

A kételkedés képessége megbízhatóbbá teheti a mesterséges intelligenciát

Nemcsak jó döntést kell hozni, hanem azt is fel kell mérni, mennyire biztos az adott döntés. A HUN-REN Wigner Fizikai Kutatóközpont Orbán Gergő vezette kutatócsoportja argentin partnerekkel együtt olyan módszert dolgozott ki, amely javítja a mélytanuló rendszerek bizonytalanságkezelését. Az eredmény a képfeldolgozó mesterséges intelligencia megbízhatóságát növelheti, ami különösen fontos lehet például az orvosi képalkotásban, emellett az emberi látás működésének pontosabb megértéséhez is közelebb vihet. Az eredményekről az [IEEE Transactions on Pattern Analysis and Machine Intelligence](#) folyóiratban számoltak be.

A kutatás a **NeuroAI** területéhez kapcsolódik, amely az idegrendszer és a mesterséges intelligencia működése közötti párhuzamokat vizsgálja. A cél kettős: az agy működésének jobb megértése segíthet hatékonyabb gépi intelligencia kialakításában, miközben a mesterséges rendszerek vizsgálata új ötleteket adhat az idegrendszer kutatásához.

A kutatók ezúttal arra keresték a választ, hogyan tudná egy mesterséges intelligencia pontosabban jelezni, mennyire biztos abban, amit egy képről kikövetkeztet. Ez azért fontos, mert nem ugyanaz a helyzet, ha egy rendszer több hasonló lehetőség közül éppen csak az egyiket tartja valószínűbbnek, mint amikor a többi lehetőséget nagy biztonsággal ki tudja zárni.

Amikor a mesterséges intelligencia túl magabiztos

A vizsgálat középpontjában a **variációs autoenkóderek**, röviden VAE-k álltak. Ezek olyan mély generatív modellek, amelyek megtanulják egy kép lényeges jellemzőit egy tömör, belső reprezentációban megragadni, majd ebből akár újra felépíteni a képet. A szakemberek számára vonzerejük az, hogy arra is alkalmasak, hogy ne csak egy következtetést adjanak, hanem annak bizonytalanságát is megbecsüljék.

A kutatók azonban azt találták, hogy a hagyományos VAE-k bizonytalanságbecslése több gyakori képfeldolgozási helyzetben hibásan működhet. Ilyen lehet például, ha megváltozik egy kép kontrasztja, ha a kép zajos vagy sérült, illetve ha a rendszer olyan képpel találkozik, amely jelentősen eltér a tanítás során látott példáktól.

A magyar-argentin kutatócsoport ezért a klasszikus gépi látásból ismert megoldással egészítette ki a modellt. Az új változatot **Explaining-Away Variational Autoencodernek, EA-VAE-nek** nevezték el. A beépített új tényező segít a modellnek elkülöníteni, hogy a képen látott változás valóban az ábrázolt objektumból ered-e, vagy például egyszerűen a teljes kép fényereje, kontrasztja változott meg. A módszert többféle képi adaton is tesztelték, és az új modell következetesebb bizonytalanságbecslést adott.

Az agyban is megtalálható megoldás jelent meg

A kutatás külön érdekessége, hogy az EA-VAE működésében megjelent a **divízió normalizáció** nevű számítási mechanizmus. Ennek lényege leegyszerűsítve az, hogy egy idegsejt aktivitását nem önmagában, hanem a környező sejtek aktivitásához viszonyítva szabályozza a rendszer. Ez az agyban, különösen a látórendszerben széles körben megfigyelhető működési elv.

Vagyis a kutatók nem egyszerűen egy mesterséges intelligencia teljesítményét javították: olyan

megoldáshoz jutottak, amelynek működési elve a biológiai idegrendszerben is jelen van.

Mi lehet a gyakorlati haszna?

A bizonytalanság helyes felmérése különösen fontos azokban az alkalmazásokban, ahol egy tévesen magabiztos döntés komoly következményekkel járhat. Egy orvosi kép elemzésekor például nemcsak az számít, hogy az algoritmus milyen elváltozást azonosít, hanem az is, hogy mennyire biztos ebben. Egy jól kalibrált rendszer jelezheti, ha az adott kép szokatlan vagy nehezen értelmezhető, és ezért emberi szakértő ellenőrzésére van szükség.

Hasonló előnyt jelenthet a módszer minden olyan képfeldolgozó rendszerben, amelynek változó fényviszonyokkal, zajjal, sérült adatokkal vagy korábban nem látott képekkel kell megbirkóznia.

A kutatás így egyszerre mutat utat a **megbízhatóbb mesterséges intelligencia** és a **biológiai látás pontosabb modellezése** felé.

Sajtókapcsolat:

- HUN-REN Kommunikáció
- media@hun-ren.hu

Eredeti tartalom: HUN-REN Magyar Kutatási Hálózat

Továbbította: Helló Sajtó! Üzleti Sajtószolgálat

Ez a sajtóközlemény a következő linken érhető el:

<https://hellosajto.hu/?p=33702>