

„A gépek kezdik megérteni a teret” — Federico Tombari az AI Symposiumon a háromdimenziós AI-ról, a megmagyarázhatóságról és az immerzív világokról

Az AI Symposium 2026 alkalmával Federico Tombarival, a Google zürichi kutatási igazgatójával, a 3D számítógépes látás, a multimodális modellek és az autonóm érzékelés egyik vezető szakértőjével beszélgettünk arról, hogyan alakítják át a teljes folyamatot lefedő AI-rendszerek a gépi érzékelést, miért válik egyre fontosabbá a megmagyarázhatóság, és mire van szükség ahhoz, hogy a térbeli AI és az immerzív 3D-jelenetgenerálás a látványos demókból valódi, széles körű alkalmazássá érjen.

- Munkája régóta arra irányul, hogy a gépek a tárgyak pusztá felismerésétől eljussanak a háromdimenziós környezetek megértéséig. Ha visszatekint az elmúlt évtizedre, mi volt a gépi érzékelés legnagyobb fordulópontja?

- 2021 körül következett be az igazi áttörés a mesterséges intelligenciában, elsősorban a gépi tanulásban. A nagy nyelvi modellek előretörésének tovagyűrűző hatásai hamar érezhetőek lettek a háromdimenziós számítógépes látásban és érzékelésben is.

A legjelentősebb változás az általános célú, a teljes munkafolyamatot lefedő modellek előretörése volt. Korábban sok feladatot egymáshoz illesztett algoritmusokból álló feldolgozási láncok kezeltek. Ma ezeket egyre inkább egyetlen modell váltja fel, amely ugyanazt a munkát elejétől a végéig önmagában elvégzi.

Ideális esetben ugyanaz a modell számos különféle feladatra is újrahasznosítható, aminek megvannak az előnyei és a hátrányai is. Előnye, hogy többféle feladat elvégzéséből kinyert tudást egyetlen modellbe integrálhatjuk. Minél több adattal tanítjuk a modellt, annál inkább képes lesz eltérő területeken is tovább tanulni és fejlődni. A hátrány viszont az, hogy ezek a modellek óriási fekete dobozokká válnak. Van ugyan egy egységes, minden részfeladatot lefedő modellünk — csak éppen nem leszünk képesek átlátni, hogy mit csinál és nem fogjuk érteni, mi történik a belsejünkben.

Ez számos valós alkalmazásban válik kritikussá: ipari folyamatokban, gyártásban, autonóm rendszerekben, robotikában, önvezetésben. Ha valami félremegy, tudni kell, hol és miért történt, és mit kell tennünk, hogy a hibás komponenst kijavítsuk. Ezért is kap most különös hangsúlyt a mesterséges intelligencia megmagyarázhatósága: fel kell nyitnunk a fekete dobozt, és meg kell értenünk, hogyan működnek a modell egyes elemei, és pontosan hol vallanak kudarcot.

- Mennyire sürgető kérdés a megmagyarázhatóság a vizuális AI-ben, például a kép- vagy videogenerálásban? A szövegben elkövetett hibákat nem mindig könnyű észrevenni, de egy hatujjú kéz azonnal szembetűnik.

- Ez valóban fontos kérdés, bár részben eltérő kihívásokra mutat rá, olyanokra, amelyek inkább az AI tanúsíthatóságával és biztonságával kapcsolatosak.

A megmagyarázhatóság ott válik igazán kritikussá, ahol egy modell közvetlenül valamilyen valós idejű rendszer döntési láncában működik. Ha például a gépi tanulási modell egy önvezető autó

viselkedéséről hoz fontos döntéseket, és valami rosszul sül el, abból veszélyes helyzet alakulhat ki. Ilyenkor pontosan érteni kell, mi történt.

Amit ön említ, az inkább a digitális tartalomkészítésben használt generatív AI kérdésköre. Ott az egyik fő kihívás annak megállapítása, hogy az adott tartalmat a mesterséges intelligencia hozta-e létre vagy sem. Ez nem csupán technikai feladat. Természetesen szükség van az AI-generált tartalom megfelelő azonosítását és osztályozását lehetővé tévő módszerekre, de kellene a szabályozási keretek is. Olyan irányelvekre van szükség, amelyeket iparági és kormányzati szereplők közösen fogadnak el, és amelyek világos szabályokat határoznak meg az AI-generált tartalmakra.

Ide kapcsolódik a nyomon követhetőség kérdése is: meg kell tudni állapítani, hogy egy kép vagy videó mikor és hogyan jött létre. Erre egyik lehetséges megoldás a vízjelezés. De számos olyan módszert vizsgálnak, amelyekkel a képeket és videókat előállító generatív modellek hatékonyan megjelölik az általuk készített tartalmat, mindenki számára nyilvánvalóvá téve, hogy AI-generált tartalomról van szó. Ez segíthet csökkenteni a deepfake-ekkel és hasonló visszaélésekkel kapcsolatos kockázatok egy részét.

- Eljutottunk már oda, hogy az átlagos felhasználó nem tudja megbízhatóan megkülönböztetni a teljesen AI által generált videót a valóditól?

- Nagyon közel járunk ehhez a ponthoz — és pontosan ezért kell ezeket a védelmi mechanizmusokat mielőbb bevezetni.

Van egy másik probléma is: a generált tartalom bizonyos esetekben nagyon szorosan korrelálhat olyan anyagokkal, amelyeket a modell korábban már látott. Éppen ezért fontos, hogy a kereskedelmi célra használt modellek képesek legyenek azonosítani a szerzői jogilag védett anyagokat, és egyértelmű szabályok szerint kezelni azokat. Az ezen a területen működő vállalatok, például a Google, óriási hangsúlyt fektetnek erre. A tanítóadatok kezelésére már születtek egyértelműbb szabályok, bár természetesen még ezen a területen is lenne mit tenni.

- Az előadásában lenyűgöző példákat mutatott arra, hogyan alakíthatók videók és fényképek immerzív 3D környezetekké. Melyik területen várja ennek első igazán átalakító hatását: navigáció, távoli együttműködés, kiskereskedelem, tervezés vagy valami egészen más?

- Az ilyenfajta technológia valóban számos alkalmazás előtt nyitja meg az utat. Az alapgondolat az, hogy a generatív AI segítségével új lehetőségeket tárjunk fel a digitális tartalomkészítésben. Amit előállítunk, az ne csupán vizuálisan meggyőző, hanem geometriailag is valóságos legyen – olyasmi, ami valóban megragadja a harmadik dimenziót.

Ennek révén számos új felhasználási mód jöhet létre. Ahogy az előadásomban is említettem, az egyik ilyen az immerzív környezetek létrehozása, amelyekben az emberek ténylegesen navigálni, virtuálisan mozogni tudnak — akár valós helyszínek rekonstrukcióiról, akár mesterségesen generált terekről van szó. Ez nemcsak a játékok és a kevert valóság szempontjából releváns, hanem az autonóm rendszerek számára is.

Most kezdenek megjelenni az úgynevezett világmodellek: olyan rendszerek, amelyek 3D környezetek interaktív digitális reprezentációit hozzák létre. Alkalmazásaik egy része a szórakoztatás felé mutat, más esetekben viszont ezek a modellek értékes szintetikus adatokat állítanak elő robotrendszerek, autonóm ágensek vagy önvezető modellek tanításához, hogy azok hatékonyabban és biztonságosabban működjenek.

Mindezekben a felhasználási esetekben a harmadik dimenzió nem pusztán vizuális többlet: a

geometriai hűség alapkövetelmény. Ha a mögöttes 3D-szerkezet torzul, azt kockáztatjuk, hogy zajos vagy félrevezető adatokkal tápláljuk a rendszereket. Ha a virtuális geometria nem követi szorosan a valós világot, a robot rossz adatokon fog megtanulni navigálni a környezetében, vagy az önvezető jármű nem lesz képes biztonságosan közlekedni – vagyis a rendszer nem látja el a feladatát, amire tervezték.

- Dolgozott az akadémiai világban és az iparban is. A két eltérő világban szerzett tapasztalatai miként formálták a gondolkodását arról, hogy mitől válik valóban értékes az AI-kutatás?

- Az elmúlt néhány évben valódi átrendeződés ment végbe az akadémiai világ és az ipar viszonyában, különösen az AI területén.

Hagyományosan az akadémiai szféra volt az innováció fő hajtóereje. Ott született a felforgató ötletek többsége, az ipar pedig inkább a technológiatranszferre összpontosított, vagyis arra, hogy a legígéretesebb ötleteket terméké vagy alkalmazásokká alakítsa.

Az utóbbi években ez az egyensúly eltolódott. Ennek egyik oka éppen az a trend, amelyet korábban említettem: egyre inkább nagy, egységes, minden feladatot elvégző modelleket tanítunk, ami óriási mennyiségű adatot és számítási kapacitást igényel. Az ezekhez való hozzáférés kritikus fontosságúvá vált.

Ma az innováció jelentős része olyan helyeken folyik, ahol ezek a kapacitások rendelkezésre állnak. Nem meglepő módon ez inkább az ipari szereplőkre igaz, és emiatt a kutatás és az innováció súlypontjai bizonyos mértékig átrendeződtek.

Azért is hívom én ezt a folyamatot átrendeződésnek és nem pedig leváltásnak, mert szerintem egyik oldal jelentősége sem csökkent. Mindkettő továbbra is alapvető szerepet játszik a mesterséges intelligencia fejlesztésében. Az akadémiai világ továbbra is nélkülözhetetlen a nagy kockázatú, felforgató ötletek követésében; az ipar viszont jobb helyzetben van a csúcstechnológiát képviselő modellek fejlesztésére és skálázására.

Ami most különösen figyelemre méltó, az a két terület között szövődő partnerségek megerősödése. Az ilyen konzorciumok azért is egyre fontosabbak, mert lehetővé teszik, hogy a különböző intézmények egyesítsék erőforrásaikat, és elérjék az adatok és a számítási kapacitás szempontjából szükséges kritikus tömeget. Az AI közelmúltbeli fejlődésének egyik következménye, hogy az ilyen jellegű együttműködések egyre fontosabbá váltak.

- Feltűnő volt látni, milyen sokféle platform képes ma már kapcsolatba lépni a 3D világgal, a telefonoktól az okosművegeken át a headsetekig. Mit tesz lehetővé egy headset vagy okosműveg, amire egy laptop képernyője nem képes?

- Az immerzív alkalmazások valódi értéke az alapvetően másfajta felhasználói élményben rejlik.

Ezekre akkor van igazán szükség, amikor a felhasználó minél inkább „jelen akar lenni”, például amikor egy tőle fizikailag távol lévő világhoz szeretne kapcsolódni vagy olyan emberekkel akar érintkezni, akik nincsenek vele egy térben. Ilyen esetekben az immerzív élmények egyértelműen hozzáadott értéket adnak. Ugyanez igaz akkor is, amikor a világot szeretnénk felfedezni vagy megérteni. Sok fogalom könnyebben megragadható, intuitívabbá és meggyőzőbbé válik, ha immerzív interfészekon keresztül, nem pedig síkképernyőkön jelenik meg.

Engem különösen az érdekel, hogyan tehetik lehetővé ezek az eszközök a gazdagabb tanulási és

mentorálási élményeket. Az AI segíthet olyan oktatási eszközök létrehozásában, amelyek megkönnyítik az összetett gondolatok megértését, és sokkal szélesebb körben elérhetővé is teheti ezeket az eszközöket — olyanok számára is, akik máskülönben nem férnének hozzájuk. Úgy gondolom, hogy ez lehet az AI egyik legfontosabb és legpozitívabb lehetséges hatása.

- Mi lesz a legfontosabb az XR és a térbeli számítástechnika szélesebb körű elterjedésének előmozdításában — jobb algoritmusok, jobb hardver vagy gazdagabb fejlesztői ökoszisztéma?

- Valójában mindez együtt, mert szorosan összefüggenek egymással.

Hardveroldalon nemcsak a mechanikai komponensekről és a szenzorokról van szó, hanem a kijelzőkről is. Az immerzív élményekhez használt kijelzőknek egyszerre kell könnyűnek, nagy felbontásúnak, energiatakarékosnak és könnyedén hordhatónak lenniük – ez pedig rendkívül komoly mérnöki kihívás.

Hasonlóan fontos összetevő a chipkészlet, vagyis azok a mobil vagy beágyazott számítási egységek, amelyek képesek ezeket a modelleket és algoritmusokat közvetlenül az eszközön futtatni, miközben megfelelnek az akkumulátoridő, a késleltetés és a pontosság terén megkövetelt szigorú követelményeknek.

És természetesen ott van az igény a korszerű gépi tanulási modellekre és algoritmusokra, mint ahogy a fejlesztői ökoszisztéma fontosságát sem szabad lebecsülni. Az okostelefon-platformok sikerének történetéből jól látszik, milyen döntő szerepet játszhat egy aktív és lelkes fejlesztői közösség. A közösség létrehozásához biztosítani kell a megfelelő eszközöket, vonzóvá kell tenni a platformot, és teret kell adni a kreativitásnak, az ötletek valóra váltásának. Ez is lényeges kihívása ennek a területnek.

- Ha öt évre előre tekint, mi győzné meg arról, hogy a térbeli AI és az immerzív 3D jelenetgenerálás valóban túljutott a látványos demókon, és széles körű, hétköznapi használatba került?

- Ilyen gyors fejlődés mellett nehéz megjósolni a jövőt, még akár csak öt év távlatában is.

Az egyik figyelemre méltó terület az autonóm ágenseké. Jelenleg egyik fő korlátjuk az, ahogyan a fizikai világgal kapcsolatba lépnek. Közeledünk ahhoz a ponthoz, ahol az ágensek már viszonylag jól megértik a világot, és egyre pontosabban képesek mozogni benne, javul az akadályok elkerülésének, a dinamikus irányítás és az általános navigáció képessége. Ezzel együtt a tárgyak manipulálására és a világgal való érdemi interakciókra való képességük továbbra is korlátozott.

A következő nagy lépés szerintem annak bizonyítása lesz, hogy a térbeli intelligencia képes az autonóm ágensek új generációját működtetni: olyanokat, amelyek már valóban cselekedni tudnak a fizikai világban, és tárgyakat is képesek kezelni. Ez számos új alkalmazást és piacot nyithat meg.

Természetesen ez nem kizárólag az AI kérdése, vagy legalábbis nem kizárólag a térbeli intelligenciáé. Ehhez jobb robotikai eszközökre is szükség van a környezettel való interakcióhoz. Vagyis két terület párhuzamos előrehaladásában gondolkodunk.

Sajtókapcsolat:

- Torda Júlia, kommunikációs vezető
- kommunikacio@hun-ren.hu



© HUN-REN
Federico Tombari, a Google zürichi kutatási igazgatója.

Eredeti tartalom: HUN-REN Magyar Kutatási Hálózat

Továbbította: Helló Sajtó! Üzleti Sajtószolgálat

Ez a sajtóközlemény a következő linken érhető el:
<https://hellosajto.hu/?p=31053>