kontextus:** milyen szövegből, szerepből, célközönségnek dolgozik (pl. „nyelvtanároknak”, „7.-es tanulóknak”).
- **Elvárt forma:** milyen szerkezetű kimenetet várunk (lista, táblázat, e-mail, JSON, rövid összefoglaló stb.).

Például:

„Képzeld el, hogy középiskolai magyar nyelv- és irodalomtanár vagy. Foglald össze a következő szöveget 5 pontban, és írd mellé 3 vitaindító kérdést a diákoknak: .....

**Few-shot prompting** esetén néhány példát is adunk a promptban, hogy „megtanítsuk” a modellt, milyen mintát kövessen. Ugyanaz a logika, mint a GPT-3 few-shot tanulásánál: *nem megérti* a feladatot, hanem felismeri a mintát, és folytatja.

A **dokumentum-beágyazás** szempontjából segít, ha a háttér-információt elkülönítve adjuk meg (pl. idézőjelek vagy backtickek között), és külön, világos kérdést fogalmazunk meg hozzá. Ezzel a modell belső reprezentációjában (embedding) jobban „helyére kerül” a szöveg, és relevánsabb választ kapunk.

Nagyon fontos a **strukturált kimenet kérése** – a modellek rengeteg táblázatot, listát, JSON- és HTML-struktúrát „láttak” a tanulás során, ezért az ilyen formák gyakran:

- átláthatóbbak,
- könnyebben továbbfeldolgozhatók (pl. Excel, más szoftverek),
- kevesebb utólagos szerkesztést igényelnek.

Például:

„Add meg a választ táblázatban, 'fogalom' és 'magyarázat' oszlopokkal.” „JSON formátumban válaszolj: { 'explanation': '...', 'answer': '...' }.”

Komplexebb feladatoknál – számolás, tervezés, kódértelmezés, oktatási magyarázat – jól működik a **gondolatmenet** (chain-of-thought) jellegű instrukció (WEI et al., 2020):

„Lépésről lépésre gondold végig a feladatot, majd a végén külön sorban írd le a végső választ.”

Ez különösen a nagyobb modelleknél hoz látványos javulást, kisebb modelleknél kevésbé hatékony. A legújabb modellekben ez tipikusan integrálva van. Hátránya, hogy hosszabb a válaszidő (a modell „gondolkodik” – valójában részekre bontja a feladatot), illetve több tokent kell generálnia, így drágább is. Az ilyen módon üzemelő modellekre hivatkozunk gyakran „Thinking” (gondolkodó) modellként.

A prompt engineering ugyanakkor **biztonsági kérdés** is. A „prompt injection”, „jailbreak” típusú technikák célja, hogy rávegyék a modellt olyan válaszokra, amelyeket alapértelmezés szerint nem adhatna ki (pl. sértő tartalom, belső instrukciók, titkos információk). Ez egyrészt mutatja, hogy a rendszerek **nem tökéletesen védettek**, másrészt figyelmeztetés felhasználóknak és fejlesztőknek egyaránt:

- ne bízunk abban, hogy ami egy modellbe bekerül, az soha nem kerül vissza máshoz,
- ne küldjünk be érzékeny, azonosításra alkalmas vagy üzletileg kritikus adatot nyilvános modellekbe,
- tudatosan tervezzük meg a „rejtett promptokat” (system prompt, háttérinstrukciók), mert ezek manipulálhatók.

Gyakorlati szempontból néhány jó gyakorlat generatív modellekhez:

- **Szerepkiosztás:** „Úgy válaszolj, mint egy középiskolai nyelvtanár / szerkesztő / szigorú bíráló / laikus magyarázó.”

- **Hangnem és célközönség megadása:** „barátságos, szakmai, röviden, B2-es angol szinten, 12 éves gyerekeknek érthetően”.
- **Konkrét feladat és korlátok:** „max. 10 sorban”, „csak felsorolás”, „ne adj új példákat, csak fogalmazd át”.
- **Nyelvi preferencia tudatosítása:** ahol lehet, érdemes tudni, hogy a modellek teljesítménye általában angolul a legjobb, mert ezen a nyelven láttak a legtöbb adatot.

Összességében a prompt engineering nem „varázslat”, hanem józan, céltudatos instrukcióadás: minél világosabban írjuk le, **ki beszél, milyen kontextusban, mi a feladat és milyen formában kérjük a választ**, annál használhatóbb lesz a kimenet. A modell attól még nem gondolkodik, de jobban a mi céljaink szolgálatába állítható (BUBECK et al., 2023).

#### KORLÁTOK ÉS KRITIKA – MIÉRT KELL ÓVATOSNAK LENNÜNK?

A nagy nyelvi modellek használóinak különösen fontos, hogy tisztában legyenek ezek működésével és korlátaival – éppen azért, mert első ránézésre annyira „emberinek” és varázslatosnak tűnnek. A velük folytatott beszélgetés sokszor teljesen természetesnek hat, és könnyű elhinni, hogy „valódi értelem” áll a válaszok mögött. Emberként ráadásul erős a torzításunk, hogy ami magabiztosan és szépen hangzik, azt igaznak is gondoljuk. A nyelvi modellek pedig szinte mindig magabiztosan fogalmaznak – kételyek nélkül.

A legfontosabb, amit érdemes megjegyezni: **ezek a rendszerek nem „igazságmodellek”, hanem nyelvi modellek.** Arra optimalizálták őket, hogy **valószínűnek tűnő szöveget** állítsanak elő, nem pedig arra, hogy mindig helyes választ adjanak. Matematikai értelemben a feladatuk az, hogy a következő token valószínűségét becsüljék – ebben rendkívül jók, és ez önmagában is óriási technológiai eredmény. Az, hogy a válaszok gyakran tényszerűen is helytállóak, inkább a tanító adatbázisok (például a Wikipédia) minőségéről tanúskodik, nem pedig a modellek „igazságérzetéről”.

A modellek további alapvető korlátja, hogy statikus tanító adatokból tanulnak:

- egy adott időpontig bezárólag rendelkeznek ismeretekkel,
- a friss, 2024–2025-ös eseményekről csak akkor tudnak, ha valamilyen módon hozzáférnek az internethez,
- még böngészési képesség mellett is előfordul, hogy rosszul értelmezik a forrásokat, vagy kitalálnak („hallucinálnak”) részleteket.

Ez különösen problémás ott, ahol aktuális, pontos információra lenne szükség (például jogszabályok, friss kutatási eredmények, napi hírek esetén). Ilyenkor a modell szöveget fog gyártani akkor is, ha nem rendelkezik megbízható adattal – és ezt ugyanolyan magabiztosan fogja előadni, mint amikor „véletlenül” épp igaza van.

### HALLUCINÁCIÓ – AMIKOR A MODELL MAGABIZTOSAN TÉVED

A mesterséges intelligenciában hallucinációnak nevezzük azt a jelenséget, amikor a modell teljesen magabiztos, de nincs mögötte valós adat. Egyszerű példa: ha a modell nem rendelkezik megbízható adattal egy cég bevételéről, akkor is „legyárthat” egy hihetőnek tűnő számot, és úgy adja elő, mintha biztos információ lenne. Ugyanez történik gyakran tudományos hivatkozásoknál is: a modell megtanulta az idézetek formátumát, felismeri a szerzők neveit, témáit, de a generált cikkek és könyvek jelentős része egyszerűen nem létezik.

Hasonló a helyzet a matekkal: a modell nem logikusan „számol”, hanem a látott példák alapján próbálja kitalálni, mi következik egy „ $2 + 3 = \dots$ ” formájú kifejezés után. Ha a tanító adatokban sok a helyes példa, jó eséllyel 5-öt ír – de komplexebb feladatoknál gyakran csak „jóformán odaillő” számokat találhat. Innen nézve az is érthető, miért olyan törékenyek a matematikai és formális feladatmegoldások.

A hallucinációk miatt minden tényállítást kritikusan kell fogadni, különösen ha:

- jogi, orvosi, pénzügyi, mérnöki döntésről van szó,
- hivatkozásokat, tudományos forrásokat generál a modell,
- olyan információt ad, amit nehezen tudunk függetlenül ellenőrizni.

Előfordulhat, hogy egy adott kérdésben **nehezebb ellenőrizni** a modell válaszát, mint ha magunk keressünk forrásokat – ezt fontos felismerni, mielőtt rutinból ráhagyunk kritikus döntéseket.

### VISSZAÉLÉS ÉS FELELŐSSÉG

A ChatGPT-hez és más nagy nyelvi modellekhez hasonló rendszerek képesek: meggyőző szöveget írni, stílust utánozni, érzelmeket és véleményeket imitálni.

Ez kettős élű fegyver:

- egyrészt rengeteg hasznos alkalmazást tesz lehetővé (oktatás, jegyzetelés, ötletelés, kódolási segítség, szövegszerkesztés),
- másrészt alkalmas álhírek, manipulatív tartalmak, hamis „szakértői” vélemények tömeges gyártására is.

Éppen ezért kulcsfontosságú, hogy a felhasználók – különösen tanárok és nyelvészek – ne csak használják, hanem tanítsák is a kritikus hozzáállást:

- a modell állításait **ellenőrizni kell**,
- a források meglétét (és minőségét) **számon kell kérni**,
- és világossá kell tenni, hogy a **modell nem helyettesíti a szakmai felelősséget**.

A nagy nyelvi modellek tehát nem gondolkodnak, nem értenek, nincsenek kétségeik, és nem vállalnak felelősséget – ez a felhasználó feladata. Nem helyettesítik, hanem kiegészítik az emberi gondolkodást.

A kérdés az, hogy mennyire okosan, tudatosan és kritikusan élünk ezzel az eszközzel.

## ZÁRÓGONDOLATOK

Nehezen állja meg az ember, hogy egy ilyen gyorsan fejlődő területet ne valamilyen nagy ívű jóslattal zárjon, de az elmúlt évek tapasztalata éppen arra tanít, hogy az MI jövőjét előre megmondani meglehetősen hálátlan feladat. A ChatGPT és a generatív modellek látványos sikere úgy söpört végig a világon, hogy valójában nem hoztak radikálisan új technikai áttörést – inkább azt mutatták meg, milyen hatalmas ereje van annak, ha a kutatási eredmények széles közönséghez jutnak el.

A mesterséges intelligencia története mindig is ciklikus volt: MI-nyarak és MI-telek váltották egymást, a túlzott várakozások és a kiábrándulás időszakai. Jelenleg kétségtelenül nyár van: forró, intenzív, mindenki próbál valamilyen módon MI-t építeni a rendszereibe. Ám a szakértők többsége egyetért abban, hogy az általános mesterséges intelligenciához (AGI) még több komoly tudományos lépés szükséges. Ahogy a villanyt sem a gyertya tökéletesítéséből találták fel, úgy a mai nyelvi modellek skálázása sem fog egyik napról a másikra „emberi értelemmé” válni.

Valószínűbb, hogy a *hype* lecsengése után a generatív modellek csendes, de tartós segédeszközzé válnak: beépülnek az oktatásba, a kutatásba, az üzleti folyamatokba – nem mint mindenhatók, hanem mint produktivitást növelő társak. Még a legnagyobb modellek esetében is tudjuk, hogy mi történik

a belsejünkben: statisztikai mintázatok, súlyok és valószínűségek dolgoznak, nem kognitív folyamatok. Charles Sutton találó megjegyzése ma is igaz: „...az általános tanulság az, hogy az elmék tényleges tartalma óriási mértékben, jövátelhetetlenül összetett...”

Az előttünk álló évek tele vannak nyitott kérdésekkel: szabályozás, kutatásetika, monopolizáció, a hamis tartalmak terjedése, a tanulómodellek átláthatósága. Ezekre nem könnyű válaszokat adni, de épp emiatt fontos, hogy értsük a technológiát, amely egyre több döntés háttérében jelenik meg.

A generatív MI gyakorlati felhasználásának listája szinte végtelen, és talán ez a legizgalmasabb benne: a kreativitásunk határozza meg, mire válik képessé. Ha tudatosan és felelősen használjuk, akkor valódi segítség lehet abban, hogy a monoton, időrabló feladatokról levegyen némi terhet, és mi azokra a dolgokra koncentrálhassunk, amelyek valóban emberivé teszik a munkánkat – a kreativitásra, a megértésre, az érzelmekre és az empátiára.

A mesterséges intelligencia jövője nem a gépek kezében van, hanem a mi döntéseinkben: abban, hogyan alakítjuk, hogyan tanítjuk, és mire használjuk. A józanság, a kritika és a kíváncsiság együtt vezethet el oda, hogy ez a technológia ne csak lenyűgözzön bennünket, hanem valóban a javunkra is váljon.

## IRODALOMJEGYZÉK

- BROWN, T. B. – MANN, B. – RYDER, N. – SUBBIAH, M. – KAPLAN, J. – DHARIWAL, P. – NEELAKANTAN, A. – SHYAM, P. – SASTRY, G. – ASKELL, A. – AGARWAL, S. – HERBERT-VOSS, A. – KRUEGER, G. – HENIGHAN, T. J. – CHILD, R. – RAMESH, A. – ZIEGLER, D. M. – WU, J. – WINTER, C. – HESSE, C. – CHEN, M. – SIGLER, E. – LITWIN, M. – GRAY, S. – CHESS, B. – CLARK, J. – BERNER, C. – MCCANDLISH, S. – RADFORD, A. – SUTSKEVER, I. – AMODEI, D. 2020. Language Models are Few-Shot Learners. In: *ArXiv*, abs/2005.14165.
- BUBECK, S. – CHANDRASEKARAN, V. – ELDAN, R. – GEHRKE, J. A. – HORVITZ, E. – KAMAR, E. – LEE, P. – LEE, Y. T. – LI, Y. – LUNDBERG, S. M. – NORI, H. – PALANGI, H. – RIBEIRO, M. – ZHANG, Y. 2023. Sparks of Artificial General Intelligence: Early experiments with GPT-4. In: *ArXiv*, abs/2303.12712.
- CHOWDHERY, A. – NARANG, S. – DEVLIN, J. – BOSMA, M. – MISHRA, G. – ROBERTS, A. – BARHAM, P. – CHUNG, H. W. – SUTTON, C. – GEHRMANN, S. – SCHUH, P. – SHI, K. – TSVYASHCHENKO, S. – MAYNEZ, J. – RAO, A. – BARNES, P. – TAY, Y. – SHAZEER, N. M. – PRABHAKARAN, V. – REIF, E. – DU, N. – HUTCHINSON, B.

- C. – POPE, R. – BRADBURY, J. – AUSTIN, J. – ISARD, M. – GUR-ARI, G. – YIN, P. – DUKE, T. – LEVSKAYA, A. – GHEMAWAT, S. – DEV, S. – MICHALEWSKI, H. – GARCÍA, X. – MISRA, V. – ROBINSON, K. – FEDUS, L. – ZHOU, D. – IPPOLITO, D. – LUAN, D. – LIM, H. – ZOPH, B. – SPIRIDONOV, A. – SEPASSI, R. – DOHAN, D. – AGRAWAL, S. – OMERNICK, M. – DAI, A. M. – PILLAI, T. S. – PELLAT, M. – LEWKOWYCZ, A. – MOREIRA, E. – CHILD, R. – POLOZOV, O. – LEE, K. – ZHOU, Z. – WANG, X. – SAETA, B. – DÍAZ, M. – FIRAT, O. – CATASTA, M. – WEI, J. – MEIER-HELLSTERN, K. S. – ECK, D. – DEAN, J. – PETROV, S. – FIEDEL, N. 2022. PaLM: Scaling Language Modeling with Pathways. In: *ArXiv*, abs/2204.02311.
- HOFFMANN, J. – BORGEAUD, S. – MENSCH, A. – BUCHATSKAYA, E. – CAI, T. – RUTHERFORD, E. – CASAS, D. D. – HENDRICKS, L. A. – WELBL, J. – CLARK, A. – HENNIGAN, T. – NOLAND, E. – MILLICAN, K. – DRIESSCHE, G. V. – DAMOC, B. – GUY, A. – OSINDERO, S. – SIMONYAN, K. – ELSEN, E. – RAE, J. W. – VINYALS, O. – SIFRE, L. 2022. Training Compute-Optimal Large Language Models. In: *ArXiv*, abs/2203.15556.
- OpenAI 2023. GPT-4 Technical Report. In: *ArXiv*, abs/2303.08774.
- RADFORD, A. – WU, J. – CHILD, R. – LUAN, D. – AMODEI, D. – SUTSKEVER, I. 2019. Language Models are Unsupervised Multitask Learners. In: *OpenAI Blog*, 8., 9.
- VASWANI, A. – SHAZEER, N. – PARMAR, N. – USZKOREIT, J. – JONES, L. – GOMEZ, A. N. – KAISER, Ł. – POLOSUKHIN, I. 2017. *Attention is All you Need*. NIPS.
- WEI, J. – WANG, X. – SCHUURMANS, D. – BOSMA, M. – CHI, E. H. – XIA, F. – LE, Q. – ZHOU, D. 2022. Chain of Thought Prompting Elicits Reasoning in Large Language Models. In: *ArXiv*, abs/2201.11903.

**ABSTRACT** • In recent years, artificial intelligence has become a central topic in public discourse, especially since the emergence of generative models, foremost ChatGPT, at the end of 2022. After the initial hype and often apocalyptic predictions, by 2025 we have a much clearer view of what large language models can – and cannot – do. The three main messages of my study: (1) today’s AI is not a single “artificial mind” but an ecosystem of many small programs trained for narrow tasks; (2) generative models are not magic, but transformer-based language models built on statistical pattern recognition that estimate the probabilities of tokens; (3) these tools do not replace but complement human work and thinking. We briefly review the basic concepts of machine learning and deep learning, the operation of language models and

their development between 2018 and 2025, their practical application areas, as well as the foundations of prompt engineering. A separate section discusses their limitations, the phenomenon of hallucination, responsible use and the risks of misuse. The central claim of the talk is that understanding generative AI has become a key digital competence today, and that its future will be shaped not by the machines themselves, but by human decisions.

**KEY WORDS** • **artificial intelligence; generative AI; large language models (LLM); transformer; machine learning; deep learning; natural language processing (NLP); prompt engineering; AI ecosystem; hallucination; digital competence; education and AI**

**SAŽETAK** • Veštačka inteligencija je poslednjih godina postala jedna od centralnih tema javne rasprave, naročito od pojave generativnih modela, pre svega ChatGPT-a krajem 2022. godine. Nakon početnog hajpa i često apokaliptičnih predviđanja, do 2025. godine mnogo jasnije vidimo šta veliki jezički modeli mogu – a šta ne mogu. Izlaganje ima tri glavne poruke: (1) današnja VI nije jedinstveni „veštački um“, već ekosistem mnogih manjih programa obučenih za usko definisane zadatke; (2) generativni modeli nisu magija, već na transformatorima zasnovani jezički modeli koji se oslanjaju na statističko prepoznavanje obrazaca i procenu verovatnoće tokena; (3) ovi alati ne zamenjuju, već dopunjuju ljudski rad i razmišljanje. Ukratko se razmatraju osnovni pojmovi mašinskog učenja i dubokog učenja, način rada jezičkih modela i njihov razvoj u periodu 2018–2025, praktične oblasti primene, kao i osnove prompt inženjeringa. Posebno poglavlje posvećeno je ograničenjima, fenomenu halucinacija, odgovornoj upotrebi i rizicima zloupotrebe. Centralna teza izlaganja glasi da je razumevanje generativne VI danas ključna digitalna kompetencija, a da se njena budućnost neće odlučivati u rukama mašina, već kroz ljudske odluke.

**KLJUČNE REČI** • **veštačka inteligencija; generativni VI; veliki jezički modeli (LLM); transformator; mašinsko učenje; duboko učenje; obrada prirodnog jezika (NLP); prompt inženjering; VI ekosistem; halucinacija; digitalna kompetencija; obrazovanje i VI**