

JÁNOKI Imre Gergely¹

A neurális hálózatokról és a magyarországi mesterségesintelligencia-kutatásról

Absztrakt

2023 februárjában megrendezésre került a AI&AUT Expo szakmai kiállítás, amin a magyarországi mesterséges intelligencia (MI) és automatizálás kutatási eredményeit mutatták be. Ehhez kapcsolódva mutatja be a cikk a MI jelentését, alapjait és alkalmazhatóságát. A neurális hálózatok ismertetését a legújabb modellek és alkalmazások rövid bemutatása, valamint a szakmai kiállításon szerepelt demonstrációk csokra zárja.

Az Európai Unió célul tűzte ki a mesterségesintelligencia-kutatás és -fejlesztés területének felzárkóztatását a vezető országok, legfőképp az USA és Kína mellé. Ennek megfelelően Magyarország is szeretné megerősíteni pozícióját, sőt a régió zászlóshajójává kíván válni, így 2020-ban létrejött a Magyarország Mesterséges Intelligencia Stratégiája direktíva,² amely a 2030-ig elérendő mérföldköveket és irányt tartalmazza, valamint a kutatásokat egybefoglaló és finanszírozó Mesterséges Intelligencia Nemzeti Laboratórium (MILAB). 2023. február 14–15-én a Budapesten megrendezett nagyszabású AI&AUT Expo³ szakmai kiállításon ismerhették meg a résztvevők – többek között – a MILAB eddig elért eredményeit.

Azért, hogy betekintést nyerhessünk a területre, először tisztáznunk kell: mi is az a mesterséges intelligencia (MI). A köznyelvben már igen régóta alkalmazzuk a fogalmat a nagyfokú automatizáltságra és az emberi viselkedés gépi másának megjelenésére. Ilyenekre példák lehetnek különböző, látványosan adaptív, de hagyományos algoritmusokkal előre programozott gépek, illetve számítógépes játékokban és a régebbi szövegalapú asszisztensekben, chatbotokban megnyilvánuló intelligencia. Egyik leghíresebb ilyen mesterséges intelligencia az *IBM Deep Blue*⁴ rendszere, amely egy hatjátzmás sakkjátékban 1997-ben megverte a sakkozás akkori világbajnokát, Garri Kaszparovot. Ezt az akkor hatalmasnak számító teljesítményével érte el, amivel az adott állásból rengeteg lépéspármélyiségig meg tudott vizsgálni minden egyes lehetséges kimenetet.

¹ A szerző a Pázmány Péter Katolikus Egyetem Információs Technológiai és Bionikai Kar Roska Tamás Műszaki és Természettudományi Doktori Iskolájában doktorandusz, fejlesztő és tudományos segédmunkatárs a SZTAKI Számítógépes Optikai Érzékelés és Feldolgozás Kutatólaboratóriumában. Érdeklődése a számítógépes látás és feldolgozás orvosi területen történő alkalmazására irányul.

² Magyarország Mesterséges Intelligencia Stratégiája,

<https://digitalisjoletprogram.hu/files/2f/32/2f32f239878a4559b6541e46277d6e88.pdf> (Letöltés: 2023. 02. 26.)

³ AI&AUT expo, <https://www.ai-expo.hu/> (Letöltés: 2023. 02. 26.)

⁴ IBM Deep Blue, <https://www.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/> (Letöltés: 2023. 02. 26.)

A tudomány területén két fő csoportot neveznek mesterséges intelligenciának. A leginkább csak gépi tanuláshoz nevezett hagyományos MI meghatározott logikák és formalizmus alapján, matematikai és statisztikai analízisen alapul, gyakran a gráfelméletet segítségül hívva tár fel összefüggéseket. Idesorolhatjuk – a teljesség igénye nélkül – a széles körben használt *K-nearest neighbor*⁵ algoritmust, a *principal component analysis*⁶ vagy a *support vector machine*-t⁷. Ez utóbbiak az adatok nagy dimenziójú tulajdonság terében (*feature space*) képesek csoportosítani valamilyen meghatározott távolság vagy metrika alapján, majd az adatok között ilyenén összefüggéseket feltárva projektálni az adatpontokat ugyanabba vagy alacsonyabb dimenziójú térbe.

A másik csoportot a modern mesterséges intelligenciák alkotják, amelyek egy tanuló adathalmazon (mintákon) való tanulás alapján érik el a kívánt működést. Ezen belül a legnagyobb, és egyben a jelenkori MI-kutatással a legjobban azonosítható a mesterséges neurális hálózatok területe. A név a biológiai neuronokból ered, amik működési elvét alapul véve hozták létre az első ilyen neurális hálózatokat. Az ötletet és az elvet már az 1940-es években publikálták, majd dolgozták ki a XX. század második felében, azonban a kutathatóságnak és alkalmazhatóságnak a számítógépek – mai szemmel nézve – korlátozott kapacitása vetett gátat, ami még nem volt elegendő egy ilyen rendszer hatékony használatához. Az áttörést a 2010-es évek eleje jelentette, amikor egy nagyobb, mélyebb hálózatot már sikeresen lehetett tanítani és használni grafikus kártyán.

De mi is tulajdonképpen a mesterséges neurális hálózat, és miért szükséges a nagy számítási kapacitás és grafikus kártya? A neurális hálózatok legfontosabb építőeleme a mesterséges neuronok, amelyek egy gráf csomópontjaiként állnak egymással összeköttetésben, minden ilyen csomópont gyakorlatilag egy műveletcsoportot jelent. A hálózat első neuronjai a bemeneti adat valamilyen számalapú reprezentációját kapják meg. A bemenet számai pontonként egy-egy súlyparaméterrel (*weight*) szorzásra kerülnek, majd ezek összegéhez egy eltolási értéket adunk (*bias*), végül pedig egy nem-lineáris függvény (*activation function*) alapján kapjuk meg a kimeneti számot, ami ezután egyik bemenete lehet egy következő neuronnak. A különböző neuronokat rétegekbe rendezzük, amik általános tulajdonsága, hogy a rétegek neuronjai egymástól nem kapnak bemenetet. A különböző rétegek közötti összeköttetések is többfélék lehetnek: egy réteg az előzőtől mindig kap bemenetet, de kaphat korábbi rétegektől is, vagy rekurrens neurális hálózatok esetén akár későbbtől is. Így tehát a neurális hálózatok egy aciklikus gráfként leírhatók (rekurrens esetben kibonthatók aciklikus gráffá). Így tehát egy neurális hálózat a gyakorlatban egy rendkívül komplex, rétegenként párhuzamos, a rétegek között keresztbe is bejárható függvények láncolatából áll, amiben minden réteg tartalmaz nagyszámú állítható paramétert a következőkből: a bemenetek súlyai és a biasok. Ha a nem-lineáris, *neurononkénti activation function*ot ehhez hozzávesszük,

⁵ Evelyn Fix – Joseph L. Hodges: *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*. University of California report, 1951, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a800276.pdf> (Letöltés: 2023. 02. 26.)

⁶ K. Pearson: *On Lines and Planes of Closest Fit to Systems of Points in Space*. *Philosophical Magazine* 2 (11): 559–572. doi: [10.1080/14786440109462720](https://doi.org/10.1080/14786440109462720)

⁷ Corinna Cortes – Vladimir Vapnik: *Support-vector networks*. *Machine Learning* 20 (3): 273–297, doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)

könnyen láthatjuk, hogy ez egy univerzális approximátor, vagyis tetszőleges függvény közelíthető vele.

A gyakorlatban tehát hogyan néz ki egy ilyen hálózat, mit várunk el tőle? A hálózatot valamilyen programozási nyelven írják, ahol előre definiálják az „építőelemeit” (például neuronok, rétegek), majd meghatározzák a neuronok és rétegek számát, ezek összeköttetéseit, kiegészítő algoritmusokat, végül ebből megalkotják az említett gráfot, architektúrát. Elmondható, hogy az első rétegek az alacsony szintű tulajdonságokat nyerik ki, majd ezeket dolgozzák fel a későbbi rétegek. Erre egy példa a képfeldolgozás területén, amikor egy arcot szeretnénk detektálni: az első rétegek kinyerik a képeken az éleket, az íveket, a színeket, majd a későbbi rétegekben a megfelelő élek és ívek megfelelő színeloszlással szemeket, szájat és orrot alkothatnak, végül ezek összességéből a hálózat megtalálja az arcot a képen. Amit pedig elvárunk a neurális hálózattól az, hogy a beadott adat, szöveg, kép számalapú reprezentációja, ha keresztülfut a hálózaton, egy megfelelő kimenetet adjon, például egy *predikciót* a tőzsdei adatokra, a szöveg valamilyen nyelvű fordítását, vagy ismerje fel, mi van a képen és hol. Ezekhez természetesen a megfelelő paraméterek szükségesek; ezek beállítását hívjuk a neurális hálózat tanításának. A tanítás a következőképpen zajlik: a paramétereknek valamilyen stratégia szerint, általában véletlenszerű módon, kezdeti értéket adunk. Ezután megnézzük, hogy a hálózat milyen kimenetet ad adott bemenetre, majd a kimenetet kiértékeljük egy előre meghatározott, úgynevezett veszteségfüggvénnyel (*loss function*), ami megadja a hibát, gyakorlatilag egy metrika, távolság a hálózat kimenete és az elvárt kimenet között. Ennek egy példája: bizonyos adatok alapján kialakult egy grafikon, legyen ez a tőzsdepiac, és az azt befolyásoló tényezők. A tanítás során beadjuk a múltbeli adatokat, majd az elvárt kimenetünk a tőzsdei árfolyam alakulása lesz. A kiadott érték és a valós érték közötti négyzetes különbség lesz a *loss function* által meghatározott hiba. Ezután ezt a hibát megpróbáljuk visszaterjeszteni: a függvényláncolatán visszafelé haladva számolunk *grádiens*t (differenciált), ez alapján próbáljuk meghatározni, hogy az aktuális bemenetre adott kimenet és az elvárt kimenet közötti különbséghez, hibához melyik paraméter milyen mértékben járult hozzá, majd ennek megfelelően módosítjuk azokat. Ez alapján, ha a megfelelő – igen nagy – mennyiségű és minőségű pontos adatot mutatunk be, sőt azokat többször is, akkor megfelelő stratégiával, megfelelő architektúra mellett a hálózat egy idő után képessé válhat általánosítani, és addig még sosem bemutatott, független adatra (teszt adathalmaz) is az elvárt kimenetet adja ki. Ez hasonlatos ahhoz, mint amikor az ember felismeri a kutyákat még akkor is, ha az adott fajtaival még sosem találkozott. A modern, sok réteggel rendelkező neurális hálózatokat gyakran nevezik mély neurális hálózatnak (*deep neural network*), a területet pedig mély tanulásnak (*deep learning*).

Egy modern, de relatíve egyszerű, feladatspecifikus neurális hálózat több millió változtatható paraméterből áll, míg a cikk írásakor talán legismertebb architektúra, a GPT-3⁸ nagy verziója 175 milliárd, míg a leghíresebbnek mondható alkalmazás, a ChatGPT⁹ 20 milliárd paraméterrel dolgozik. Ez az elképesztő mennyiség hatalmas számítási kapacitást igényel, és

⁸ Tom B Brown et al: *Language Models are Few-Shot Learners*. Advances in Neural Information Processing Systems. Curran Associates, Inc. 33: 1877–1901. doi: 10.48550/arXiv.2005.14165

⁹ ChatGPT <https://openai.com/blog/chatgpt/> (Letöltés: 2023. 02. 26.)

a futtatás és maga a tanítási folyamat időigényes. Azonban a rengeteg, kisebb számítás feladat masszívan párhuzamosítható; ez azt jelenti, hogy a különböző magok egyesével egy, de összességében számtalan részfeladatot végeznek el, amivel sokszorosára gyorsítható a folyamat grafikus kártyák használatával. A grafikus kártyák továbbá maguk is egymás mellé köthetők, így végül még ezeknek a hatalmas neurális hálózat modelleknek a tanítási ideje is lecsökkenthető napokra, természetesen az akár több százezer dolláros géppark fenntartási vagy bérleti költsége mellett.

Láthatjuk, hogy a legújabb, főként generatív modellekhez szükséges a drága, nagykapacitású géppark, valamint nagyon nagy adatmennyiségre van szükség. Ez utóbbi általánosságban is elmondható, a több adatból jobb MI készül. Innen már könnyen látható, hogy versenyelőnyből indul például a Tesla az önvezető autók terén, hiszen saját flottájából gigászi saját adatbázist tud kinyerni. A neurális hálózatok másik híres piacvezetője az OpenAI, ami a saját fejlesztésű – már említett – GPT-3 architektúra kitalálója, a DALL-E¹⁰ nevezetű szövegből kép (*text-to-image*) mesterséges intelligencia és a – már szintén említett – ChatGPT chatbot megalkotója. A vállalat több alkalommal is igen nagy befektetést kapott a Microsofttól. A ChatGPT egy intelligens chatbot. A tanítóhalmaza az internetről kinyert hatalmas információmennyiség és chat-beszélgetések, így képes nemcsak emberszerű írásos társalgásra, hanem a tanult tudást visszaadva esszéyszerű választ ad akár szakmai kérdésekre is – sikeresen teljesített amerikai jogi és gazdasági vizsgákat is.¹¹ A ChatGPT sikerességét mutatja, hogy a Microsoft több szolgáltatásába is már építi be az MI-t, valamint a Google a keresőmotorjával szemben olyan ellenfelet lát, amire válaszként saját verziójukat kezdték el kifejleszteni.

Egy MI kimenetét a siker, a széles körű érdeklődés és nyilvánosság ellenére azonban jelenleg érdemes fenntartásokkal kezelni. Egyrészt mondhatjuk, hogy a neurális hálózatok egyfajta fekete dobozok: maga a hálózat és a paramétereinek tanítása nem feleltethető meg egy hagyományos, kézzel írt algoritmusnak, gyakorlatilag lehetetlen determinálni és ellenőrizni, hogy minden bemenetre jó kimenetet ad-e. Így tehát ha a ChatGPT-vel íratnánk meg a szakdolgozatunkat, első ránézésre igen jó írást tartanánk a kezünkben, de könnyen előfordulhat, hogy az esszé tárgyi tévedést tartalmaz. Ugyan igen nagy valószínűséggel ad jó választ, de inkonzisztens módon nagyon súlyos tévedéseket tartalmazó, hibás kimenetek is előfordulnak, amik a „feketedoboz” jellegből adódóan nem kiszámíthatók és határozhatók meg előre. A jelenlegi verzióban egy példa erre a téves érvelés: amikor egy hamis állítás miértjére kérdezzük rá, a *chatbot* megpróbál érvekkel válaszolni, annak ellenére, hogy az állítás eleve hamis, ezt nem képes felismerni. Egy másik népszerű kérdéskör a MI részrehajlása. Lehet-e rasszista egy *chatbot*? A válasz igen és nem. Nem, hiszen az MI nem egy gondolkodó lény, azonban amennyiben a *chatbot* tanító halmaza rasszista beszélgetéseket, megnyilvánulásokat tartalmazott jelentős részben, úgy ezt a mintázatot fogják tükrözni a MI válaszai is, tehát rasszista lehet. Természetesen a tanító adatok gyűjtése, megfelelő címkézése, válogatása, tisztítása újfent mind igen jelentős idő és költség.

¹⁰ DALL-E 2 <https://openai.com/dall-e-2/> (Letöltés: 2023. 02. 26.)

¹¹ ChatGPT passes exams from law and business schools <https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html> (Letöltés: 2023. 02. 26.)

Tény tehát, hogy a legmodernebb és legnevesebb – tipikusan generatív – neurális hálózatok kutatása és fejlesztése igen költséges vállalkozás. Azonban a feladat specifikusabb, nem generatív MI-hoz használt hálózatok szerencsére jóval kisebbek, és a kisebb hálózatoknak kevesebb adat is elég a megfelelő tanításhoz és működéshez. Magyarországon a MILAB egyfajta gyűjtő, esernyő szervezatként osztja el a forrásokat, ezzel támogatva a MI-kutatást és -fejlesztést. A jelenlegi kutatások jelentős része a következő területekhez kapcsolódik: önvezető drónok, önvezető autó, orvosi alkalmazás, mezőgazdaság; tehát leginkább az automatizáláson van a hangsúly.

A cikk további részében néhány, a 2023-as AI&AUT Expón megjelent demonstráció kerül bemutatásra, mint a magyarországi MI-kutatás és-fejlesztés példái.

Az Eötvös Loránd Tudományegyetem (ELTE) kutatói olyan mesterséges intelligenciát hoztak létre, amely automatizált módon, egy egyszerű színes kamera segítségével jó pontossággal képes meghatározni a tejlő szarvasmarhák testállapotindexét (*body condition index*). Szintén a mezőgazdaságban a Xnetik¹² XAG okos drónokat kínál, amelyek mesterséges intelligencia segítségével képesek feltérképezni a biomassa-változást a levegőből, illetve felismerve a növényeket, földi vagy légi, célzott permetezést vagy granulátumszórást hajtanak végre a kijelölt területen. A Clementine Hanga nevű virtuális asszisztense hangalapon dolgozik, kiváltva ezzel az előre programozott, választásalapú interaktív hangmenü (IVR) callcenter-rendszereket. Külön neurális hálózatok dolgozzák fel és alakítják szöveggé az emberi hangot, kiértékelik a szöveget és alakítanak rá választ, végül azt szövegből hanggá alakítják vissza.

A Bosch és a Continental is saját önvezető technológia fejlesztésén dolgozik. Egy ilyen rendszer részét képezhetik a Széchenyi Egyetemen fejlesztett, kameraalapú útfelszín, sáv, jármű- és gyalogosfelismerő neurális hálózatok.

A Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI), a MILAB és a kiállításon szintén részt vevő Autonóm Rendszerek Nemzeti Laboratórium (ARNL) koordinátora is, így több demonstrációval is jelen voltak a kiállításon. A SZTAKI is bemutatta saját kutatását az önvezető autók témakörében. Magyarországon ezek tesztelésére épült ki a ZalaZONE¹³ tesztpálya és kutatás-fejlesztés segítő központ, ahol többek között a Széchenyi Egyetem és a SZTAKI technológiáit is tesztelik. Fontos szerepet kaptak a drónok is: a SZTAKI fejlesztésében mesterséges intelligencia koordinálja több drón vagy egy drónraj kooperatív mozgását, segít feltérképezni és tájékozódni ezeknek az önvezető repülő járműveknek.

Jelen cikk szerzője a SZTAKI Számítógépes Optikai Érzékelés és Feldolgozás Kutatólaboratórium¹⁴ tagja. A labor a kiállításon három demonstrációval jelent meg. A hagyományos fénymikroszkópok mélységélessége a pár mikronos tartományban mozog, csak a tárgylemezzel párhuzamos síkot látja át. A labor digitális holografikus mikroszkóp rendszere viszont képes felvételt készíteni akár egy 1 milliméter vastag térfogatról is, amiben

¹² Xnetik <https://www.xnetik.hu/> (Letöltés: 2023. 02. 26.)

¹³ ZalaZONE <https://zalazone.hu/> (Letöltés: 2023. 02. 26.)

¹⁴ SZTAKI Számítógépes Optikai Érzékelés és Feldolgozás Kutatólaboratórium <https://www.sztaki.hu/tudomany/reszlegek/analogic> (Letöltés: 2023. 02. 26.)

folyadék vagy gáz található. A prezentált rendszerben mikrofluidikai csatornán folyik át a tesztelni kívánt folyadékminta, úgymint ivóvíz, szennyvíz, természetes vizek mintája, vagy például sör. Az átfolyó folyadékról egyidejűleg készített volumetrikus képeken utólag, digitálisan lehet „végigfókuszálni” a Z-síkon – hasonlóan, mint ahogy a fénymikroszkóp fókuszát tekerjük. A leképezett mintában mesterséges intelligencia találja meg és osztályozza a keresett részecskéket, például algákat, algatelepeket, baktériumokat, egysejtűeket, élesztőt, törmeléket, illetve osztályozza, rendszerezi ezeket. Az eredményből a szoftver végül statisztikákat készít.

A labor szorosan együttműködik a Semmelweis Egyetemen és az inkubátorokat és orvosi eszközöket gyártó Medicor Zrt.-vel. Ennek keretében inkubátoros, koraszülött babák kameraalapú monitorozásának kutatása folyik. A koraszülött csecsemők bőre még nagyon érzékeny, sérülékeny, amit az érzékelők könnyen megsérthetnek. A kábelek egyébként zavarják a babákat is, csakúgy, mint az orvosokat, a nővéreket és a szülőket. A pulzuszórákat, a légzésmérést EKG-elektrodák segítségével lehet végezni, azonban ez utóbbiak folyamatos leszedése és felhelyezése miatt az eszközt csak indokolt esetben használják az orvosok. Ezeket helyettesíti egy infraszűrő nélküli színes kamera. Már a legtöbb kereskedelmi kamera, például egy mobiltelefon kamerája is képes felvenni a bőr emberi szem számára láthatatlan, apró színváltozásait, amiket a vér pulzálása okoz. Ebből megfelelő megvilágítás és zavaró mozgás nélkül már hagyományos algoritmusokkal is kinyerhető a pulzusráta, azonban egy mozgó koraszülött változó fényviszonyok között, a felnőtteknél gyengébb pulzussal már olyan kihívás, amihez mesterséges intelligenciát kell alkalmazni. A labor által készített rendszerben a kamera képe alapján több neurális hálózat értékeli ki, hogy a mérési körülmények adottak-e, és hogy kell-e mérni: a baba az inkubátorban van, épp nem történik sem orvosi beavatkozás, sem vizsgálat vagy etetés, illetve a fényviszonyok, képélesség stb. megfelelő-e. Ezután a képsorozatból neurális hálózat határozza meg a pulzusjelet, majd ebből a pulzus rátát. A légzésráta mérése – szintén az említett kiértékeléstől függően – a has mozgása alapján történik. Ehhez külön hálózat szegmentálja, maszkolja ki a baba hasát és hátát a végtagok nélkül, majd hagyományos algoritmus figyel ezen terület mozgását. A projekt folytatásaként, a labor jelenlegi kutatása a babák alvás-ébrenlét ciklusának figyelésére irányul. A megfelelő mennyiségű és minőségű alvás nagyon fontos a koraszülöttek agyi fejlődéséhez ebben a kritikus életszakaszban. Az ilyen kisbabáknál az alvás és ébrenlét megállapítása még nem teljesen egyszerű, és ehhez az inkubátorokat folyamatosan figyelnie kellene a kórházi személyzetnek. Ehelyett újfent a kamerakép alapján határozná meg, és jelezne vissza a rendszer: a baba mikor aludt eleget és mikor van ébren, mikorra kell a vizsgálatot és az etetést időzíteni.

A labor egy harmadik demonstrációval is jelen volt a kiállításon. A projekt egy személy, tipikusan autósófor vagy ember-gép kooperációban dolgozó gyártósori munkás figyelmének és fáradtságának meghatározását kutatja. Az egy kamerából és számítógépből álló rendszerben egy neurális hálózat érzékeli az emberi arcot, majd egy másik kijelöli az arc nevezetes pontjait: az arcéleket, orrot, szájat, szemeket, szemöldököt. Ezután algoritmikus úton, a pontok alapján már meghatározható lesz a fej irányultsága, a szemhéj zártsága és a zárás karakterisztikája (ti. „laposakat pislog”), valamint ha a megfigyelt személy ásít. Az előző projekthez hasonlóan ezúttal is meghatározható a pulzus. A demonstrációban még nem

implementált pupilladetekcióval a nézés iránya is megállapítható. Ezen adatok időbeli összességéből a jövőben meghatározható lesz a fáradtság mértéke, valamint az, hogy az illető például körülnéz-e egy kereszteződésben, figyel-e az utcatáblákat, a robot mellett odafigyel-e arra, amit csinál.

Összességében tehát elmondható, hogy a magyarországi MI-kutatás lendületbe jött, a fókuszban a feladatspecifikusabb neurális hálózatok állnak. A MI jövőjének fényét egyelőre a nehéz tesztelhetőség és determinálhatóság tompítja, így ugyan a jelenlegi technológia is már igen sok feladat automatizálására alkalmas, az emberi kritikai ellenőrzést és felügyeletet továbbra is mindenképpen igénylik ezek a rendszerek. Az iparban így a kooperatív, „ember-robot” megoldások jelentik a fő alkalmazási módszert.