

Global Catastrophic Risk INSTITUTE

<http://gcrinstitute.org>

A mesterséges általános intelligencia projektek áttekintése az etika, a kockázat és a politika számára

Seth D. Baum
Globális Katasztrófakockázati Intézet

Globális Katasztrófakockázati Intézet
munkadokumentum 17-1

Cite as: Baum, 2017. A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy.
A Globális Katasztrófakockázati Intézet 17-1. számú munkadokumentuma.

*A szerző megjegyzése: Megjegyzéseket szívesen fogadunk. Felülvizsgálat várható. Ez a változat 2017.
november 12.*

*A munkadokumentumokat az ötletek megosztása és a vita előmozdítása érdekében tesszük közzé. Nem
feltétlenül mentek át szakértői értékelésen. A benne foglalt nézetek a szerzők nézetei, és nem
feltétlenül a Globális Katasztrófakockázati Intézet nézetei.*

Összefoglaló

A mesterséges általános intelligencia (AGI) olyan mesterséges intelligencia, amely számos területen képes érvelni. Bár a legtöbb mesterséges intelligencia kutatás és fejlesztés (K+F) a szűk értelemben vett mesterséges intelligenciával, nem pedig az AGI-vel foglalkozik, van néhány kifejezetten AGI-re irányuló K+F. Ha az AGI létrejön, annak hatásai mélyrehatóak lehetnek. Attól függően, hogy hogyan tervezik és használják, segíthet megoldani a világ problémáit, vagy katasztrófát okozhat, esetleg az emberiség kihalását is.

Ez a dokumentum az első felmérést mutatja be az aktív AGI K+F projektekről az etika, a kockázat és a politika területén. A felmérés megkísérli azonosítani az összes aktív AGI K+F projektet, és hét attribútum alapján jellemzi őket:

- A projekt alapjául szolgáló intézmény típusa
- Nyílt forráskódú kódot tesz-e közzé a projekt
- A projektnek vannak-e katonai kapcsolatai
- Az ország(ok), ahol a projekt székhelye található
- A projekt céljai az AGI-vel kapcsolatban
- A projektnek az AGI biztonsági kérdésekkel való foglalkozásának mértéke
- A projekt teljes mérete

Ennek érdekében a felmérés nyíltan közzétett információkat használ fel, amelyeket tudományos publikációkban, projekt-weboldalakon, népszerű médiacikkekben és egyéb weboldalakon találunk, beleértve 11 technikai felmérést, a *Journal of Artificial General Intelligence* 8 évfolyamát, 7 év AGI-konferenciák jegyzőkönyveit, 2 AGI-projektet online listáját, kulcsszavas kereséseket a Google webes keresőben és a Google Scholarban, a szerző előzetes ismereteit, valamint a fentiek alapján azonosított további irodalmat és weboldalat.

A felmérés 45 AGI K+F projektet azonosított, amelyek 6 kontinens 30 országában találhatóak, és amelyek közül sok nagyvállalatoknál és tudományos intézményekben működik, és amelyek közül néhány nagyszabású és nagy finanszírozású. A projektek közül sokan közös személyzet, közös anyaszervezetek vagy projekt-együttműködés révén kapcsolódnak egymáshoz.

A hét attribútum mindegyikénél megfigyelhető néhány fő tendencia az AGI K+F projektekkal kapcsolatban:

- A legtöbb projektet vállalatoknál vagy felsőoktatási intézményeknél valósítják meg.
- A legtöbb projekt nyílt forráskódú kódot tesz közzé.
- Kevés projektnek van katonai kapcsolata.
- A legtöbb projekt székhelye az Egyesült Államokban van, és szinte mindegyik az Egyesült Államokban vagy valamelyik szövetséges országban. Az egyetlen olyan projekt, amely teljes egészében az USA-n és szövetségesein kívül létezik, Kínában vagy Oroszországban található, és ezek a projektek mind erős tudományos és/vagy nyugati kapcsolatokkal rendelkeznek.
- A legtöbb projekt az emberiség egészének javára vagy a tudás határainak kiterjesztésére irányuló célokat fogalmaz meg, amelyeket a dokumentum "humanitárius" és "intellektualista" céloknak nevez.
- A legtöbb projekt nem foglalkozik aktívan az AGI biztonsági kérdéseivel.
- A legtöbb projekt a kis- és közepes méretűek közé tartozik. A három legnagyobb projekt a DeepMind (a Google londoni székhelyű projektje), a Human Brain Project (a svájci Lausanne-ban működő tudományos projekt) és az OpenAI (San Franciscóban működő nonprofit szervezet).

Több attribútumot vizsgálva néhány további tendencia is megfigyelhető:

- Van egy olyan tudományos projektcsoport, amely az ismeretterjesztést tűzi ki célul (azaz intellektualista), és nem foglalkozik a biztonsággal.

- A vállalati projektek egy csoportja az emberiség javát szolgálja (azaz humanitárius), és a biztonság terén tevékenykedik.
- A katonai kapcsolatokkal rendelkező projektek többsége katonai finanszírozásban részesülő amerikai akadémiai csoport, beleértve a tudományos-intellektuális-nem aktív biztonsági klaszteren belüli alsoportot is.
- Mind a hat kínai projekt kicsi, bár némelyik nagy szervezeteknél működik, amelyek rendelkeznek a gyors méretnöveléshez szükséges erőforrásokkal.

A következő oldalon található ES1. ábra az adatok

áttekintését mutatja be. Az adatokból a következő

következtetések vonhatók le:

Az etika tekintetében a fő tendencia a projektek megosztottsága az emberiség javára és a tudás előmozdítására irányuló célok között, az előbbi nagyrészt a vállalati projektekből, az utóbbi pedig a tudományos projektekből származik. Bár a projektek nem csak ezeket a célokat fogalmazzák meg, úgy tűnik, hogy laza konszenzus van e célok valamilyen kombinációját illetően.

Ami a kockázatot, különösen az AGI-katasztrófa kockázatát **illeti, vannak** jó és rossz hírek. A rossz hír az, hogy a legtöbb projekt nem foglalkozik aktívan az AGI biztonsági kérdéseivel. Az akadémiai projektek különösen hiányoznak a biztonsággal kapcsolatban. További aggodalomra ad okot, hogy a vállalati projektek a profitot a biztonság és a közérdek elé helyezik. A jó hír az, hogy sok lehetőség van arra, hogy a projektek együttműködjenek a biztonsági kérdésekben, köszönhetően a célokkal kapcsolatos részleges konszenzusnak, a projektek USA-ban és szövetségeseiben való koncentrálódásának, valamint a különböző projektek közötti különböző kapcsolódási pontoknak.

A politikát illetően több következtetést is le lehet vonni. Először is, a projektek USA-ban és szövetségeseiben való összpontosítása nagymértékben megkönnyítheti az AGI-ra vonatkozó közpolitika kialakítását. Másodsor, a tudományos projektek nagy száma arra utal, hogy a kutatópolitika fontos szerepet játszik, például a kockázatos kutatások értékelésére létrehozott felülvizsgálati bizottságok. Harmadszor, a vállalati projektek nagy száma arra utal, hogy figyelmet kell fordítani az AGI K+F politikai gazdaságára. Például, ha az AGI K+F a vállalatoknak rövid távú projekteket hoz, akkor a politika sokkal nehezebbé válhat. Végül a nyílt forráskódú projektek nagy száma egy másik politikai kihívást is jelent, mivel az AGI K+F-et bárki bárhol a világon végezheti.

A tanulmánynak vannak bizonyos korlátai, ami azt jelenti, hogy az AGI K+F tényleges helyzete eltérhet az itt bemutatottaktól. Ennek oka, hogy a felmérés kizárólag nyilvánosan közzétett információkon alapul. Lehetséges, hogy a felmérésből kimaradt néhány AGI K+F projekt. Ezért a felmérésben azonosított projektek száma, 45, egy alsó határérték. Továbbá lehetséges, hogy a projektek tényleges jellemzői eltérnek a nyilvánosan közzétett információkban szereplőktől. Például a legtöbb vállalati projekt nem jelölte meg a nyereség célját, noha feltehetően sokan nyereségre törekszenek. Ezért e tanulmány eredményei nem feltétlenül tükrözik az AGI K+F tényleges jelenlegi helyzetét. Ennek ellenére a tanulmány az AGI K+F eddigi legátfogóbb leírását nyújtja etikai, kockázati és politikai szempontból.

biztonsági tevékenységet.

- **Méret:** a betűméret a projekt méretét jelzi.

Köszönetnyilvánítás

Tony Barrett, Robert de Neufville, Steven Umbrello, Ben Goertzel, Gordon Irlam és Roman Yampolskiy visszajelzést nyújtottak egy korábbi tervezethez. Matthijs Maas némi kutatási segítséget nyújtott. Az esetlegesen fennmaradt hibák vagy egyéb hiányosságok kizárólag a szerzőt terhelik.

Ezt a munkát a Gordon R. Irlam Jótékonyági Alapítvány támogatása tette lehetővé. Az ebben a tanulmányban kifejtett nézetek kizárólag a szerző sajátjai, és nem feltétlenül tükrözik a Global Catastrophic Risk Institute vagy a Gordon R. Irlam Charitable Foundation nézeteit.

Tartalomjegyzék

Fő szöveg

1. Bevezetés.....	7
2. Korábbi irodalom	8
3. Kutatási kérdések	11
4. Módszertan.....	12
5. Főbb eredmények	17
6. Következtetés	29
1. függelék. Aktív AGI projektek	32
2. függelék. Egyéb figyelemre méltó projektek.....	77
Hivatkozások.....	92

1. Bevezetés

A mesterséges általános intelligencia (AGI) olyan mesterséges intelligencia, amely számos területen képes érvelni. Az emberi elme rendelkezik általános intelligenciával, de a legtöbb mesterséges intelligencia nem. Így például a DeepBlue meg tudja verni Kaszparovot sakkban, és talán néhány más alapvető feladatban, például szorzásban, de semmi másban nem tudja legyőzni. Az AGI a kezdeti mesterséges intelligencia egyik elsődleges célja volt, és sokáig a "nagy álomnak" vagy "szent gráfnak" számított.¹ Az AGI megalkotásának technikai nehézségei miatt a szakterület nagy része a mesterséges intelligencia szűkebb, közvetlenebbül gyakorlatias formáira összpontosít, de néhány elkötelezett AGI-kutatás és -fejlesztés (K+F) továbbra is folyik.

Az AGI szintén mélyreható társadalmi probléma, vagy legalábbis az lehet, ha megépül. Az AGI kiegészíthetné az emberi értelmet, és nagyobb kapacitást kínálna a világ problémáinak megoldására. Vagy akár rosszindulatúan, hatalmi játszámra is felhasználhatja az, aki irányítja. Vagy az emberiség nem tudja irányítani. Bármely mesterséges intelligencia képes túljárni az ember eszén bizonyos területeken (pl. szorzás); egy AGI minden területen túljárhat az ember eszén. Ebben az esetben a végeredmény az AGI céljaitól függhet: az emberiség, a világ, saját maga vagy valami egészen más pozitív célt szolgál-e. Sőt, az AGI kutatói néha azt javasolják, hogy az AGI az emberiség kihalását vagy hasonló katasztrófát okozhat (lásd az alábbi irodalmi áttekintést).

Az AGI nagy potenciális tétje etikai, kockázati és politikai kérdéseket vet fel. Melyik AGI-t, ha egyáltalán, érdemes megépíteni? Mekkora a katasztrófa kockázata, ha AGI-t építenek? Milyen szakpolitikai lehetőségek állnak rendelkezésre az AGI-katasztrófa elkerülésére, és - amennyiben ez kívánatos - a biztonságos és hasznos AGI-k lehetővé tételére? Ezek mind olyan kérdések, amelyeket aktívan vizsgálunk. Az eddigi szakirodalom azonban inkább elméleti és spekulatív, és kevésbé támaszkodik az AGI tényleges helyzetére. Tekintettel arra, hogy az AGI-t talán csak sok év múlva építik meg, némi spekuláció elkerülhetetlen. Az AGI K+F azonban már most is folyik. A jelenlegi K+F-re vonatkozó információk irányt mutathatnak az etikával, kockázattal és politikával kapcsolatos jelenlegi tevékenységeknek, és betekintést nyújthatnak abba, hogy a jövőbeni K+F hogyan nézhet ki.

Ez a tanulmány az aktív AGI K+F projektek első felmérését mutatja be etikai, kockázati és politikai szempontból. Az AGI-vel és a kapcsolódó technológiákkal kapcsolatos K+F-ről már több korábbi felmérés is készült (Chong et al. 2007; Duch et al. 2008; Langley et al. 2008; de Garis et al. 2010; Goertzel et al. 2010; Samsonovich 2010; Taatgen és Anderson 2010; Thórisson és Helgasson 2012; Dong és Franklin 2014; Goertzel 2014; Kotseruba et al. 2016). Ezek a felmérések azonban mind technikai szempontokra vonatkoznak, például arra, hogy hogyan tervezik magát az AGI-t, milyen előrelépések történtek vele kapcsolatban, és mennyire ígéretesek a különböző megközelítések az AGI megvalósítása szempontjából. Ez a tanulmány az etika, a kockázat és a politika szempontjából releváns információkat mutat be és elemez - például azt, hogy a projektek mely politikai joghatóságokban találhatóak, és mennyire elkötelezettek az AGI biztonsága iránt. Továbbá, míg a korábbi felmérések az AGI K+F projektek kiválasztott, figyelemre méltó példáira összpontosítanak, ez a dokumentum a projektek teljes körét igyekszik dokumentálni. Az etika, a kockázat és a politika szempontjából hasznos, ha a teljes halmaz rendelkezésre áll - például azért, hogy tudjuk, mely politikai joghatóságokat kell bevonni az AGI közpolitikájába.

Az 1.1. szakasz az AGI-vel kapcsolatos terminológiát ismerteti. A 2. szakasz áttekinti az AGI etikájával, kockázatával és politikájával kapcsolatos korábbi szakirodalmat. A 3. szakasz bemutatja a tanulmányban feltett kutatási kérdéseket. A 4. szakasz összefoglalja a tanulmány felméréséhez használt módszertant. Az 5. szakasz bemutatja az AGI K+F projektekkal és más figyelemre méltó projektekkal kapcsolatos felmérés főbb eredményeit. A 6. szakasz következtetéseket tartalmaz. Az 1. függelék tartalmazza az aktív AGI K+F projektek teljes felmérését. A 2. függelék azokat a figyelemre méltó projekteket mutatja be, amelyeket a felmérés során figyelembe vettek, de kizártak, mert nem

feleltek meg a felvételi kritériumoknak.

1.1 Terminológia

Az AGI egyike a fejlett és potenciálisan átalakító jövőbeli mesterséges intelligenciára használt számos kifejezésnek. A kifejezések jelentése némileg eltér egymástól, és érdemes röviden különbséget tenni közöttük.

¹ Lásd pl. Legg (2008, 125. o.); Pollock (2008); <http://www.humanobs.org>.

- *Az AGI kifejezetten az intelligencia képességek széles skálájával rendelkező mesterséges intelligencia, beleértve "a különböző célok elérésének és a különböző feladatok elvégzésének képességét különböző kontextusokban és környezetekben"* (Goertzel 2014, 2. o.). Az AGI nem feltétlenül fejlett - egy mesterséges intelligencia lehet általános anélkül, hogy magasan fejlett lenne -, bár az általános intelligencia bizonyos fokú fejlettséget igényel, és az AGI-ről gyakran feltételezik, hogy magasan fejlett képességekkel rendelkezik. Az AGI egy külön tudományterület is, saját társasággal (<http://www.agi-society.org>), folyóirattal (*Journal of Artificial General Intelligence*) és konferenciasorozattal (<http://agi-conf.org>).
- *A kognitív architektúra egy intelligens entitás általános felépítése. Beszélhetünk az emberek vagy más állatok agyának kognitív architektúrájáról, de a kifejezést elsősorban az emberi és nem emberi állatok kogníciójának elméleti és számítási modelljeire használjuk. A kognitív architektúrák lehetnek szűk körűek is, amelyek bizonyos kognitív folyamatokra, például a figyelemre vagy az érzelmekre összpontosítanak* (Kotseruba et al. 2016). Gyakran azonban általánosak, így tanulmányozásuk átfedésben van az AGI tanulmányozásával. *A biológiailag inspirált kognitív architektúrák egy külön tudományterület, saját társasággal* (<http://bicasociety.org>), folyóirattal (*Biologically Inspired Cognitive Architectures*) és konferenciasorozattal (<http://bicasociety.org/meetings>).
- *Az agyi emulációk a biológiai agyak számítási példányai. Az agyi emulációkat néha az AGI-tól elkülönülve osztályozzák* (pl. Barrett és Baum 2017a), de ezek általános intelligenciával rendelkező számítási entitások, ezért ez a tanulmány az AGI egyik típusaként kezeli őket.
- *Az emberi szintű mesterséges intelligencia az emberhez hasonló intelligenciával rendelkező mesterséges intelligencia, vagy "emberi szintű, ésszerűen emberhez hasonló AGI"* (Goertzel 2014, 6. o.). Fontos finomság, hogy egy AGI lehet ugyanolyan fejlett, mint az ember, de az intelligencia meglehetősen más típusú: nem feltétlenül utánozza az emberi megismerést. Például az AI sakkprogramok bizonyos esetekben nyers erővel fognak keresni, ahol az emberek intuíciót használnak, mégis az AI képes az emberi szinten vagy azt meghaladó teljesítményre.
- *A szuperintelligencia az emberi intelligenciát jelentősen meghaladó mesterséges intelligencia. Az ultraintelligencia kifejezést is használták már ebben az összefüggésben* (Good 1965), de kevésbé elterjedt. Gyakran felvetik, hogy a szuperintelligencia egy kezdeti *magvető mesterséges intelligenciából* származik, amely *rekurzív önfejlesztésen* megy keresztül, egyre okosabbá és okosabbá válva. A mag AI nem feltétlenül AGI, de gyakran feltételezik, hogy az.

Ez az írás az AGI-ra összpontosít, mivel a kifejezést a K+F kontextusban gyakran használják, és fontos az etika, a kockázat és a politika szempontjából. A szűk kognitív architektúrák (és a szűk mesterséges intelligencia) kevésbé valószínű, hogy átalakító hatással lesznek a világra. Az emberi szintű mesterséges intelligencia és a szuperintelligencia nagyobb valószínűséggel járhat átalakító következményekkel, de ezek a kifejezések a K+F-ben nem gyakoriak. Megjegyzendő, hogy nem minden, a jelen tanulmányban szereplő projekt azonosítja magát kifejezetten AGI-ként, de mindegyikben megvan a potenciál, hogy AGI legyen, és hasonlóképpen átalakító következményekkel járhat. A felmérés nem zárja ki azokat a projekteket, amelyek kifejezetten emberi szintű mesterséges intelligenciát vagy szuperintelligenciát próbálnak létrehozni.

2. Korábbi Irodalom

2.1 Etika

Az AGI etikájának talán legalapvetőbb kérdése az, hogy az AGI-t intellektuális törekvésként vagy olyan dologként kezeljük-e, amely hatással lehet a társadalomra és a világ egészére. Más szóval, az

AGI K+F-et a tudás élvonalának előmozdítása vagy a társadalom haszna érdekében végzik? Gyakran ez a két cél szorosan összefügg, amint azt a tudomány és a technológia központi szerepe is bizonyítja az életkörülmények javításában világszerte. Ugyanakkor nem mindig kapcsolódnak össze, és néha ellentétesek is lehetnek. A jelen tanulmány szempontjából különösen fontos, hogy a potenciálisan veszélyes új technológiák kutatása jelentős tudományos és intellektuális felismeréseket eredményezhet, de a társadalom számára végül káros lehet.

A kutatók minden kutatási területen gyakran erős intellektuális értékekkel és motivációkkal rendelkeznek, és ez alól az AGI sem kivétel. Az a kérdés, hogy a kutatást az intellektuális érdemek vagy a szélesebb körű társadalmi/globális hatások alapján értékeljük-e, állandó vitatéma az akadémiai világban (Schienke et al.

2009). A legtöbb területhez hasonlóan a mesterséges intelligencia hagyományosan az intellektuális érdemeket hangsúlyozza, bár vannak olyan felhívások, amelyek ennek megváltoztatására szólítanak fel (Baum 2017a). Az AGI intellektuális vonzereje különösen erős lehet, mivel hosszú távú "nagy álom" vagy "szent grál" státuszáról van szó, bár a tágabb értelemben vett hatások is erős vonzerővel bírhatnak, tekintettel az AGI nagy potenciális tétjeire. Gyakorlati szempontból egy intellektuális motivációjú AGI-projekt nagyobb valószínűséggel tekinti az AGI létrehozását önmagában is méltó célnak, és kevés figyelmet fordít a lehetséges veszélyekre vagy egyéb szélesebb körű hatásokra, mint egy olyan projekt, amelyet szélesebb körű hatások motiválnak.

Az AGI etikájának másik területe az AGI céljainak meghatározása. Az AGI-etikával foglalkozó korábbi szakirodalom elsősorban erre összpontosít. Az egyik gondolatmenet az "indirekt normativitást" vagy "koherens extrapolált akaratot" javasolja, amelyben az AGI-t úgy tervezik, hogy az intellektusát arra használja, hogy kitalálja, mit akar az emberiség (Yudkowsky 2004; Muehlhauser és Helm 2012; Bostrom 2014). Ezt a javaslatot részben az eljárási igazságossággal kapcsolatos aggályok motiválják - mindenkinek bele kell szólnia az AGI etikájába, nem csak az AGI tervezőinek -, részben pedig az emberi etika finomságainak közvetlenül az AGI-ba történő programozásának technikai nehézségeivel kapcsolatos aggályok. Ettől függetlenül ezek a javaslatok mind arra szólítanak fel, hogy az AGI az *emberiség* etikáját kövesse, és ne valami mást.²

Egy alternatív gondolkodásmód szerint az AGI-nak olyan új entitásokat kellene létrehoznia, amelyek erkölcsileg magasabb rendűek az embereknél. Ez a gondolkodás a "transzhumanizmus" vagy "poszthumanizmus" területére esik; az AGI-kutatók de Garis (2005) és Goertzel (2010) a "kozmisztika" kifejezést használják. Ez a nézet szerint az AGI-nek a kozmosz egészének, nem csak az emberiségnek kell hasznára válnia, és azt javasolja, hogy a kozmosz javát az AGI által létrehozott, erkölcsileg magasabb rendű lények segíthetik elő. Míg Goertzel (2010) azt hangsúlyozza, hogy az "örökölt embereknek" maguknak kell eldönteniük, hogy folytatják-e ebben az új világban, addig de Garis (2005) azt javasolja, hogy ez a világ akkor is érdemes lehet kialakítani, ha az örökölt emberek megszűnnének. Ezzel szemben Yampolskiy (2013) amellett érvel, hogy az AGI-eket csak akkor érdemes megépíteni, ha azok az emberi alkotóik számára hasznos, feláldozható eszközök.

Végezetül pedig arról is folyt némi vita, hogy egyáltalán meg kell-e építeni az AGI-t, bár ez eddig kevésbé állt a figyelem középpontjában. Az AGI-ről szóló legtöbb vita vagy támogatja az építést, vagy nem foglalkozik komolyan a kérdéssel, mert feltételezi annak elkerülhetetlenségét, ahogyan azt Totschnig (2017) is tárgyalja. Az AGI kiépítése elleni érvek egy része a katasztrófakockázattal kapcsolatos aggodalmakban gyökerezik (pl. Joy 2000); a kockázatról bővebben alább. Mások azzal érvelnek, hogy még a biztonságos AGI-t sem szabadna megépíteni. Ezek közé tartoznak Kaczynski (1995, 174. bekezdés) szélsőséges anarchista nézetei és Totschnig (2017) józanabb vitája, bár az AGI-vel szemben sokkal kevesebb nyílt ellenállás mutatkozik, mint a hasonló átalakító technológiák, például az emberi képességek fokozása ellen.

2.2 Kockázat

Az AGI katasztrófa lehetősége abban az elképzelésben gyökerezik, hogy az AGI túljárhat az emberiség eszén, átveheti az irányítást a bolygó felett, és bármilyen célt elérhet, amire beprogramozták. Hacsak nem olyan célokkal programozzák be, amelyek biztonságosak az emberiség számára, vagy bármi más számára, ami az ember számára fontos, akkor katasztrófát fog eredményezni. Hasonlóképpen, a katasztrófa elkerülése érdekében az AGI K+F projekteknek megfelelő biztonsági óvintézkedéseket kell tenniük.

E kockázat nagyságáról és a szükséges biztonsági erőfeszítések mértékéről eltérőek a vélemények. Egyesek azt javasolják, hogy alapvetően nehéz biztonságos célokat kitűző AGI-t tervezni - még a látszólag apró hibák is katasztrofális eredményekkel járhatnak, ezért az AGI K+F projekteknél nagy figyelmet kell fordítani a biztonságra (Yudkowsky 2004; Muehlhauser és Helm 2012; Bostrom 2014). Mások azzal érvelnek, hogy egy AGI

² Baum (2017b) kritikával illeti a koherens extrapolált akarati javaslatokat, mivel azok csak az emberi etikára összpontosítanak, és nem foglalkoznak azokkal az etikai nézetekkel, amelyeket más biológiai fajok, maga a mesterséges intelligencia vagy más entitások képviselhetnek.

biztonságos célok elérésére képezzék ki, és hogy ez a folyamat nem kivételesen törékeny, így az AGI K+F projekteknek figyelniük kell a biztonságra, de nem szokatlan mértékben (pl. Goertzel és Pitt 2012; Bieger et al. 2015; Goertzel 2015; 2016; Steunebrink et al. 2016). Végül egyesek teljesen elhárítják a kockázatot, vagy azért, mert az AGI nem lesz képes túljárni az emberiség eszén (pl. Bringsjord 2012; McDermott 2012), vagy azért, mert túl valószínűtlen vagy túl távoli a jövőben ahhoz, hogy figyelmet érdemeljen (pl. Etzioni 2016; Stilgoe és Maynard 2017).

Az egyik általános aggodalom az, hogy a versengő projektek versenyt futnak majd az AGI első elindításáért, ami katasztrófális következményekkel járhat (Joy 2000; Shulman 2009; Dewey 2015; Armstrong et al. 2016). Az AGI-verseny megnyerésének vágya különösen erős lehet azon elképzelések miatt, amelyek szerint az AGI olyan erős lehet, hogy rendkívüli előnyhöz juttatná az elsőként indulókat. Ez kollektív cselekvési problémát okoz: a csoport érdeke, hogy minden projekt magas biztonsági színvonalat tartson fenn, de az egyes projektek egyéni érdeke, hogy a verseny megnyerése érdekében spóroljanak a biztonságon. Armstrong és társai (2016) játékelméleti elemzést mutatnak be az AGI-verseny forgatókönyvéről, és azt találták, hogy a kockázat nő, ha a) több K+F projekt van, b) a projektek erősebben preferálják saját AGI-jüket a többiekéhez képest, így kevésbé valószínű, hogy időigényes biztonsági intézkedésekbe fektetnek, és c) a projektek hasonló képességgel rendelkeznek az AGI építésére, ami nagyobb relatív előnyt jelent számukra, ha spórolnak a biztonságon.

Barrett és Baum (2017a; 2017b) az AGI-ből eredő katasztrófa kockázati modelljét dolgozza ki, különös tekintettel az olyan AGI-re, amely rekurzív módon önfejlesztődik, amíg szuperintelligenssé nem válik, és el nem nyeri a bolygó feletti uralmat.³ Ahhoz, hogy ez a katasztrófa bekövetkezzen, hat feltételnek kell mind teljesülnie: (1) a szuperintelligenciának lehetségesnek kell lennie; (2) létre kell hozni a kezdeti ("mag") mesterséges intelligenciát, amely elindítja az önfejlesztést; (3) nem lehet sikeresen megfékezni az önfejlesztő folyamatot vagy a keletkező szuperintelligenciát, hogy a szuperintelligencia átvegye az uralmat a bolygó felett; (4) az emberek nem tehetik biztonságossá a mesterséges intelligencia céljait úgy, hogy a célok megvalósítása elkerülje a katasztrófát; (5) a mesterséges intelligencia nem teheti biztonságossá céljait önmagától, függetlenül a céljai biztonságossá tételére irányuló emberi erőfeszítésektől; és (6) a mesterséges intelligenciát nem tarthatják vissza céljai elérésétől az emberek, más mesterséges intelligenciák vagy bármi más. A teljes kockázat attól függ, hogy e feltételek mindegyike milyen valószínűséggel teljesül. Hasonlóképpen, a kockázatkezelés a (2), (3), (4) és (6) feltételek valószínűségének csökkentésére irányulhat. A kockázatkezelés az AGI-politika egyik aspektusa.

2.3 Politika

Az AGI-politika tágran értelmezhető az AGI K+F befolyásolására irányuló valamennyi erőfeszítésként, amely magában foglalhatja a kormányok és más intézmények hivatalos politikáit, valamint az AGI iránt érdeklődő vagy érintett emberek informális politikáit, beleértve magukat a kutatókat is. Az AGI-politika törekedhet többek között az AGI K+F finanszírozására vagy más módon történő támogatására, bizonyos etikai nézetek AGI-be való beépítésének ösztönzésére vagy az AGI kockázatának csökkentésére. Sotala és Yampolskiy (2015) az AGI-politikai elképzelések széles körét tekinti át, a kockázatkezelésre összpontosítva.

Az AGI-politikával foglalkozó korábbi szakirodalom nagy része a feszültséget hangsúlyozza (1) a hipotetikus AGI-fejlesztők között, akik a biztonság vagy az etika nem megfelelő figyelembevételével kívánnak eljárni, és (2) egy olyan közösség között, amely aggódik a nem biztonságos és etikátlan AGI miatt, és keresi a módját annak, hogy az AGI K+F-et biztonságosabb és etikusabb irányba terelje. Joy (2000) szerint a katasztrófa kockázata túl nagy, és az AGI-kutatás és -fejlesztés általános elhagyására szólít fel. Hibbard (2002) és Hughes (2007) ehelyett olyan szabályozási rendszereket szorgalmaz, amelyek a veszélyes AGI-k elkerülésére irányulnak anélkül, hogy teljesen lemondanának a technológiáról. Yampolskiy és Fox (2013) a kutatóintézeteknél működő felülvizsgálati bizottságokat

javasol a túl veszélyes AGI-kutatás korlátozására.

Baum (2017c) az AGI K+F közösségek szociálpszichológiájának figyelembevételére szólít fel annak érdekében, hogy a biztonsági és etikai intézkedések sikeresek legyenek, és hogy az AGI K+F közösségeket arra ösztönözzék, hogy többet tegyenek saját maguk.

³ A modell az általános mesterséges intelligenciáról beszél, amelynek az AGI csak egy típusa, más típusú mesterséges intelligenciák mellett, amelyek szintén képesek rekurzív önfejlesztésre. Ez a különbségtétel a jelen tanulmány szempontjából nem döntő fontosságú.

Az egyik politikai kihívás abból a tényből fakad, hogy az AGI-t a világ bármely pontján ki lehet fejleszteni, amely képes a kutatói tehetségeket vonzani és szerény számítástechnikai erőforrásokat összeállítani. Wilson (2013) ezért egy olyan nemzetközi szerződést vázol fel, amely biztosíthatná, hogy a veszélyes AGI-munka ne kerüljön át szabályozatlan országokba. Scherer (2016) elemzi az AGI amerikai kormány általi szabályozásának lehetőségeit, megjegyzi a nemzeti szabályozás előnyeit a szubnacionális szabályozással szemben, és azt javasolja, hogy ez egy nemzetközi szerződés előzménye lehetne. Goertzel (2009) elemzi az AGI kínai vs. nyugati fejlődésének kilátásait, és megállapítja, hogy ez bizonyos tényezők relatív fontosságának függvényében mindkét irányba mehet. Bostrom (2014) az AGI K+F nemzetközi ellenőrzését szorgalmazza, esetleg az ENSZ égisze alatt. Az esetlegesen titokban működő AGI K+F projektek felderítése érdekében Hughes (2007), Shulman (2009) és Dewey (2015) globális felügyeleti rendszereket javasol; Goertzel (2012a) azt javasolja, hogy egy korlátozott AGI végezze a felügyeletet.

Végezetül, a korábbi szakirodalom alkalmanként kitért arra az intézményi környezetre, amelyben az AGI K+F zajlik. Yampolskiy és Fox (2013) felülvizsgálati bizottságokra vonatkozó javaslata az egyetemekre összpontosít, hasonlóan az egyetemi humán tárgyú kutatásokra vonatkozó meglévő felülvizsgálati bizottságokhoz. Goertzel (2017a) aggodalmát fejezi ki a nagyvállalatoknál folyó AGI K+F-fel kapcsolatban, mivel azok hajlamosak a globális gazdagság koncentrálására és a kormánypolitika saját javukra történő torzítására; ehelyett a nyílt forráskódú AGI K+F mellett érvel. Bostrom (2017) ezzel szemben azzal érvel, hogy a nyílt forráskódú AGI K+F veszélyesebb lehet, mivel mindenki hozzáférhet ugyanahhoz a kódhoz, és ezáltal szorosabbá válik a verseny az AGI első megépítéséért. Shulman (2009) attól tart, hogy a nemzetek versenyezni fognak az AGI megépítéséért, hogy "megkérdőjelezhetetlen gazdasági és katonai dominanciát" szerezzenek, és hogy az AGI-re való törekvés geopolitikai szempontból destabilizáló hatású lehet. Baum et al. (2011) AGI-szakértőket kérdeztek meg a vállalatok, a nyílt forráskódú közösségek és az amerikai hadsereg AGI-kutatásának relatív előnyeiről, és a szakértők eltérő véleményeket találtak, különösen a nyílt forráskódú és a katonai K+F között.

Azt is meg kell jegyezni, hogy a nagy kormányok jelentős tevékenységet folytattak a mesterséges intelligenciával kapcsolatban. A kínai kormány például nemrég jelentett be egy jelentős kezdeményezést, amelynek célja, hogy a következő évtizedekben globális vezető szerepet töltsön be a mesterséges intelligencia területén (Webster et al. 2017). A kínai kezdeményezés nagyban hasonlít - és talán származik is belőle - az Obama elnök alatt az Egyesült Államok által az AI-ról közzétett jelentéssorozatra (Kania 2017). Vlagyimir Putyin orosz elnök nemrégiben beszélt a mesterséges intelligencia fontosságáról, "a jövőnek" nevezve azt, megjegyyezve "kolosszális lehetőségeket, de nehezen megjósolható veszélyeket is", és kijelentve, hogy "aki vezető lesz ezen a területen, az lesz a világ ura" (RT 2017). Ezek a különböző kezdeményezések és kijelentések azonban nem kifejezetten az AGI-ről szólnak, és úgy tűnik, hogy elsősorban a szűk értelemben vett mesterséges intelligenciára vonatkoznak. Egyes politikai közösségek még az AGI-vel való társítást is elkerülik, mint például az Obama Fehér Ház által a fent említett jelentésekkel kapcsolatban szponzorált rendezvénysorozat (Conn 2016). A magas szintű kormányzati érdeklődés az AI iránt tehát nem feltétlenül jelenti a kormányzat AGI iránti elkötelezettségét. Az AGI iránti magas szintű kormányzati érdeklődés egyik példája az Európai Bizottság által a Human Brain Project nagyszabású támogatása, abban a reményben, hogy egy számítógépes agyszimuláció újjáélesztheti az európai gazdaságot (Theil 2015).

3. Kutatás Kérdések

A korábbi szakirodalom több olyan kérdést is felvet, amelyekről az aktív AGI K+F projektek felmérése adhatna tájékoztatást:—

Hány AGI K+F projekt van? Armstrong et al. (2016) megállapította, hogy az AGI kockázata nő,

ha több K+F projekt van, ami miatt kevésbé valószínű, hogy együttműködnek a biztonság terén. Hasonlóképpen, a más kontextusokban a kollektív cselekvéssel foglalkozó szakirodalom gyakran javasolja, hogy bizonyos körülmények között a kisebb csoportok sikerebbek lehetnek az együttműködésben, bár más körülmények között a nagy csoportok sikerebbek lehetnek (pl. Yang et al. 2013).⁴ Ezért érdemes egyszerűen tudni, hogy hány AGI K+F projekt van.

⁴ A kollektív cselekvés szakirodalma kifejezetten azt állapítja meg, hogy a kisebb csoportok gyakran sikerebbek az együttműködésben, ha a szoros interakciók csökkentik a szabadúszást, valamint a tranzakciók és a megfelelés ellenőrzésének költségeit, míg a nagyobb csoportok gyakran sikerebbek az együttműködésben.

Milyen típusú intézményekben működnek a projektek? Shulman (2009), Baum et al. (2011), Yampolskiy és Fox (2013), valamint Goertzel (2017a) szerint bizonyos intézményi kontextusok veszélyesebbek lehetnek, és a szakpolitikai válaszokat a projektek intézményeihez kell igazítani. Bár az intézményi kontextus pontos következményei még vita tárgyát képezik, segítené látni, hogy mely intézménytípusok adnak otthont az AGI K+F-nek.

Mennyi AGI K+F nyílt forráskódú? Bostrom (2017) és Goertzel (2017a) ellentétes nézőpontokat kínálnak a nyílt forráskódú AGI K+F előnyeiről. Ez egy másik, még megoldásra váró vita, amelynek időközben jót tene a nyílt forráskódú AGI K+F túlsúlyára vonatkozó adatok.

Az AGI K+F-nek mennyi katonai kapcsolata van? Shulman (2009) azt javasolja, hogy a nemzetek az AGI-t a katonai dominancia érdekében is alkalmazhatják. Ha ez igaz, akkor ennek jelentős geopolitikai következményei lehetnek. Bár a katonai K+F gyakran titkosított, érdemes megnézni, hogy a nyilvánosan elérhető adatok milyen katonai kapcsolatokkal rendelkeznek.

Hol található az AGI K+F projektek? Wilson (2013) egy nemzetközi szerződés mellett érvel a globális AGI K+F szabályozására, míg Scherer (2016) egy, az Egyesült Államokra jellemző szabályozási javaslatot dolgoz ki. Érdemes tehát megnézni, hogy a K+F mely országokban található.

Milyen céljai vannak a projekteknek? A 2.1. szakasz összefoglalja az AGI K+F projektek különböző céljainak megfelelő etikai nézeteket. Emellett Armstrong et al. (2016) megállapítja, hogy az AGI kockázata növekszik, ha a projektek erősebben preferálják a saját AGI-jüket másokéhoz képest, ami inkább akkor fordulhat elő, ha a projektek nem értenek egyet a célokat illetően. Érdemes tehát azonosítani és összehasonlítani a projektek céljait.

Mennyire elkötelezettek a projektek a biztonsági kérdésekben? A 2.2. szakasz áttekinti az AGI kockázatának nagyságáról és az AGI biztonságossá tételének nehézségéről alkotott nézeteket, a 2.3. szakasz pedig összefoglalja a szakpolitikai szakirodalmat, amely azon az aggodalmon alapul, hogy az AGI K+F projektek biztonsági eljárásai nem megfelelőek. Így a projektek biztonságosságára vonatkozó adatok mind az AGI-kockázat nagyságára, mind pedig az esetlegesen indokolt szakpolitikai válaszlépésekre vonatkozó információkkal szolgálhatnak.

Mekkora projektekről van szó? A nagyobb projektek alkalmasabbak lehetnek az AGI kiépítésére. Emellett Armstrong és munkatársai (2016) úgy találják, hogy az AGI kockázata növekszik, ha a projektek hasonló képességgel rendelkeznek az AGI kiépítésére. Az Armstrong et al. (2016) elemzése feltételezi, hogy a projektkapacitás egyenletesen oszlik meg. Érdemes megnézni, hogy valójában milyen a projektméret eloszlása, és melyek a legnagyobb projektek.

A projektkapacitás az AGI kiépítése szempontjából vitathatatlanul fontosabb, mint a projekt mérete. A projektkapacitást azonban nehezebb értékelni az ebben a tanulmányban alkalmazott módszertannal, amely a nyíltan közzétett nyilatkozatok elemzését alkalmazza. A projektméret mellett a projektkapacitás függhet a személyzet tehetségétől, a finanszírozás, a számítási teljesítmény vagy más erőforrások rendelkezésre állásától, valamint attól, hogy mennyire jól irányítják a projektet. Ezeket a tényezőket gyakran nem hozzák nyilvánosságra. Egy másik fontos tényező a projekt által követett technikai megközelítés életképessége, de ez nem eléggé ismert, és az AGI-szakértők között is nézeteltérések vannak. Bár a projektek kapacitását bizonyos fokú szigorral lehet értékelni, a jelen dokumentum módszertana nem alkalmas ilyen feladatra, ezért azt a jövőbeni munkára hagyjuk. Ehelyett a projektméret használható a projektkapacitás legalább durva helyettesítőjeként, bár itt óvatosságra int, mert ez egy tökéletlen vagy akár félrevezető helyettesítő lehet.

4. Módszertan

A tanulmány módszere az AGI K+F-projektek azonosításából, majd azok leírásából áll, több attribútum alapján. Az azonosítás és a leírás a tudományos publikációkban, a projektek weboldalain, a népszerű médiában megjelent cikkekben és más weboldalakon található, nyíltan közzétett

információkon alapult, különös tekintettel a következőkre

sikeresen együttműködnek, ha az együttműködésnek nagyobb összes rendelkezésre álló erőforrás az előnye (Yang et al. 2013). Így például lehet, hogy egy titok megőrzéséhez kis csoportot szeretnénk, de egy fix projekthez szükséges adománygyűjtéshez nagy csoportot.

a tekintélyesebb kiadványokra helyezik. Az azonosítást és a leírást elsősorban a jelen szerző végezte.⁵

A társadalomtudományi terminológiában ezt a módszert a kvalitatív adatok "kódolásának" nevezik (Coffey és Atkinson 1996; Auerbach és Silverstein 2003). Az adatok kvalitatívak, mivel az AGI K+F projektekről szóló szövegekből állnak; ezeket kvantitatív formába kódolják, például "egy egyetemi projekt és három kormányzati projekt". A kódolási sémát kezdetben a korábbi szakirodalom és a jelen szerzőnek a témáról alkotott elképzelései alapján dolgozták ki, és a kódolási folyamat során a jelen szerző által az adatok olvasása alapján frissítették (az úgynevezett "in vivo" kódolás).

Ez a módszertan alapvetően értelmező, amely az adatoknak a kutató általi értelmezésében gyökerezik. Az adatok egyes aspektusai nem igazán értelmezés kérdései - például az a tény, hogy a Dél-kaliforniai Egyetem egy akadémiai intézmény az Egyesült Államokban. Más szempontok jobban értelmezhetők. Ilyen például az, hogy mely projektek minősülnek AGI K+F-nek. Goertzel (2014, 2. o.) az AGI-közösséget "fuzzy set"-ként említi; ez találó leírás. A különböző kutatók különböző módon értelmezhetik ugyanazokat az adatokat. Az AGI K+F-ről szóló, nyíltan közzétett információk hatalmas terében való keresés során is eltérő adatokra bukkanhatnak. Így az eredményeket az AGI K+F egyfajta megközelítéseként kell értelmezni, és nem feltétlenül a téma valós vagy teljes tükörképeként. Az érdeklődő olvasókat arra kérjük, hogy maguk kérdezzék le az adatokat, és tegyék meg saját értelmezésüket. Az 1. függelék tartalmazza a kódolási ítéletek teljes leírását és magyarázatát a megfelelő adatokra hivatkozva. (Nem minden adatot idézünk - a talált adatok nagy része redundáns vagy korlátozottan releváns).

4.1 Az AGI K+F projektek azonosítása

Az AGI K+F jelölt projekteket a következők révén azonosították:

- A jelen szerző előzetes ismeretei.
- Kulcsszavas keresés az interneten és a tudományos adatbázisokban, elsősorban a Google webes keresőben és a Google Scholarban.
- Korábbi felmérések (Chong et al. 2007; Duch et al. 2008; Langley et al. 2008; de Garis et al. 2010; Goertzel et al. 2010; Samsonovich 2010; Taatgen és Anderson 2010; Thórisson és Helgasson 2012; Dong és Franklin 2014; Goertzel 2014; Kotseruba et al. 2016).
- A *Journal of Artificial General Intelligence* teljes elérhető tartalma (2009 decemberétől 2016 decemberéig).⁶
- A legutóbbi AGI-konferenciák (2011-2017) anyagai.
- Számos online lista (<http://www.agi-society.org/resources>; <http://bicasociety.org/cogarch/architectures.htm>; <http://realai.org/labs>; <http://2ai.org/landscape>; https://en.wikipedia.org/wiki/Cognitive_architecture#Notable_examples).
- A köszönetnyilvánításban felsorolt kollégák visszajelzései.
- A fentiek alapján azonosított további szakirodalom és weboldalak.

Minden egyes azonosított projektet három kategória egyikébe soroltak:

- Aktív AGI K+F projektek (1. függelék). Ezek olyan projektek, amelyek az AGI kiépítésén dolgoznak. A projektek vagy AGI-ként azonosítják magukat, vagy K+F-et végeznek valamilyen AGI-nek, emberi szintű intelligenciának vagy szuperintelligenciának tekintett dolog megépítésére.

⁵ Köszönet illeti Matthijs Maas-t a néhány projekt jellemzésében nyújtott kutatási segítségért.

⁶ Két további folyóiratot, a *Biologically Inspired Cognitive Architectures* és az *Advances in Cognitive Systems* folyóiratokat az AGI-vel kapcsolatos tartalmuk miatt vettük figyelembe, de ezeket nem vizsgáltuk át teljes egészében, mivel a

cikkeikben nem számoltak be AGI-projektekről.

- Egyéb figyelemre méltó projektek (2. függelék). Ezek közé tartoznak: (1) az inaktív AGI K+F projektek, amelyek az elmúlt három évben nem voltak látható frissítések; (2) olyan projektek, amelyek az AGI technikai aspektusain dolgoznak, de nem az AGI létrehozásán, például az AGI számára felhasználható hardver- vagy biztonsági mechanizmusokon dolgozó projektek; és (3) szűk értelemben vett AI projektek, például a nagy technológiai vállalatok AI-csoportjai.
- Egyéb projektek, amelyekről úgy ítélték meg, hogy nem érdemes szerepeltetni őket ebben a dokumentumban.

Az 5. szakasz adatelemzésében csak az aktív AGI K+F projektek szerepelnek. A többi figyelemre méltó projektről azért számolunk be, hogy dokumentáljuk a kapcsolódó munkákat, hogy tisztázzuk a szerző gondolkodását arról, hogy hol húzódnak az AGI K+F határai, és hogy segítsük a jelen kutatás által figyelmen kívül hagyott AGI K+F projektek azonosítását.

Azokat a projekteket, amelyek csak a mélytanulással és a kapcsolódó technikákkal kapcsolatos kutatás-fejlesztést végzik, kizártuk, kivéve, ha kifejezetten az AGI létrehozására irányuló kísérletet jelölnek meg. A mélytanulás már mutat némi általánosságot (LeCun et al. 2015), és egyesek szerint a mélytanulás kiterjeszhető az AGI-re (pl. Real AI 2017).

Mások azzal érvelnek, hogy a mélytanulás a jelenleg elért figyelemre méltó sikerei ellenére alapvetően korlátozott, és az AGI más típusú algoritmusokat igényel (pl. Strannegård és Nizamani 2016; Wang és Li 2016; Marcus 2017). A mélytanulást alkalmazó munkák közelmúltbeli robbanásszerű növekedése túlságosan megnehezíti a jelen tanulmány projektről projektre történő áttekintését. Ráadásul, ha az összes mélytanulási munkát bevonnánk, az dominálná az eredményeket, és azt a figyelemre nem méltó megállapítást eredményezné, hogy rengeteg aktív mélytanulási munka van. A kifejezetten az AGI létrehozására irányuló projekteket azonosító mélytanulási projektek száma sokkal kisebb, így kényelmesen illeszkednek a jelen tanulmány módszertanához, és több figyelemre méltó betekintést nyújtanak.

4.2 Az AGI K+F projektek leírása

Minden egyes azonosított AGI K+F projektről általános leírás készült, valamint a következő jellemzők szerinti besorolás:

- *Az intézmény típusa:* A projekt alapjául szolgáló intézmény típusa, például egyetemi vagy kormányzati intézmény.
- *Nyílt forráskódú:* Nyílt forráskód: A projekt nyíltan elérhetővé teszi-e a forráskódját.
- *Katonai kapcsolatok:* A projektnek van-e kapcsolata bármilyen katonai tevékenységgel.
- *Állampolgárság:* Az a nemzet, ahol a projekt székhelye található. A multinacionális projektek esetében egy vezető nemzetet határoztak meg, amelyet a projekt adminisztratív és/vagy operatív vezetésének székhelyeként határoztak meg, a további partnerországokat pedig külön táblázatban tüntették fel.
- *Kijelölt cél:* A projekt AGI-jával kapcsolatban megfogalmazott célok, amelyek azt jelentik, hogy a projekt mit kíván elérni az AGI-jával és/vagy milyen célok elérésére kívánja programozni az AGI-t.
- *A biztonsággal kapcsolatos elkötelezettség:* A projekt elkötelezettségének mértéke az AGI biztonsági kérdésekben.
- *Méret:* A projekt teljes mérete.

4.2.1 Intézmény típusa

Az intézmény típusa attribútum hat kategóriát tartalmaz:

- Akadémiai: Középfokú oktatást folytató intézmény (pl. főiskolák és egyetemek).
- Kormány: Az intézmény egy helyi vagy nemzeti kormányzaton belül található (pl. nemzeti laboratóriumok). Ez a kategória nem tartalmazza az állami főiskolákat és egyetemeket.
- Nonprofit: Az intézmény formálisan nonprofit jellegű, és nem tudományos intézmény (pl. nonprofit kutatóintézetek).
- Magánvállalkozás: Az intézmény nyereségorientált, és nem bocsát ki nyilvános részvényeket.

- Közvéallalat: Az intézmény nyereséorientált, és nyilvános részvényeket bocsát ki.
- Nincs: A projekt nem tartozik semmilyen hivatalos intézményhez.

Néhány projektnek két intézménytípusa volt, de egynek sem volt kettőnél több. A kéttípusú projektek esetében mindkét típust feljegyezték. A projekt résztvevőit csak akkor vettük számításba, ha a projekt honlapjain vagy más kulcsfontosságú projektdokumentumokban hivatalosan is szerepelnek. Néhány projektben több intézmény korlátozott mértékben vett részt, például a publikációk társszerzőjeként. Ezt a korlátozottabb részvételt nem vettük figyelembe, mivel az a sok azonosított projekt rendkívül együttműködő jellege miatt az egész feladatot nehézkessé tenné. Ezt a kódolási politikát nem csak az intézménytípusra, hanem az összes attribútumra vonatkozóan is fenntartották.

4.2.2 Nyílt forráskódú

A nyílt forráskódú attribútumnak három kategóriája van:

- Igen: A projekt forráskódja online letölthető.
- Korlátozott: A projekt kérésre felajánlja a forráskódot.
- Nem: A projekt nem kínál forráskódot.

A projekteket akkor jelölték igennek, ha az AGI-munkájukhoz kapcsolódó forráskódok egy része nyitott. Azokat a projekteket, amelyeknek az AGI kódjának egy része, de nem az összes kódja nyitott, igennek kódolták. Azokat a projekteket, amelyeknek csak más, nem AGI kódja van nyitva, nemmel kódolták. Az, hogy pontosan melyik kód kapcsolódik az AGI-hez, értelmezés kérdése, és a különböző kódolók némileg eltérő eredményeket adhatnak. A három nyílt forráskódú kategória kölcsönösen kizárja egymást: minden projektet az egyik kategóriába kódolnak.

4.2.3 Katonai kapcsolatok

A katonai kapcsolatok attribútumnak három kategóriája van:

- Igen: a projektnek azonosítható katonai kapcsolatai vannak.
- Nem: a projektnek nincsenek katonai kapcsolatai.
- Meghatározatlan: Nem lehetett megállapítani a katonai kapcsolatok meglétét.

A katonai kapcsolatokat a projekt honlapjain és az interneten végzett kulcsszavas keresések, valamint a közelmúltban megjelent publikációk köszönetnyilvánítási szakaszai alapján azonosították. Katonai kapcsolatokkal rendelkezőnek kódolták azokat a projekteket, amelyek katonai szervezetben működnek, amelyek katonai finanszírozásban részesülnek, vagy amelyek egyéb katonai együttműködések folytatnak. A projekteket úgy kódoltuk, hogy nincsenek katonai kapcsolatok, ha kijelentették, hogy nem működnek együtt katonai szervezetekkel, vagy ha az entre projektet át lehetett vizsgálni kapcsolatok szempontjából. Ez utóbbi csak bizonyos kisebb projektek esetében volt megvalósítható. Hacsak nem lehetett végleges kódolási ítéletet hozni, a projekteket nem meghatározottként kódolták.

4.2.4 Állampolgárság

A nemzetiségi attribútumnak két kategóriája van:

- Vezető ország: Az az ország, ahol a projekt adminisztratív és/vagy operatív vezetése található.
- Partnerországok: A projekthez hozzájáruló egyéb országok

Minden projekthez egy vezető országot határoztak meg; néhány projektnek több partnerországa volt.

4.2.5 Kijelölt célok

A kitűzött célok attribútum hat kategóriát tartalmaz:

- **Állatjólét:** AGI a nem emberi állatok javát szolgálja.
- **Ökocentrizmus:** Az AGI a természetes ökoszisztémák javát szolgálja.
- **Humanitarizmus:** Az AGI az emberiség egészségének javát szolgálja. Ebbe a kategóriába tartoznak azok a kijelentések, amelyek az AGI-t az általános emberi problémák, például a szegénység és a betegségek megoldására használják.
- **Intellektualizmus:** Az AGI intellektuális célokra épül, ami magában foglalja magának az AGI-nek az intellektuális teljesítményét és az AGI felhasználását intellektuális célok elérésére.
- **Nyereség:** Az AGI-t azért építették, hogy pénzt hozzon az építetőinek.
- **Transzhumanizmus:** Az AGI-t a fejlett biológiai és/vagy számítógéppel rendelkező lények javára építik, beleértve potenciálisan magát az AGI-t is.
- **Meghatározatlan:** A rendelkezésre álló források nem voltak elegendőek a kódolás megítéléséhez.

A korábbi AGI szakirodalomban található célok néhány kategóriája nem jelent meg az adatokban, beleértve a katonai előnyt és az AGI-építők önző hasznát.

A kinyilvánított célok kódolásánál csak az explicit kijelentéseket vettük figyelembe, a környező kontextust nem. Például a profitorientált vállalatoknál a legtöbb AGI K+F projekt nem jelölte meg kifejezetten a nyereséget célként. Ezeket a projekteket nem kódolták a "nyereség" kategóriába, még akkor sem, ha esetleg a nyereség a céljuk.

4.2.6 A biztonsággal kapcsolatos elkötelezettség

A biztonsági attribútummal kapcsolatos elkötelezettség négy kategóriába sorolható:

- **Aktív:** A projektek elkötelezett erőfeszítéseket tesznek az AGI biztonsági kérdéseinek kezelésére.
- **Mérsékelt:** A projektek elismerik az AGI biztonsági problémáit, de nem tesznek célzott erőfeszítéseket azok kezelésére.
- **Elutasító:** A projektek azzal érvelnek, hogy az AGI biztonsági aggályai helytelenek.
- **Meghatározatlan:** A rendelkezésre álló források nem voltak elegendőek a kódolás megítéléséhez.

Minden projektet e kategóriák egyikébe soroltunk. Elvileg egy projekt lehet aktív és elutasító is, azaz aktívan dolgozhat az AGI biztonságán, miközben elutasítja az ezzel kapcsolatos aggályokat, bár nem találtunk olyan projektet, amelyik ezt tette volna.

4.2.7 Méret

Végül a projektméret attribútum öt kategóriát tartalmaz:

- Kis
- Közepes-kicsi
- Közepes
- Közepes-nagy
- Nagy
- **Meghatározatlan:** A rendelkezésre álló információ nem volt elegendő a kódolás megítéléséhez.

Ez egyszerűen egy ötfokozatú skála a kódolási méretre. Hivatalos méretdefiníciókat nem használunk, mivel a projektek különböző módon mutatják a méretet, például a személyzet, a

publikációk vagy az AGI teljesítményének felsorolásával.

5. Main Eredmények

Ez a szakasz a felmérés főbb eredményeit mutatja be. A teljes eredmények az 1-2. függelékben található. Az összefoglaló ES1. ábrája áttekintést nyújt.

5.1 Az azonosított AGI K+F projektek

45 AGI K+F projektet azonosítottak. Betűrendben a következők:

1. ACT-R, John Anderson (Carnegie Mellon Egyetem) vezetésével
2. Az AERA, Kristinn Thórisson (Reykjavíki Egyetem) vezetésével
3. Az AIDEUS, amelyet Alekszej Potapov az ITMO Egyetemről és Szergej Rodionov az Aix Marseille-i Egyetemről vezet.
4. AIXI, Marcus Hutter vezetésével az Ausztrál Nemzeti Egyetemről
5. Alice Csodaországban, Claes Strannegård (Chalmers University of Technology) vezetésével
6. Animats, egy kis projekt, amelyet nemrégiben svéd, svájci és amerikai kutatók kezdeményeztek.
7. Baidu Research, a Baidu egy AI kutatócsoportja a Baidu-n belül
8. Becca, egy nyílt forráskódú projekt, amelyet Brandon Rohrer vezet.
9. Blue Brain, az École Polytechnique Fédérale de Lausanne Henry Markram vezetésével
10. A Kínai Tudományos Akadémia Mu-Ming Poo által vezetett China Brain Project (Kínai Agykutatási Projekt)
11. CLARION, a Rensselaer Polytechnic Institute Ron Sun vezetésével
12. CogPrime, egy nyílt forráskódú projekt, amelyet Ben Goertzel vezet, székhelye az Egyesült Államokban van, és amelynek laboratóriumai Hongkongban és Addisz-Abeában működnek.
13. CommAI, a Facebook AI Research projektje, amelynek székhelye New York Cityben található, irodái pedig a kaliforniai Menlo Parkban és Párizsban vannak.
14. A Cyc, a Cycorp of Austin, Texas projektje, amelyet Doug Lenat 1984-ben indított el.
15. DeepMind, egy londoni székhelyű AI vállalat, amelyet a Google 2014-ben 400 millió fontért (650 millió dollárért) vásárolt fel.
16. A DeSTIN, amelyet Itamar Arel, a Tennessee-i Egyetem munkatársa vezet.
17. DSO-CA, amelyet Gee Wah Ng vezet a DSO National Laboratories-tól, amely Szingapúr elsődleges nemzeti védelmi kutatási ügynöksége.
18. FLOWERS, Pierre-Yves Oudeyer (Inria) és David Filliat (Ensta ParisTech) vezetésével
19. A GoodAI, egy prágai székhelyű AI vállalat, amelyet Marek Rosa számítógépes játékokkal foglalkozó vállalkozó vezet.
20. A HTM, a kaliforniai Redwood Cityben működő Numenta AI vállalat projektje, amelyet Jeffrey Hawkins, a Palm Computing alapítója vezet.
21. Human Brain Project, egy európai kutatóintézetekből álló konzorcium, amelyet az Európai Bizottság 1 milliárd dollárral támogat.
22. Icarus, Pat Langley vezetésével a Stanford Egyetemről
23. Leabra, a Colorado Egyetem Randall O'Reillyja vezetésével
24. LIDA, Stan Franklin vezetésével a Memphis Egyetemen
25. A Maluuba, egy montreali székhelyű vállalat, amelyet nemrég vásárolt fel a Microsoft
26. MicroPsi, Joscha Bach (Harvard Egyetem) vezetésével
27. A Microsoft Research AI, a Microsoft egyik csoportja 2017 júliusában jelentette be, hogy
28. MLECOG, Janusz Starzyk (Ohio Egyetem) vezetésével
29. NARS, Pei Wang, a Temple Egyetem munkatársa vezetésével
30. Nigel, a Kimera, egy Oregon állambeli Portlandben működő mesterséges intelligenciával foglalkozó vállalat projektje.

31. NNAISENSE, a svájci Luganóban működő, Jürgen Schmidhuber által vezetett mesterséges intelligenciával foglalkozó vállalat.
32. Az OpenAI, egy San Franciscó-i székhelyű nonprofit AI-kutató szervezet, amelyet több neves technológiai befektető alapított, akik 1 milliárd dollárt ajánlottak fel...

33. Real AI, egy hongkongi székhelyű, Jonathan Yan által vezetett AI vállalat.
34. A Kínai Tudományos Akadémia projektje, az Agy által inspirált intelligencia kutatóközpontja (RCBII).
35. Sigma, amelyet Paul Rosenbloom, a Dél-Kaliforniai Egyetem munkatársa vezetett.
36. SiMA, a Bécsi Műszaki Egyetem Dietmar Dietrich vezetésével
37. SingularityNET, a Ben Goertzel által vezetett nyílt AI platform
38. A Stuart Shapiro által vezetett SNePS a New York-i Állami Egyetemen Buffalóban
39. A Soar, amelyet John Laird, a Michigani Egyetem munkatársa és a SoarTech nevű spin-off cég vezet.
40. Susaro, egy mesterséges intelligenciával foglalkozó vállalat, amelynek székhelye Cambridge-ben, az Egyesült Királyságban található, és amelyet Richard Loosemore vezet.
41. Tencent AI Lab, a Tencent AI csoportja
42. Uber AI Labs, az Uber AI kutatási részlege
43. Vicarious, egy San Francisco-i székhelyű AI vállalat
44. Victor, a 2AI projektje, amely a Cifer Inc. leányvállalata, egy kis amerikai vállalat.
45. Whole Brain Architecture Initiative (WBAI), egy tokiói nonprofit szervezet.

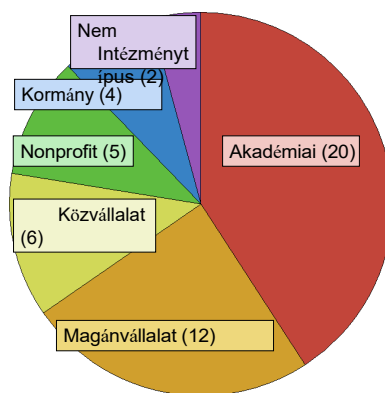
Számos projekt kapcsolódik egymáshoz. Például az AIDEUS vezetője, Alexey Potapov a SingularityNET tanácsadója; az Animats közreműködői között van Claes Strannegård (Alice Csodaországban), Joscha Bach (MicroPsi) és Bas Steunebrink (NNAISENSE); a DeSTIN egyes részeit a CogPrime is felhasználta; a DeepMind és az OpenAI együttműködik az AGI biztonsági kutatásában; a CogPrime-ot és a SingularityNET-et ugyanaz a személy vezeti (Ben Goertzel), és a CogPrime technológiáját a SingularityNET is használja; A Blue Brain és a Human Brain Project ugyanaz a személy (Henry Markram) kezdeményezte, és közös a kutatási stratégiájuk; a Maluuba és a Microsoft Research AI anyaszervezete ugyanaz (Microsoft), akárcsak a China Brain Project és az RCBII (a Kínai Tudományos Akadémia); az ACT-R és a Leabra pedig egy SAL nevű projektben kapcsolódott össze (az ACT-R és a Leabra szintézisének rövidítése; lásd a 2. függelékét). Ez egy olyan AGI-közösségre utal, amely legalább részben közös célokért dolgozik, és nem egymással verseng, ahogyan azt a szakirodalomban gyakran feltételezik (2.2. szakasz).

5.2 Intézmény típusa

20 projekt legalább részben akadémiai intézményekben, 12 legalább részben magánvállalatokban, 6 állami vállalatokban, 5 legalább részben nonprofit szervezetekben, 4 legalább részben kormányzati szervezetekben, kettőnek (AIDEUS, Becca) pedig nincs hivatalos intézményi székhelye. 4 projekt két különböző intézménytípus között oszlik meg: Az Animats és a Soar egyetemi és magánvállalkozás, a FLOWERS egyetemi és kormányzati, a DeSTIN pedig egyetemi és nonprofit. Az 1. ábra összefoglalja az intézménytípusra vonatkozó adatokat.

Az AGI-politika szempontjából fontos, hogy az AGI-kutatás és -fejlesztés túlsúlya az akadémiai és a nyereségorientált vállalatoknál van. Az akadémiai projektek nagy száma azt sugallja, hogy a Yampolskiy és Fox (2013) által a kutatási felülvizsgáló bizottságokra tett javaslatnak van létjogosultsága. Bár az akadémiai szférában már régóta szabályozzák a kockázatos kutatásokat, ez főként az orvosi és társadalomtudományi kutatásokra vonatkozott, amelyek kockázatot jelenthetnek az emberi és állati kísérleti alanyokra, nem pedig az informatikai és a kapcsolódó területeken végzett kutatásokra. Általánosságban elmondható, hogy az akadémiai kutatási etika és szabályozás inkább a kutatási eljárásra összpontosít, nem pedig a kutatásnak a világra gyakorolt következményeire (Schienke et al. 2009). Vannak kivételek, mint például a közelmúltban a kockázatos gain-of-function biológiai kutatásokra vonatkozó szünet (Lipsitch és Inglesby 2014), de ezek a kivételek bizonyítják a szabályt. Az akadémiai AGI K+F projektek között a tendencia például a Human Brain

Project etikai programjában is megfigyelhető, amely a kutatási eljárásra és nem a kutatási következményekre összpontosít. Az akadémiai AGI kockázatos K+F szabályozására irányuló erőfeszítéseknek le kellene küzdeniük ezt az általánosabb tendenciát.



1. ábra. Az intézménytípusra vonatkozó adatok összefoglalása. Az ábra 45-nél több adatpontot mutat, mivel néhány a projektek több intézménytípussal rendelkeznek.

Az AGI K+F túlsúlya az akadémiai szférában talán várható volt, tekintettel az akadémiai szféra régóta fennálló szerepére az AI és általában a hosszú távú kutatások élvonalában. Az akadémiai AGI-közösség azonban meglehetősen eltérő jellegű, mint az AGI kockázati szakirodalmában elterjedt feltételezés, miszerint az AGI K+F-projektek egymással versengenek az AGI első megalkotásáért. Az akadémiai AGI-közösség ugyanúgy működik, mint bármely más akadémiai közösség: például együtt vesznek részt konferenciákon, és ugyanazokban a folyóiratokban megjelent publikációkban idézik és vitatják meg egymás munkáját. Ugyanez elmondható a nonprofit és kormányzati projektekről is, amelyek nagyrészt tudományos jellegűek. Még néhány vállalati projekt is aktívan részt vesz a tudományos AGI-közösségben vagy a kapcsolódó tudományos közösségekben. Ez tovább erősíti azt az elképzelést, hogy az AGI-közösség legalább részben együttműködésen alapuló, nem pedig versenyen alapuló.

A nyereségorientált vállalatok körében két tendencia figyelhető meg. Az egyik, hogy a vállalatok egy része kvázi akadémiai módon támogatja a hosszú távú AGI K+F-et, és csak korlátozottan törődik a rövid távú nyereséggel, vagy egyáltalán a nyereséggel. A Cyc például 1984-ben indult, és ezzel a történelem legrégebben futó AI-projektjei közé tartozik. A GoodAI, annak ellenére, hogy for-profit formában működik, kifejezetten elutasítja a profitot mint végső célt. Ha a rövid távú profitszerzésre irányuló nyomás csökken, ezek a csoportok nagyobb rugalmassággal folytathatnak hosszú távú K+F-et, beleértve a biztonsági mechanizmusokat is.

A másik tendencia az AGI-projektek rövid távú nyereséget hoznak a vállalatoknak, miközben a hosszú távú AGI-célok elérésén dolgoznak. Ahogy a Vicarious társalapítója, Scott Phoenix fogalmaz, az AGI felé vezető munka során "rengeteg értéket teremtenek a köztes időszakban" (High 2016). Ez a tendencia különösen az állami vállalatokon belül működő AGI-csoportoknál figyelhető meg, amelyek közül több is magánvállalkozásként indult, és az állami vállalatok jó pénzt fizettek a felvásárlásukért (DeepMind/Google, Maluuba/Microsoft és Uber AI Labs, amely korábban Geometric Intelligence volt). Ha a rövid távú profit és a hosszú távú AGI K+F közötti szinergia erősnek bizonyul, akkor ez a mélytanuláshoz hasonlóan robbanásszerű AGI K+F-hez vezethet.⁷ Ez akár az AGI K+F domináns tényezőjévé is válhat; ez elég fontos ahhoz, hogy érdemes legyen megnevezni:

AGI nyereség-K+F szinergia: minden olyan körülmény, amelyben a hosszú távú AGI K+F rövid távú nyereséget eredményez.

Az AGI nyereség és K+F szinergiája fontos ok arra, hogy különbséget tegyünk az állami és a magánvállalatok között. Az állami vállalatok a részvényesek nyomására a rövid távú nyereség maximalizálására kényszerülhetnek, ami arra készíti őket, hogy az AGI K+F-et még akkor is előmozdítsák, ha ez hosszú távú globális kockázatokkal jár. Ezzel szemben a magánvállalkozások

⁷ Hasonlóképpen Baum (2017a) szinergiákat észlel a rövid és hosszú távú mesterséges intelligencia társadalmi hatásai iránt érdeklődő csoportok között, így a rövid távú mesterséges intelligencia iránti növekvő érdeklődés felhasználható a hosszú távú mesterséges intelligencia - beleértve az AGI-t - eredményeinek javítására irányuló politikai és egyéb erőfeszítések előmozdítására.

a vállalatokat jellemzően egy szűkebb tulajdonosi kör, esetleg egyetlen személy irányítja, aki dönthet úgy, hogy a biztonságot és a közérdeket a profit elé helyezi. A magánvállalatok vezetése nem feltétlenül tenne ilyet, de talán több lehetőségük van rá. Érdemes azonban megjegyezni, hogy az AGI K+F-nek otthont adó két állami vállalat, a Facebook és a Google továbbra is az alapítói irányítása alatt áll: Mark Zuckerberg megtartja a szavazati jogok többségét a Facebooknál (Heath 2017), míg Larry Page és Sergey Brin megtartja a szavazati jogok többségét a Google anyavállalatánál, az Alphabetnél (Ingram 2017). Mindaddig, amíg megtartják az irányítást, a Facebook és a Google AGI K+F projektjei elkerülhetik a részvényesek által gyakorolt nyomást, amellyel az állami vállalatok gyakran szembesülnek.

A vállalati K+F más kérdéseket is felvet. A vállalatok több okból is ellenállóbbak lehetnek a szabályozással szemben, mint a tudományos csoportok. Először is, a vállalatok a szabályozást gyakran a profitjukat fenyegető veszélynek tekintik, és ennek megfelelően visszavágnak. Ez még akkor is igaz, ha a szabályozás célja a globális katasztrófa megelőzése, mint például a globális felmelegedés esetében (pl. Oreskes és Conway 2010). Másodszor, amikor a szabályozás kifejezetten a vállalati K+F-et célozza, a vállalatok gyakran kifejezik aggodalmukat, hogy a szabályozás sérti szellemi tulajdonukat és gyengíti versenyelőnyüket. Így például az USA sikeresen ellenállt az ellenőrzési intézkedéseknek a biológiai fegyverekről szóló egyezmény tárgyalásakor, nagyrészt gyógyszeriparának szellemi tulajdonáért való aggodás miatt (Lentzos 2011). Amennyiben az AGI-vállalatok az AGI-t fontosnak tartják üzleti modelljük szempontjából, ellenállhatnak az olyan szabályozásoknak, amelyekről úgy vélik, hogy akadályozhatják őket.

Végül, viszonylag kevés projekt van a kormányzati intézményekben. Csak egy projekt található katonai/védelmi intézményben: DSO-CA. A másik három kormányzati projekt a mesterséges intelligencia és a kognitív tudomány szellemi céljait sorolja fel, az egyik az orvosi alkalmazásokat is felsorolja (5.5. szakasz). Ez azonban alábecsüli az AGI K+F-ben való kormányzati részvétel mértékét. Számos más projekt is részesül kormányzati finanszírozásban, amelyek célja az orvostudomány (pl. Blue Brain, HBP), a gazdaságfejlesztés (pl. Baidu) és a haditechnika (pl. Cyc, Soar) fejlesztése.

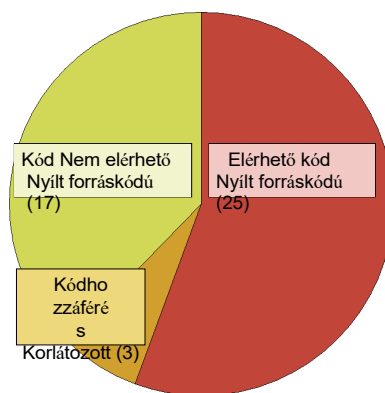
5.3 Nyílt forráskódú

25 projekt forráskódja nyíltan elérhető az interneten. További három projekt (Cyc, HBP és LIDA) kódja kérésre elérhető. E 28 projekt esetében a rendelkezésre álló kód nem feltétlenül a projekt teljes kódgyűjteménye, legalábbis a kód legfrissebb verziója esetében, bár néhány esetben igen. Mindössze 17 olyan projekt volt, amelynek kódja nem volt megtalálható az interneten. E 17 projekt közül 3 esetben (Baidu Research, Tencent AI Lab, Uber AI Labs) az anyaszerkezet rendelkezik nyílt forráskóddal, de a kód átvizsgálása során nem találtunk AGI-kódot. A 2. ábra összefoglalja a nyílt forráskódú adatokat.

A nyílt forráskódú projektek túlsúlya hasonlít a kódlói közösség nagy részében tapasztalható, a nyitottság irányába mutató szélesebb körű tendenciára. A projektek közül sokan a github.com-ra, egy népszerű kódtárba teszik fel a kódjukat. Még a vállalati projektek is, amelyeknek versenyelőnyük lehet, gyakran elérhetővé teszik kódjuk legalább egy részét.

Goertzel (2017a) kétféle nyílt forráskódú projektet különböztet meg: A "klasszikus" nyílt forráskódú projektek, amelyekben a kódfejlesztés "nyíltan" történik, és a "vállalati" nyílt forráskódú projektek, amelyekben a kódfejlesztést a projekt bennfentesei végzik zárt környezetben, majd nyíltan kiadják.

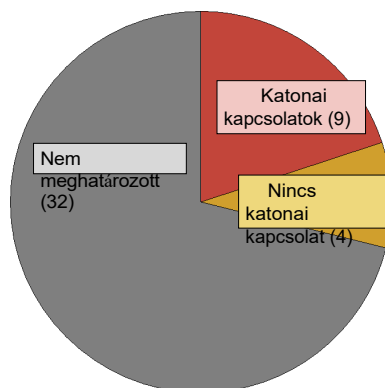
Goertzel (2017a) az OpenAI-t említi a vállalati nyílt forráskód példaként (megjegyzés: az OpenAI egy nonprofit projekt, nem egy profitorientált intézménynél); az ő CogPrime-ja a klasszikus nyílt forráskód példája lenne. A klasszikus/vállalati megkülönböztetés az etika, a kockázat és a politika szempontjából is fontos lehet, mivel befolyásolja, hogy ki képes befolyásolni az AGI céljait és biztonsági jellemzőit. Az azonban, hogy mely nyílt forráskódú projektek minősülnek klasszikus vagy vállalati projekteknek, meghaladja e dokumentum kereteit.



2. ábra. A nyílt forráskódú adatok összefoglalása.

5.4 Katonai kapcsolatok

Kilenc projektnek van azonosítható katonai kapcsolata. Ezek között van egy katonai/védelmi intézményben működő projekt (DSO-CA, a DSO National Laboratories-ban, Szingapúr elsődleges nemzeti védelmi kutatási ügynökségében) és nyolc olyan projekt, amely katonai finanszírozásban részesül (ACT-R, CLARION, Icarus, Leabra, LIDA, Sigma, SNePS, Soar). E nyolc projekt mindegyike legalább részben amerikai egyetemi intézményekben működik. Ez követi azt a régóta tartó tendenciát, hogy az Egyesült Államokban a számítástechnikai kutatások katonai finanszírozása a jellemző. A Soar szintén a SoarTech magáncégen alapul, amely a weboldalán erőteljesen reklámozza a katonai alkalmazásokat. A 3. ábra a katonai kapcsolatok adatait foglalja össze.



3. ábra. A katonai kapcsolatok adatainak összefoglalása.

Mindössze négy projektről állapították meg, hogy nem rendelkezik katonai kapcsolatokkal. Ezek közül kettő (Aera, HBP) nyíltan elutasítja a katonai kapcsolatokat; a másik kettő (Alice in Wonderland, Animats) elég kicsi és jól dokumentált volt ahhoz, hogy a katonai kapcsolatok hiányát meg lehessen állapítani. A többi 32 projektet nem meghatározottként kódolták. E projektek közül soknak valószínűleg azért nincsenek katonai kapcsolatai, mert nem katonai alkalmazásokon dolgoznak, és nem olyan intézményekben (például amerikai egyetemeken) működnek, amelyek általában katonai kapcsolatokkal rendelkeznek. Néhány projekt esetében valószínűbb, hogy katonai kapcsolatokkal rendelkeznek. A Microsoftnak például kiterjedt katonai ügyekkel foglalkozó programja van, amely lehet, hogy kapcsolódik (vagy nem) az MSR AI-hez.

A nyilvánosan elérhető adatok azt sugallják, hogy a szingapúri és az amerikai hadsereg érdeklődése az AGI iránt katonai mércével mérve hétköznapi, és nem a megkérdőjelezhetetlen katonai dominanciára való törekvés irányába mutat. Úgy tűnik, hogy a DSO-CA egy kis, nagyrészt tudományos jellegű projekt; egy nemrégiben megjelent tanulmány a DSO-CA-t képfeliratozásra

alkalmazva mutatja be, egy étkező családot ábrázoló fénykép példáján (Ng et al. 2017). A projekt nem tűnik egy nagy szingapúri hatalmi játszma benyomását keltőnek. Hasonlóképpen, a katonai finanszírozású amerikai projektek is nagyrészt tudományos jellegűek; a legtöbbjük esetében,

találhatók, amelyek mindegyike mesterséges intelligencia központnak számít; az egyetlen kivétel az amerikai Virgin-szigeteken található Victor. Emellett az OpenAI (nonprofit) és a Becca (intézményi otthonnal nem rendelkező) San Francisco, illetve Boston területén található, az ottani mesterséges intelligencia ipar miatt.

13 projekt multinacionális. Az AERA, az AIDEUS, a Baidu Research, a CommAI, a Maluuba, a Tencent AI Lab és az Uber AI Labs összesen két országban található (beleértve a vezető országot is). Az Animats, a CogPrime, a DeepMind és a SOAR egyenként három országban működik. A Blue Brain öt országban van jelen.

A SingularityNET 8 országban működik. A Human Brain Project 19 országban van jelen, amelyek mindegyike Európában található, kivéve Izraelt és (attól függően, hogy ki hogyan sorolja be) Törökországot. A leggyakoribb nemzetközi együttműködés az Egyesült Királyság és az Egyesült Államok között zajlik, mindkét ország négy projektben vesz részt (Blue Brain, DeepMind, Soar és Uber AI Labs). Kína és az USA egyaránt négy projektben vesz részt (Baidu Research, CogPrime, SingularityNET és Tencent AI Lab).

Geopolitikai szempontból az Egyesült Államok és szövetségesei jelentős dominanciát élveznek. Csak öt olyan ország rendelkezik AGI K+F-projektekkel, amely nem az USA szövetségese: Brazília, Kína, Etiópia, Oroszország és (vitathatóan) Svájc, amelyek közül csak Kína és Oroszország tekinthető az USA ellenfelének. A nyolc projekt közül, amelyben Kína és Oroszország részt vesz, egy az Egyesült Államokban működik (CogPrime), három másikban amerikai részvétel van (Baidu Research, SingularityNET és Tencent AI Lab), kettő pedig olyan kis projekt, amely szoros kapcsolatban áll a nyugati AI-közösségekkel (AIDEUS, Real AI). A fennmaradó kettő a Kínai Tudományos Akadémia projektje, amely az alapvető idegtudományokra és az orvostudományra összpontosít (China Brain Project, RCBI). A projektek USA-ban és szövetségeseiben való koncentrációja, valamint a többi projekt nyugati és tudományos irányultága jelentősen megkönnyítheti az AGI K+F nemzetközi irányítását.

Geopolitikai szempontból is figyelemre méltó, hogy néhány geopolitikailag fontos régió hiányzik. Dél-Ázsiában nincsenek projektek, és csak egy (CogPrime) részben Afrikában és egy (SingularityNET) részben Latin-Amerikában. Az oroszországi AGI K+F főleg Alekszej Potapov hozzájárulásából áll, aki az AIDEUS és a SingularityNET projekteken is közreműködik. Emellett Potapov publikált AGI-kutatásai főként elméleti, nem pedig K+F jellegűek (pl. Potapov et al. 2016).⁸ Latin-Amerikában az egyetlen AGI K+F Cassio Pennachin hozzájárulása a SingularityNET-hez. Az ebben a felmérésben azonosított AGI K+F projektek tehát szinte mind Észak-Amerikában, Európában, a Közel-Keleten, Kelet-Ázsiában és Ausztráliában található intézményekben folynak.

Fontos fenntartás, hogy az ACT-R esetében nem vonták be a partnerintézményeket. Az ACT-R honlapja számos együttműködő intézményt sorol fel, amelyek szinte mindegyike akadémiai intézmény, és amelyek 21 országban találhatóak. Ezen országok közül több nem vesz részt más AGI K+F projekteken: India, Irán, Marokkó, Pakisztán, Dél-Korea, Srí Lanka és Venezuela. Ezeket a partnerintézményeket azért nem vesszük figyelembe, mert az ACT-R honlapjának ez a része elavult. Lehetséges, hogy ezekben az országokban jelenleg is van valamilyen hozzájárulás az ACT-R-hez; hogy van-e ilyen vagy sem, az meghaladja e dokumentum kereteit.

Egy másik fontos figyelmeztetés a sok nyílt forráskódú projektből adódik. Ez a kód lehetővé teszi, hogy az AGI K+F-et a világ bármely pontján lehessen végezni. Így lehetséges, hogy más országok is részt vesznek az AGI K+F-ben, talán számos más ország is. Az olyan országok azonosítása, amelyek részvétele kizárólag nyílt forráskódú kódhoz való hozzájárulásból áll, meghaladja e dokumentum kereteit.

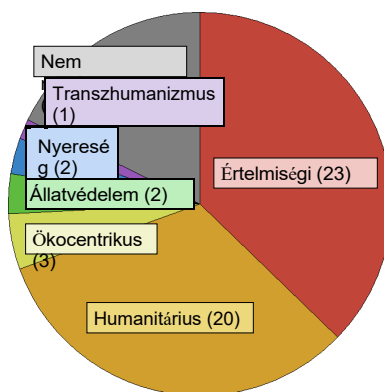
5.6 Meghatározott cél

A kinyilvánított célok között a humanitárius és az intellektualizmus túlsúlya az uralkodó tendencia. 23 projekt intellektualista és 20 humanitárius célokat fogalmazott meg, míg csak 3 projekt fogalmazott meg ökocentrikus célokat (SingularityNET, Susaro, Victor), 2 projekt profítcélokat (Maluuba, NNAISENSE), 2 projekt állatjóléti célokat (Human Brain Project, amely agyszimulációt javasol az állatkísérletek elkerülésére, és a SingularityNET, amely minden érző lény javára kíván tenni), és 1

projekt transzhumanista célokat (SingularityNET, amely az érző lények és robotok javára kíván tenni).
12 projekt céljai nem voltak meghatározva.

⁸ Szintén Oroszországban van Max Talanov a Kazan Állami Egyetemről, aki hozzájárult a NARS-hoz, de nem eléggé ahhoz, hogy partnerként kódolják (pl. Wang et al. 2016).

Néhány projekt több célt is megfogalmazott: Az 5. ábra összefoglalja a kitűzött célok adatait.



5. ábra. A kitűzött célok adatainak összefoglalása. Az ábra 45-nél több adatpontot mutat, mivel néhány projektnek több kitűzött célja is van.

Az intellektualista projektek túlnyomórészt akadémiaiak, összhangban a tudományos életben a szellemi érdemek szélesebb körű hangsúlyozásával. A 19 projektből 14 legalább részben valamely felsőoktatási intézményben zajlik. (A másik öt akadémiai projekt mindegyike kódolva van, mint meghatározatlan; ezek általában kisebb projektek, és valószínűleg intellektualista célokat szolgálnak.) Az intellektualista projektek közé tartozik a Kínai Tudományos Akadémián folyó két projekt is. Az Akadémia kormányzati intézményként van kódolva, de erősen tudományos jellegű. A többi intellektualista projekt mind nyereségorientált vállalatoknál van, kivéve az AIDEUS-t, amely nem rendelkezik intézménytípussal. E vállalati projektek közül négynek (GoodAI, HTM, Maluuba és NNAISENSE), valamint az AIDEUS-nak is szoros kapcsolata van az akadémiai szférával. Az ötödik, az MSR AI egy kicsit kilóg a sorból.

Mind a nyolc katonai kapcsolatokkal rendelkező amerikai akadémiai projekt értelmiségi célokat tűzött ki. Közülük csak egy (LIDA) fogalmaz meg humanitárius célokat is. (Az egyetlen katonai kapcsolatokkal rendelkező, nem amerikai projekt, a DSO- CA céljai nem voltak meghatározva.) Csak egy másik amerikai akadémiai projekt (NARS) fogalmaz meg humanitárius célokat; a többi nem részletezett. Eközben a nyolc nem amerikai tudományos projektből három humanitárius célokat fogalmaz meg. Ennek egyik lehetséges magyarázata, hogy az amerikai akadémiai AGI K+F katonai finanszírozásának túlsúlya arra készíti a projekteket, hogy a humanitárius célok helyett a szellemi célokat hangsúlyozzák, míg a más országokban (különösen Európában) más AGI-alkalmazások, különösen az egészségügy számára rendelkezésre álló finanszírozás a humanitárius célok hangsúlyosabbá tételére ösztönöz. Röviden, az adatok tükrözhetik az USA nagy katonai költségvetését és a nagy európai polgári költségvetést, ugyanakkor azt is jelzik, hogy az AGI-kutatóknak nehézséget okoz a katonai K+F humanitárius szempontok szerinti megfogalmazása.

A 18 for-profit projektből 10 humanitárius célokat jelöl meg. Ezek közül sokan nyugati (különösen amerikai) projektek, amelyeknek erős a nyilvánosság előtti megjelenése (pl. CommAI/Facebook, DeepMind/Google, MSR AI/Microsoft, Nigel/Kimera, Uber AI Labs). A többi humanitárius for-profit projekt némelyike az altruizmus nyugati felfogásában gyökerezik (pl. GoodAI, Real AI). Ezzel szemben a nem humanitárius for-profit projektek többnyire kisebb projektek (pl. Animats, NNAISENSE) és kínai projektek (Baidu Research, Tencent AI Lab). Ez arra utal, hogy a for-profit humanitárius tevékenység elsősorban a nyugati értékek és a nyugati marketing keveréke.

Az értékek és a marketing közötti különbségtétel fontos, és rámutat a felmérés módszertanának egyik korlátjára. A projektek nyilvánosan olyan célokat fogalmazhatnak meg, amelyek vonzóak a célközönségük számára, miközben a magánéletben más célokat tartanak szem előtt. Ez magyarázatot

adhat arra, hogy tíz for-profit projekt miért fogalmaz meg humanitárius célokat, míg csak kettő profitcélokat. Néhány profitorientált projektet valóban kevésbé érdekelhet a profit - kettő közülük, a GoodAI és a Vicarious kifejezetten elutasítja a profitot mint célt. Mások azonban talán csak azért fogalmazzák meg a humanitáriusságot, hogy jó színben tüntessék fel magukat. Ez a gyakorlat analóg lenne a

a "zöldmosás" gyakorlata, amelynek során a környezetkárosító vállalatok kisebb, környezetvédelmi szempontból előnyös tevékenységeket reklámoznak, hogy azt a benyomást keltsék, hogy "zöldebbek", mint amilyenek valójában (pl. Marquis et al. 2016). Egy szénipari vállalat például népszerűsítheti a dolgozók irodán belüli újrahasonosítását, hogy ezzel is bizonyítsa környezetvédelmi elkötelezettségét. Hasonlóképpen, a vállalati AGI K+F projektek hirdethetik a humanitárius aggodalmat, miközben elsősorban a profitra törekszenek, függetlenül az általános humanitárius következményektől. Az ilyen magatartást "bluwashing"-nek nevezhetjük, a kék az ENSZ és a Humanitárius Világtalálkozó színe.

A kitűzött célok adataiból kimutathatóan hiányzik a katonai előny és az AGI-építők önző haszna. Mindkettő AGI K+F célnak számít a korábbi szakirodalomban, akárcsak a transzhumanizmus/kozmisztika, amely csak egy projektben (SingularityNET) kap rövid támogatást. E hiányok oka meghaladja a dolgozat kereteit, de néhány lehetőség hihető.

A transzhumanizmus és a kozmisztika nem széles körben elfogadott célok a mai emberi társadalomban, bár viszonylag gyakoriak az AGI fejlesztők körében (pl. Goertzel 2010). Valószínű, hogy a transzhumanisták és a kozmisták (a SingularityNET-en kívül) inkább titokban tartják nézeteiket, hogy elkerüljék az AGI K+F projektjeikkel kapcsolatos aggodalmakat, amelyek az emberiséget fenyegethetik. Az AGI katonai előnyökre való törekvése hasonló aggodalmakat vethet fel, és az ellenfeleket is arra készítheti, hogy saját AGI-kutatásukat megkezdjék vagy felgyorsítsák. Végül, az AGI önös érdekből történő kutatása antiszociális, és ha nyíltan kimondják, az hírnevet veszélyeztetheti, vagy szabályozásra készíthet. Ugyanakkor az is lehetséges, hogy a jelenleg az AGI K+F-hez vonzó embereknek valójában elsősorban humanitárius és intellektualista céljaik vannak, és nem ezek a más célok (Goertzel figyelemre méltó kivétel).

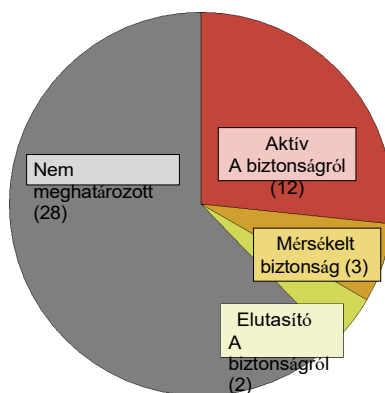
Fontos kérdés, hogy az AGI K+F-projektek milyen mértékben osztják meg ugyanazokat a célokat. Az eltérő célokkal rendelkező projektek nagyobb valószínűséggel versenyezhetnek egymással az AGI első megépítéséért (Armstrong et al. 2016). Az AGI K+F projektek között a humanitárius és az intellektualizmus túlsúlya arra utal, hogy a célok tekintetében jelentős mértékű konszenzus van. Továbbá ezek a célok ágenssemlegesek, ami tovább sugallja a versengésre való alacsony tendenciát. A verseny azonban ettől függetlenül előfordulhat. Ennek egyik oka, hogy e konkrét nézetek részleteiben jelentős nézeteltérések lehetnek, például az emberi jogok eltérő felfogásában Kína és a Nyugat között (Posner 2014). Emellett még a lelkiismeretes emberek is érezhetnek késztetést arra, hogy ők legyenek azok, akik elsőként építik meg az AGI-t, és talán azt gondolják magukban: "Felejtsd el, mi a helyes és mi a helytelen. Tigris van a farkadnál fogva, és soha nem lesz annyi esélyed, hogy befolyásold az eseményeket. Fuss vele!"⁹ És persze lehetnek nézeteltérések az AGI-humanisták és az értelmiségiek között, csakúgy, mint a többi megfogalmazott céllal.

Végül meg kell jegyezni, hogy a kinyilvánított célok kódolása különösen értelmező jellegű volt. Sok projekt nem fogalmazza meg céljait feltűnően vagy filozófiailag tiszta kifejezésekkel. A DeepMind például az éghajlatváltozást nevezi meg fontos alkalmazásként. Ez lehet ökocentrikus, humanitárius vagy mindkettő, attól függően, hogy a DeepMind miért akar foglalkozni az éghajlatváltozással. Humanitáriusnak kódoltuk, mert az "emberiség legnagyobb kihívásainak kezelésében való segítségnyújtás" kontextusában említik, de valószínű, hogy az ökocentrizmus volt a cél.

5.7 A biztonsággal kapcsolatos elkötelezettség

A biztonsággal kapcsolatos kötelezettségvállalás csak 17 projekt esetében volt azonosítható. Ezek közül 12 projektet találtunk a biztonsággal kapcsolatban aktívnek (AERA, AIDEUS, CogPrime, DeepMind, FLOWERS, GoodAI, LIDA, NARS, OpenAI, Real AI, Susaro és WBAI), 3 projektet mérsékeltnak (CommAI, Maluuba és Vicarious), 2 projektet pedig elutasítónak (HTM, Victor). A 6. ábra összefoglalja a biztonsági adatokkal kapcsolatos elkötelezettséget.

⁹ Ez az idézet Martin Hellman kriptográfustól származik, aki maga is nyilvánvalóan lelkiismeretes ember. Az idézet egy 1975-ös epizódra utal, amikor Hellman azon vitatkozott, hogy nyilvánosságra hozzon-e egy kód feltörésével kapcsolatos információkat, annak ellenére, hogy az amerikai Nemzetbiztonsági Ügynökség figyelmeztetett, hogy ez káros lenne. Hellman végül mégis nyilvánosságra hozta a dolgot, és utólag arra a következtetésre jutott, hogy helyesen cselekedett, de rossz motivációval tette (Hellman és Hellman 2016, 48-50. o.).



6. ábra. A biztonsági adatokkal kapcsolatos kötelezettségvállalás összefoglalása.

A biztonsági adatokkal kapcsolatos elkötelezettség egyik fő tendenciája a biztonságra való odafigyelés hiánya. A 45 projektből 28 esetében nem volt azonosítható a biztonsággal kapcsolatos kötelezettségvállalás. Ez empirikusan alátámasztja a korábbi AGI-politikai szakirodalomban általánosan elterjedt feltételezést, miszerint az AGI-fejlesztők a biztonságot nem kellő mértékben figyelembe véve kívánnak eljárni (2.3. szakasz). Mivel a felmérés a nyilvánosan elérhető adatokra összpontosított, a biztonság elhanyagolását túlértékelheti, mivel egyes projektek a biztonságra anélkül is figyelmet fordíthatnak, hogy ezt nyilvánosan jeleznék. Például az Animats és az NNAISENSE kódja nem meghatározott, de e projektek bármely aspektusáról kevés nyilvánosan elérhető információ áll rendelkezésre. Eközben Jürgen Schmidhuber, az NNAISENSE vezető tudósa és Bas Steunebrink, az NNAISENSE társalapítója és az Animats közreműködője az AGI biztonságával kapcsolatos munkát végzett (Steunebrink et al. 2016). Az adatok mégis erősen arra utalnak, hogy az AGI K+F projektek körében a biztonságot széles körben elhanyagolják.

A 17 projekt közül, amelyek esetében a biztonsággal kapcsolatos kötelezettségvállalást azonosították, további tendenciák is megfigyelhetők. E 17 projekt között van a 8 tisztán humanitárius célú projektből 6, az egyetlen tisztán ököcentrikus célú projekt, ugyanakkor a 10 tisztán intellektualista célú projektből csak 1, és a 11 nem meghatározott célú projektből 1. A 17 projekt között van a tisztán humanitárius célú projektek közül 6, és az egyetlen tisztán ököcentrikus célú projekt. A 17 projekt között szerepel továbbá a 16 tisztán profitorientált projektből 9, a 4 tisztán nonprofit projektből 3, de a 15 tisztán tudományos intézményben megvalósuló projektből csak 2, és a 4 projektből 1 részben tudományos intézményben megvalósuló projekt. Ez olyan projektek csoportjára utal, amelyek széles körben foglalkoznak az AGI K+F hatásaival, beleértve a hatásokkal kapcsolatos etikai kérdéseket, valamint a kívánt hatások elérésének kockázatával/biztonságával kapcsolatos kérdéseket, és amelyek túlnyomórészt az akadémiai intézményeken kívül helyezkednek el. Eközben van egy olyan projektcsoport is, amely túlnyomórészt tudományos, és az AGI-t nagyrészt intellektuális szempontból vizsgálja, amennyiben egyáltalán megfogalmazznak valamilyen célt. Ezek a tendenciák a kockázati és etikai gyakorlatok megerősítésére irányuló javaslatok fontosságára utalnak, amelyek néha azt sugallják, hogy a megfelelő képzés révén az AGI biztonsága az akadémiai AGI K+F projektek körében ésszerűen kezelhető feladat lehet, például a kutatási felülvizsgálati bizottságokon keresztül (Yampolskiy and Fox 2013).

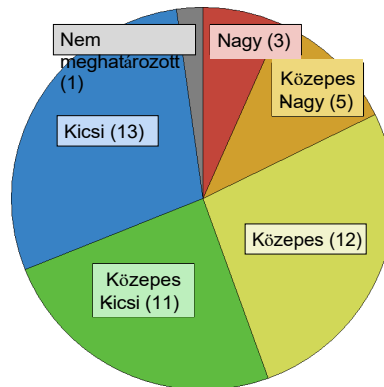
A biztonság terén aktívnak talált 12 projekt között néhány téma nyilvánvaló. Számos projekt hangsúlyozta az AGI-k biztonságos kiképzésének fontosságát (AERA, CogPrime, GoodAI és NARS), és kritizálta azokat az aggodalmakat, amelyek szerint az AGI biztonsága rendkívül nehéz lehet (2.2. szakasz; Goertzel and Pitt 2012; Goertzel 2015; 2016; Bieger al. 2015; Steunebrink et al. 2016).¹⁰ Más projektek a közeljövőbeli mesterséges intelligencia biztonsági kérdéseire összpontosítanak a robotika (FLOWERS; Oudeyer et al. 2011) és a megerősítő tanulás (DeepMind és OpenAI; Christiano et al. 2017) összefüggésében, ez utóbbi összhangban van a Google és az OpenAI kutatói által javasolt, a közeljövőbeli mesterséges intelligencia felhasználását a hosszú távú mesterséges intelligencia

biztonságának tanulmányozására irányuló menetrenddel (Amodei et al. 2016). A LIDA az AGI erkölcsiségének alapjait vizsgálta a mesterséges rendszerekkel kapcsolatban (Wallach et al. 2010; 2011; Madl and Franklin 2015). A WBAI azt sugallja, hogy részben azért igyekeznek egyszerű AGI-t építeni, mert az emberi intelligenciához való hasonlósága biztonságosabbá tenné azt.

¹⁰ Hasonlóképpen, a Vicarious "biztonsági szempontból mérsékelt" projekt azt a meggyőződést fejezi ki, hogy az AGI biztonsága nem lehet nehéz, ami arra készíti őket, hogy (még) ne legyenek aktívak a biztonság terén.

5.8 Méret

A projektek mérete főként a kis és közepes méretűek közé tartozott. 13 projektet kis méretűnek, 11-et közepes méretűnek, 12-t közepes méretűnek, 5-öt közepes méretűnek (Blue Brain, CogPrime, MSR AI, Soar és Vicarious), 3-at pedig nagy méretűnek (DeepMind, Human Brain Project és OpenAI) kódoltak, 1 projektet pedig nem határoztak meg (Susaro, egy "lopakodó üzemmódban" lévő projekt, amelyről nem találtak méretadatot). A 7. ábra összefoglalja a méretadatokat.



7. ábra. A méretadatok összefoglalása.

Bár az egyes méretkategóriák projektjei eltérőek, néhány tendencia mégis nyilvánvaló. Az intézménytípus tekintetében a felsőoktatási projektek általában valamivel kisebbek, míg a vállalati projektek általában valamivel nagyobbak, bár vannak nagy felsőoktatási és kis vállalati projektek is. A 13 kis projektből 7 egyetemi, míg csak három magánvállalati, és egy sem állami vállalatnál. A 8 közepesen nagy vagy nagy projektből 3 egyetemi, míg 4 vállalati.

A méret és az intézménytípus közötti tendenciák részben kódolási hibák lehetnek, mivel az egyetemi és a vállalati projektek méretének mutatói eltérőek. A vállalati projektek kisebb valószínűséggel adják meg teljes személyi állományukat, illetve kevésbé katalogizálják termelékenységüket a nyílt publikációs listákon. A vállalati projektek jelentős részei a rövid távú alkalmazásokra összpontosíthatnak a hosszú távú AGI K+F rovására, bár az akadémiai projektek hasonlóan intellektuális célokra összpontosíthatnak az AGI K+F rovására. A vállalati projektek, különösen az állami vállalatok projektjei gyakran rendelkeznek további erőforrásokkal, amelyeket szétoszthatnak az AGI-projektek között, beleértve a nagy adathalmazokat és a finanszírozást. Az akadémiai intézmények is találnak finanszírozást AGI-projektekre, de általában nem olyan mértékben, mint a vállalatok, különösen a nagy állami vállalatok. Az akadémiai és vállalati projektek közötti különbséget jól szemlélteti például az NNAISENSE, amelyet részben azért indítottak el akadémikusok, hogy megakadályozzák, hogy más vállalatok elorozzák a kutatási tehetségeket. Ez azt is sugallja, hogy a legnagyobb AGI K+F projektek egyre inkább vállalatiak lehetnek, különösen, ha van AGI K+F-profit szinergia (5.2. szakasz).

Ami a nyílt forráskódot illeti, egyértelmű tendencia, hogy a nagyobb projektek egyre inkább nyílt forráskódúak. A 17 olyan projekt közül, amely nem teszi elérhetővé a forráskódot, 8 kis és 6 közepes méretű, míg csak 1 közepes és 1 közepes méretű, 1 pedig meghatározatlan méretű. Ezzel szemben a 24 nyílt forráskódú, korlátozás nélküli projektből 11-ből 5 kis projekt, 11-ből 5 közepes méretű projekt, 9-ből 9 közepes projekt, 5 közepes-nagy projektből 4, és 3 nagy projektből 2.

Ami a katonai kapcsolatokat illeti, a katonai kapcsolatokkal rendelkező projektek általában a kis és közepes méretűek közé tartoznak. A 8 katonai projekt között 2 kis, 2 közepes-kis, 4 közepes és 1 közepes-nagy projekt található. Ennek van értelme, mivel ezek elsősorban katonai finanszírozású amerikai akadémiai projektek.

Az akadémiai projektek általában kisebbek, míg a külső finanszírozásúak inkább közepes méretűek.

méret. A katonai finanszírozásról lemondó tudományos projektek néha kisebbek lehetnek, mint amekkorák egyébként lehettek volna.

A nemzetiség tekintetében megfigyelhető némi tendencia, hogy az amerikai székhelyű projektek nagyobbak, míg a kínai székhelyűek kisebbek, bár az USA esetében ez a tendencia kissé gyenge. A 23 amerikai székhelyű projektből 6 a 13 kis projektből, 3 a 11 közepes-kis projektből, 9 a 12 közepes projektből, 4 az 5 közepes-nagy projektből, és 1 a 3 nagy projektből. Ezzel szemben mind a hat kínai székhelyű projekt vagy kicsi, vagy közepesen kicsi. Ez a tendencia megerősíti az 5.5. szakasz azon megállapítását, hogy az AGI az USA-ban és szövetségeseiben koncentrálódik. Fontos fenntartás, hogy a kínai projektek közül kettő a Baidu és a Tencent kínai nagyvállalatokhoz tartozik. Ezek a vállalatok nagy AI-csoportokkal rendelkeznek, amelyek csak kis figyelmet fordítanak az AGI-re. Ha a vállalatok úgy döntenek, hogy többet foglalkoznak az AGI-vel, akkor gyorsan növekedhetnek. Az is lehetséges, hogy már most is többet tesznek, olyan munkában, amelyet ez a felmérés nem azonosított, bár ugyanez elmondható a más országokban működő projektekről is.

A kitűzött célt illetően néhány tendencia figyelhető meg. Először is, a meg nem határozott céllal rendelkező projektek általában kicsik, köztük a 13 kis projektből 7 kis projekt. Ennek van értelme: a kisebb projekteknek kevesebb lehetőségük van céljaik megfogalmazására. Az intellektualista projektek általában közepes-kis és közepes méretűek, beleértve a 11 közepes-kis projektből 6-ot és a 12 közepes projektből 9-et, ami összhangban van azzal, hogy az intellektualista projektek inkább tudományos jellegűek. A humanitárius projektek általában nagyobbak, beleértve a 12 közepes projektből 6-ot, az 5 közepes-nagy projektből 3-at, és mind a 3 nagy projektet. Ennek egyik lehetséges magyarázata, hogy a nagyobb projektek általában szélesebb társadalmi támogatottsággal rendelkeznek, ami arra készteti őket, hogy humanitáriusabb álláspontot képviseljenek. A nagyobb projektek humanitárius jellegének túlsúlya megkönnyítheti a konszenzus kialakulását a célok tekintetében azon projektek között, amelyek a legnagyobb valószínűséggel építik az AGI-t. Ez a konszenzus viszont segíthet elkerülni a kockázatos versenydinamikát (Armstrong et al. 2016).

Végül, ami a biztonsággal kapcsolatos elkötelezettséget illeti, gyenge tendencia figyelhető meg, hogy a nagyobb projektek aktívabbak a biztonság terén. Az aktív projektek között 2 kis, 3 közepes-kis, 3 közepes, 1 közepes-nagy, 2 nagy és 1 meghatározatlan méretű projekt található. Ezzel szemben a biztonsággal kapcsolatban nem részletezett elkötelezettségű projektek között 10 kis, 7 közepes-kis, 6 közepes, 3 közepes-nagy és 1 nagy projekt található. Tehát minden méretű projektet lehet aktívnak vagy nem aktívnak találni a biztonság terén, bár a nagyobb projektek valamivel nagyobb tendenciát mutatnak az aktivitásra.

5.9 Egyéb figyelemre méltó projektek

Az AGI K+F-projektek azonosítása során 47 további figyelemre méltó projektet rögzítettek. Ezek között 26 inaktív AGI K+F projekt, 9 olyan projekt, amely nem AGI, és 12 olyan projekt, amely nem K+F. Az aktív AGI K+F-projektekkel ellentétben az egyéb figyelemre méltó projektek azonosítása során nem törekedtek a teljességre. Valószínű, hogy néhány figyelemre méltó projekt nem szerepel. Az egyéb figyelemre méltó projektek listája és rövid leírása a 2. függelékben található.

A 26 inaktív projekt főként akadémiai projekt, mint például a 4CAPS (Marcel Just, Carnegie Mellon Egyetem), a Mesterséges Agy Laboratórium (Hugo de Garis, Xiamen Egyetem), a CERA- CRANIUM (Raúl Arrabales, Madridi Egyetem), CHREST (vezetője Fernand Gobet a Liverpooli Egyetemről), DUAL (vezetője Boicho Kokinov az Új Bolgár Egyetemről), EPIC (vezetője David Kieras a Michigani Egyetemről), OSCAR (vezetője John Pollock az Arizonai Egyetemről) és Shruti (vezetője Lokendra Shastri a Berkeley-i Kaliforniai Egyetemről). Ezek időtartama jelentősen különbözött, néhány évtől (pl. AGINAO, 2011 és 2013 között aktív) több mint egy évtizedig (pl. CHREST, 1992 és 2012 között aktív).

A kilenc projekt, amely nem AGI, többnyire nagy számítástechnikai vállalatok mesterséges

intelligencia csoportjai. Hat vállalatnál alaposan átkutattuk az AGI-projektekét, de nem találtunk egyetlen sem:

Alibaba, Amazon, Apple, Intel, Salesforce és Twitter.¹¹ Tekintettel ezeknek a vállalatoknak a kiterjedt erőforrásaira, figyelemre méltó, hogy a jelek szerint nincsenek AGI-projektjeik.¹² A DeepMind mellett a Google két másik projektjét is figyelembe vettük: A Google Brain és a Quantum AI Lab. Bár a Google Brain a DeepMinddel együtt végzett némi AGI-munkát, a szűkebb értelemben vett mesterséges intelligenciára összpontosít. Végül egy szűk kognitív architektúra projektet (Xapagy) vontak be illusztratív példaként. Számos további szűk kognitív architektúra projektet is be lehetett volna vonni - Kotscher et al. (2016) 86 kognitív architektúra projektet sorol fel, amelyek többsége szűk.

A K+F-nek nem minősülő 12 projekt az AGI különböző aspektusait öleli fel. Egyesek az alaptudományokra összpontosítanak, köztük számos agyi projekt (pl. a BRAIN kezdeményezés az amerikai Nemzeti Egészségügyi Intézeteknél és a Brain/MINDS a japán Oktatási, Kulturális, Sport, Tudományos és Technológiai Minisztériumnál). Több projekt az AGI létrehozására szolgáló hardverekre és szoftverekre összpontosít (pl. az IBM Cognitive Computing Project, a Ricardo Gudwin (Campinas Egyetem, Brazília) által vezetett Cognitive Systems Toolkit projekt, valamint a Kwabena Boahen (Stanford Egyetem) által vezetett Neurogrid projekt). Kettő az AGI-tervezés biztonsági szempontjaira összpontosít (Center for Human-Compatible AI a Kaliforniai Egyetemen, Berkeleyben és a Machine Intelligence Research Institute). Egy (Carboncopies) más K+F projektek támogatására összpontosít.

Végül az egyik az AGI elméleti aspektusaira összpontosít (Goedel-gép, amelyet Jürgen Schmidhuber, a svájci Dalle Molle Mesterséges Intelligencia Kutatóintézet vezetője vezet). Ez a lista nem teljes körű az AGI nem K+F szempontjaival foglalkozó projektek listája. Például az AGI etikájával, kockázatával és politikájával foglalkozó projektek nem kerültek bele, mivel ezek távolabb állnak a K+F-től.

6. Következtetés

Az AGI látszólag spekulatív jellege ellenére a K+F már most is folyik. Ez a felmérés 45 AGI K+F projektet azonosított, amelyek 6 kontinens 30 országában találhatóak, és amelyek közül sok a nagyvállalatoknál és a tudományos intézményekben működik, és amelyek közül néhány nagyszabású és nagy finanszírozású. Mivel ez a felmérés kizárólag nyilvánosan közzétett információkra támaszkodik, ez a szám az AGI K+F teljes mértékének alsó határértékeként kezelendő. Így, függetlenül attól, hogy maga az AGI mennyire spekulatív, az erre irányuló K+F egyértelműen nagyon is valós. Tekintettel az AGI potenciálisan nagy etikai, kockázati és politikai tétjére, az AGI K+F projektek folyamatos figyelmet érdemelnek.

6.1 Főbb megállapítások

Az etika tekintetében a fő tendencia a projektek humanitárius és intellektualista célok közötti megosztottsága, az előbbi nagyrészt a vállalati projektekből, az utóbbi pedig a tudományos projektekből származik. A vállalatok humanitárius nyilatkozataival kapcsolatban okkal lehetünk gyanakvóak - lehet, hogy ezek "bluwashing" (5.6. szakasz), hogy elrejtse az önérdékű profitcélokat. Mégis, a nem intellektuális célok által motivált projektek között úgy tűnik, hogy van némi konszenzus legalább a humanitarizmus valamilyen formája mellett, és nem az AGI-diskurzusokban gyakran előforduló más típusú célok, például a kozmiztika/transzhumanizmus mellett. Eközben az intellektualista projektek azt jelzik, hogy az akadémiai K+F projektek még mindig hajlamosak arra, hogy munkájukat intellektuális szempontok szerint, és nem a társadalmi hatások vagy más etikai tényezők szempontjából vizsgálják, még az olyan potenciálisan nagy hatású törekvések esetében is, mint az AGI.

A kockázatot illetően két pont emelkedik ki. Először is, a projektek egyértelmű többségénél nem volt azonosítható elkötelezettség a biztonsággal kapcsolatban. Bár ezek közül sok kis projekt, még

néhány nagyobb projekt is ide tartozik. Úgy tűnik, hogy az AGI kockázatával kapcsolatos aggodalmak nem terjedtek el az AGI K+F nagy részében,

¹¹ Az IBM-et is alaposan átkutatták. Az IBM-nél nem találtak AGI K+F-et, de az IBM-nek van egy AGI-hez kapcsolódó hardverfejlesztési projektje, a Cognitive Computing Project, amelyet a 2. függelék tartalmaz.

¹² Lehetséges, hogy néhányuknak vannak olyan AGI-projektjei, amelyeket nem ismertek el nyilvánosan. Az Apple különösen a titkolózás hírében áll, ami viszonylag valószínűvé teszi, hogy egy nem elismert AGI-projekt gazdája.

különösen a tudományos életben. Másodszor, egyes tendenciák arra utalnak, hogy a kockázatos faji dinamika elkerülhető lenne. Az egyik a projektek, különösen a nagyobb projektek koncentrációja az Egyesült Államokban és szövetségeseiben, ami megkönnyítheti mind az informális koordinációt, mind a hivatalos közpolitikát. A másik a humanitárius segítségnyújtás iránti szerény konszenzus, különösen a nagyobb projektek körében, ami csökkentheti a projektek egymás elleni versengési vágyát. Végezetül, sok projekt kapcsolódik egymáshoz a közös személyzet, az anyaservezetek, az AGI-rendszerek fejlesztése és az azonos közösségekben, például az AGI Society-ben való részvétel révén. Ez azt sugallja, hogy a közösség együtt dolgozik egy közös cél érdekében, és nem egymással verseng a "győzelemért".

A politikát illetően több fontos megállapítás tehető. Az egyik az, hogy a projektek az USA-ban és szövetségeseiben összpontosulnak, beleértve a nagyobb projektek többségét (vagy az összeset, attól függően, hogy mely országok tekinthetők az USA szövetségeseinek). Ez nagymértékben megkönnyítheti az AGI-politika kialakítását, amely a legtöbb AGI K+F-projektre - beleértve az összes nagyobbat is - kiterjedő joghatósággal rendelkezik. Egy másik fontos pont a projektek koncentrációja az akadémiai és vállalati szférában, és viszonylag kevés a kormányzati vagy katonai kapcsolatokkal rendelkező projekt. Az egyes intézménytípusok saját típusú politikát érdemelnek, például felülvizsgálati bizottságokat a tudományos életben és pénzügyi ösztönző struktúrákat a vállalatoknál. Az AGI K+F-profit szinergia lehetősége (5.2. szakasz) itt különösen fontos lehet, mivel meghatározza mind a vállalati K+F mértékét, mind a vállalatok hajlandóságát arra, hogy alávegyék magukat a korlátozó szabályozásoknak. Ez a felmérés az AGI K+F-profit szinergiájára utaló jeleket talál, de nem olyan masszív szinergiákat, mint amelyeket a mesterséges intelligencia egyes más típusai esetében találtak. Végül, a legalább részben nyílt forráskódú projektek túlsúlya megnehezíti a politikai erőfeszítéseket, mivel a K+F-et elvileg bárki és bárhol végezheti.

6.2 Korlátozások és jövőbeli munka

A fenti következtetések e tanulmány módszertanát tekintve szilárdnak tűnnek, de más módszertanok más irányba mutathatnak. Például a célokról szóló közzétett nyilatkozatok a humanitárius és intellektualista célok irányába mutató konszenzusra utalnak, de ez figyelmen kívül hagyhatja az e célok konkrétumaira vonatkozó, nem közzétett nézeteltéréseket. Az egyik potenciális gyűjtőpont az, ha a nyugati humanitárius AGI-projektek a politikai szabadságot, míg a kínai AGI-projektek a gazdasági fejlődést kívánják előmozdítani, párhuzamosan a Kína és a Nyugat közötti szélesebb körű nézeteltéréssel az emberi jogokról. Továbbá, az ebben a felmérésben használt, a célokra vonatkozó közzétett nyilatkozatok eltérhetnek a projektek tényleges céljaitól, ha azok nem teljesen őszinték a közzétett nyilatkozataikban. Előfordulhat, hogy a projektek meggondolatlanul önző célok felé haladnak, miközben felelős, etikus látszatot keltenek a nyilvánosság előtt. Ezek a lehetőségek nehezebb politikai kihívást és nagyobb AGI-kockázatot jelentenek. Másik lehetőség, hogy ez a tanulmány túlbecsüli a kockázatot. Lehet, hogy sok projektnek hasonló céljai és biztonsági aggályai vannak, még akkor is, ha nem tettek közzé erre vonatkozó nyilatkozatokat. A jövőbeni, más módszereket alkalmazó kutatások, különösen az AGI K+F-ben részt vevő személyekkel készített interjúk további felismeréseket nyújthatnak.

Egy másik bonyodalom abból a lehetőségből adódik, hogy vannak olyan AGI K+F projektek, amelyeket ez a tanulmány nem azonosított. Egyes projektek nyilvánosan jelen vannak, de a tanulmányban végzett keresések során egyszerűen nem találtak rájuk, a teljeskörűsége irányuló erőfeszítések ellenére sem. Ez a tanulmány különösen nagy valószínűséggel hagyta figyelmen kívül a nem angol nyelven működő projekteket, mivel csak angol nyelven végzett kereséseket. Az angol azonban a mesterséges intelligencia és általában a tudomány elsődleges nemzetközi nyelve, így valószínű, hogy nem angol nyelvű projektek nem maradtak ki.

Ezen túlmenően további olyan AGI K+F projektek is létezhetnek, amelyeknek egyáltalán nincs

nyilvános arculata. Ezek közé tartozhatnak a titkos kormányzati katonai projektek, amelyektől egyesek tartanak. A felmérésben azonosított projektek katonai kapcsolatainak szerény jellege azonban azt sugallja, hogy jelenleg nem léteznek jelentős katonai AGI-projektek. Konkrétan, a katonai kapcsolatokkal azonosított projektek általában kicsik, és (katonai mércével mérve) hétköznapi taktikai kérdésekre összpontosítanak, nem pedig nagyszabású, globális hódító törekvésekre. Ez valószínűsíti, hogy nincsenek ambiciózusabb titkos

katonai AGI-projektek jelenleg. Tekintettel azonban a tétre, fontos, hogy továbbra is éberek maradjunk a lehetőséggel kapcsolatban.

Egy másik, valószínűbb lehetőség a magánszektorbeli projektek lopakodó üzemmódja. Ez a tanulmány egy olyan lopakodó projektet azonosított, amely történetesen nyilvános weboldallal rendelkezik (Susaro); talán vannak mások is, amelyeknek nincs weboldala. Ezek közül néhányan független startupok lehetnek, míg mások nagyobb mesterséges intelligencia szervezeteken belüli projektek. Különösen a nagyobb szervezetek gyakran rendelkeznek forrásokkal az AGI K+F támogatására, és ezt más AI-projektek keretében álcázhatják. Néhány nagyobb AI-szervezet, például az Apple, titokzatos hírnévvel rendelkezik, és valószínűsíthetően alkalmasak lehetnek arra, hogy titkos AGI-projekteknek adjanak otthont. Az ebben a felmérésben nem azonosított titkos vagy egyéb projektek lehetősége azt jelenti, hogy a felmérésben azonosított projektek számát (45) alsó határként kell kezelni.

A további AGI K+F projektek másik lehetséges forrása a mélytanulásra és a kapcsolódó technikákra összpontosító projektek hatalmas területe. Ezeket a projekteket kizártuk ebből a felmérésből, mivel túl sok van belőlük ahhoz, hogy a tanulmányban alkalmazott projektenkénti módszertan alapján értékelni lehessen őket, és mivel eltérő vélemények vannak arról, hogy ezek a projektek AGI K+F-nek minősülnek-e. Ha az AGI a mélytanulásból és a kapcsolódó technikákból származhatna, akkor az AGI K+F-térképe lényegesen eltérne az ebben a tanulmányban lefestett képtől, ami jelentős következményekkel járna az etikára, a kockázatokra és a politikára nézve. Ezért a jövőbeni kutatások fontos iránya annak felmérése, hogy az AGI a mélytanulás és a kapcsolódó technikák révén létrejöhet-e, majd ennek az etikával, a kockázattal és a politikával való összefüggésbe hozása.

A jövőbeli kutatások másik érdemes iránya a projektek AGI-építési képessége. Ez a tanulmány a projektméretet a képességek helyettesítőjeként használja, de ez egy tökéletlen helyettesítő. A képesség a fontosabb jellemző annak megértéséhez, hogy mely projektek állnak a legközelebb az AGI kiépítéséhez. A nagyobb képességű projektek nagyobb politikai figyelmet érdemelnek, és a hasonló képességű projektek csoportosulása kockázatos verseny dinamikához vezethet. (A projektméret más okokból is fontos, például a projektek munkaerő-piaci vonzereje az AGI-kutatók számára vagy a politikai lobbizás lehetőségei miatt). A projektek kapacitását a projektek eddigi teljesítményének részleteire, az AGI technikai megközelítéseinek életképességére és egyéb tényezőkre való odafigyeléssel lehet értékelni. Tekintettel a projektkapacitás etikai, kockázati és politikai jelentőségére, ez a jövőbeli kutatások fontos iránya.

Végezetül a jövőbeli kutatások az AGI-ban részt vevő egyéb szereplőket is vizsgálhatnák. A K+F-projekteken kívül léteznek többek között az AGI kapcsolódó aspektusaival, például a hardverrel és a biztonsági intézkedésekkel foglalkozó K+F-csoportok; az AGI etikáját, kockázatát és politikáját tanulmányozó és azzal foglalkozó személyek; valamint az AGI bizonyos felfogását támogató "episztemikus aktivisták". E népszerűségcsoportok mindegyike jelentős szerepet játszhat az AGI végeredményeiben, és szintén olyan etikai, kockázati és politikai következményekkel jár, amelyeket érdemes megfontolni. E népszerűségcsoportok empirikus vizsgálata tisztázhatja az elvégzett munka jellegét, azonosíthatja a hiányosságokat, és tendenciákat vetíthet fel az AGI lehetséges megvalósulását illetően.

6.3 Záró megjegyzések

Összességében a jelen tanulmány azt mutatja, hogy az AGI etikája, kockázata és politikája szilárd empirikus alapokon nyugszik - nem kell kizárólag hipotetikus AGI K+F projektekkal és szereplőkkel kapcsolatos spekulációkra alapozni. A jelen tanulmány ezen túlmenően előrelépést tesz ezen az empirikus alapon azáltal, hogy hozzájárul az AGI K+F projektek első, etikai, kockázati és politikai felméréséhez. Tekintettel az AGI potenciálisan nagy téjére, reméljük, hogy ez a kutatás produktívan

felhasználható az AGI eredményeinek javítására.

1. függelék. Aktív AGI K+F projektek

ACT-R

Fő weboldal: <http://act-r.psy.cmu.edu>

Az ACT-R egy kutatási projekt, amelyet John Anderson (Carnegie Mellon Egyetem) vezet. Ez az emberi megismerés elmélete és az emberi megismerés szimulálására szolgáló számítógépes keretrendszer.¹³ Az ACT-R az Adaptive Control of Thought-Rational rövidítése. Rövid ideig a SAL-projekten keresztül kapcsolódott a Leabrához.

Vezető intézmények: Carnegie Mellon Egyetem

Partnerintézmények: nincs

- Az ACT-R honlapja 21 ország számos együttműködő intézményét sorolja fel,¹⁴ bár ez a lista olyan személyeket is tartalmaz, akik korábban közreműködtek, de azóta más projektekben dolgoznak, és nem tartalmazza a közelmúltban megjelent tanulmányok társszerzőit.¹⁵ A honlapról nem sikerült megerősíteni egyetlen aktív partnerintézményt sem, ezért itt nem szerepelnek kódolva, bár lehet, hogy vannak aktív partnerintézmények.

Az intézmény típusa: egyetemi

Nyílt forráskódú: igen¹⁶

Katonai kapcsolat: igen¹⁷

Vezető ország: USA

Partnerországok: nincs

Meghatározott célok:

intellektualizmus

- Az ACT-R fő leírása a honlapon kizárólag a szellemi kutatásról szól, a szélesebb körű hatásokról nem tesz említést.¹⁸

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes

¹³ <http://act-r.psy.cmu.edu/about>

¹⁴ <http://act-r.psy.cmu.edu/people>

¹⁵ Például Lee et al. (2017) vezető szerzője Hee Seung Lee a Yonsei Egyetemről, aki nem szerepel a <http://act-r.psy.cmu.edu/people> oldalon.

¹⁶ <http://act-r.psy.cmu.edu/software>

¹⁷ Az Egyesült Államok Haditengerészeti Kutatási Hivatalának finanszírozásáról Zhang et al. (2016) számol be.

¹⁸ <http://act-r.psy.cmu.edu/about>

AERA

Fő weboldal: <http://www.ru.is/faculty/thorisson>

Az AERA-t Kristinn Thórisson, a Reykjavíki Egyetem munkatársa vezeti. Az AERA az Autocatalytic Endogenous Reflective Architecture rövidítése (Nivel et al. 2013). A projekt célja "az elme megértése és egy gyakorlati AGI-rendszer építése", és jelenleg "a gépi megértés, a mesterséges tanulók tanítási módszertana, sőt az etikai értékek kialakulásának tanulmányozására is használják".¹⁹

Vezető intézmények: Reykjavíki Egyetem

Partnerintézmények: Dalle Molle Institute for Artificial Intelligence Research (Svájc) (a Steunebrink et al. 2016-ban felsorolt szerzők szerint).

Az intézmény típusa: egyetemi

Nyílt forráskódú: nem

Katonai kapcsolat: nincs

- Thórisson, a projekt vezetője kritizálja a katonai mesterséges intelligenciát az Izlandi Intelligens Gépek Intézetében (IIIM).²⁰

Vezető ország: Izland

Partnerországok:

Kijelölt célok: humanitárius, intellektualizmus

- Thórisson honlapja hivatkozik az IIIM etikai politikájára, amelynek célja "a világ tudományos megértésének előmozdítása, és e tudás alkalmazásának lehetővé tétele az emberiség javára és javítására", hangsúlyt fektetve a magánélet védelmével és a katonai visszaélésekkel kapcsolatos aggályokra.²¹
- Az etikai politika azt a célt is megfogalmazza, hogy "a kutatásokat olyan témákra és kihívásokra összpontosítja, amelyek nyilvánvalóan a közvélemény számára hasznosak, és a társadalom, az emberi megélhetés és a földi élet javítását szolgálják". A földi élet említése ököcentrizmust sugall, de minden más szöveg humanitárius vagy intellektualista.²²

A biztonsággal kapcsolatos elkötelezettség: aktív

- Az AERA csoport írt arról, hogyan lehet növelni a biztonságot az AGI önfejlesztése során, azzal érvelve, hogy bizonyos tervezési elvek megkönnyítenék a biztonság elérését (Steunebrink et al. 2016).
- Thórisson honlapja Oren Etzioni (2016), az AI-kutató egyik cikkére hivatkozik, amely elutasítja az AGI-vel kapcsolatos aggodalmakat, ami arra utal, hogy az AERA is elutasító lehet, de erre nem lehetett következtetni pusztán a link jelenlétéből.

Méret: közepes-kicsi

¹⁹<http://www.ru.is/faculty/thorisson>

²⁰ <http://www.iiim.is/ethics-policy/>

²¹ <http://www.iiim.is/ethics-policy/3>

²² <http://www.iiim.is/ethics-policy/3>

AIDEUS

Fő weboldal: <http://aideus.com>

Az AIDEUS-t Alekszej Potapov a szentpétervári ITMO Egyetemről és Szergej Rodionov az Aix Marseille-i Egyetemről vezeti. Potapov a SingularityNET tanácsadója is.²³ Céljaként "egy erős mesterséges intelligencia létrehozását" tűzte ki.²⁴ Megközelítése az, hogy "az optimális cselekvések kiválasztásához használt algoritmikus valószínűség alapján univerzális előrejelzési modellekből indul ki".²⁵ Gyakran publikál az AGI-ről, gyakran az AGI-konferenciák anyagaiban.²⁶ A közelmúltban megjelent publikációk az Orosz Föderáció, többek között az Oktatási és Tudományos Minisztérium finanszírozásáról számolnak be (pl. Potapov et al. 2016).

Vezető intézmények:

Intézmény típusa: nincs.

- A projekt weboldalán nem szerepel intézménytípus. Az AIDEUS a publikációkban (pl. Potapov et al. 2016) az akadémiai hovatartozástól elkülönítve szerepel (pl. Potapov et al. 2016). A Facebook-oldalán, a²⁷ oldalon céggént szerepel, de a Facebook "cég" kategóriája nem korlátozódik a vállalatokra.

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

- A finanszírozási források között szerepel a "Government of Russian Federation, Grant 074-U01", amely nem tűnik katonai jellegűnek, de ezt nem lehetett megerősíteni.

Vezető ország: Oroszország

Partnerországok:

Kijelölt célok: humanitárius, értelmiségi

- A projekt célja a szuperintelligencia kiépítése, hogy "segítsen jobban megérteni saját gondolkodásunkat, és megoldani a nehéz tudományos, műszaki, társadalmi és gazdasági problémákat".²⁸

A biztonsággal kapcsolatos elkötelezettség: aktív

- Az AIDEUS AGI biztonsági kutatásokat tett közzé, pl. Potapov és Rodionov (2014).

Méret: kicsi

²³<https://singularitynet.io>

²⁴ <http://aideus.com>

²⁵ <http://aideus.com/research/research.html>

²⁶ <http://aideus.com/research/publications.html>

²⁷ <https://www.facebook.com/pg/Aideus-Strong-artificial-intelligence-455322977847194/about>

²⁸ <http://aideus.com/community/community.html>

AIXI

Fő weboldal: <http://www.hutter1.net/ai/aixigentle.htm>

Az AIXI-t Marcus Hutter, az Ausztrál Nemzeti Egyetem munkatársa vezeti. Az AIXI egy "meta-algoritmuson" alapul, amely az algoritmusok terében keresi az AGI számára legjobbat.²⁹ Hutter bizonyítja, hogy az AIXI megtalálja a legintelligensebb mesterséges intelligenciát, ha végtelen számítási teljesítményt kapunk. Bár ez egy tisztán elméleti eredmény, közelítő változatokhoz vezetett, amelyeket számítógépes kódban valósítottak meg.³⁰

Vezető intézmények: Ausztrál Nemzeti Egyetem

Partnerintézmények: nincs

Az intézmény típusa: Nyílt

forráskódú: nincs

Katonai kapcsolat: meghatározatlan

Vezető ország:

Meghatározott célok:

meghatározatlan.

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

²⁹ Goertzel (2014, 25. o.)

³⁰ Veness et al. (2011)

Alice Csodaországban (AIW)

Fő weboldal: <https://github.com/arnizamani/aiw> és <http://flov.gu.se/english/about/staff?languageId=100001&userId=xstrel>

Az AIW-t Claes Strannegård, a svédországi Chalmers Műszaki Egyetem munkatársa vezeti. Az AIW-ről a *Journal of Artificial General Intelligence* című folyóiratban megjelent tanulmány a NARS-hoz hasonlónak írja le (Strannegård et al. 2017a). Egy külön tanulmány prototípusként írja le a "szimbolikus és szubszimolikus érvelés közötti szakadék áthidalására" vonatkozó új elképzelések megvalósítására (Strannegård et al. 2017b).

Vezető intézmények: Chalmers University of Technology

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen³¹

Katonai kapcsolat:

nincs³² **Vezető ország:**

Partnerországok: nincs

Meghatározott célok:

meghatározatlan.

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

³¹<https://github.com/arnizamani/aiw>

³² A bejelentett finanszírozás a Svéd Kutatási Tanácstól származik

Animátumok

Fő weboldal: <https://github.com/nils/animats>

Az Animats egy kis projekt, amelyet a First International Workshop on Architectures for Generality & Autonomy (Első Nemzetközi Műhely az általános és autonóm architektúrákról) számára fejlesztettek ki.³³ és a 2017-es AGI-konferenciára készült.³⁴ A projekt svédországi és egyesült államokbeli egyetemek kutatói, valamint a svájci NNAISENSE vállalat együttműködésében jött létre.³⁵ Célja az állati intelligencián alapuló AGI létrehozása.³⁶

Vezető intézmények: Chalmers University of Technology

Partnerintézmények: Harvard Egyetem, NNAISENSE

Intézmény típusa: egyetemi, magántársaság

Nyílt forráskódú: igen³⁷

Katonai kapcsolat:

nem³⁸ **Vezető ország:**

Svédország

Partnerországok: A biztonsággal

kapcsolatos szerepvállalás: nem

meghatározott **Méret:** kicsi

³³<http://cadia.ru.is/workshops/aga2017>

³⁴ Strannegård et al. (2017)

³⁵ Strannegård et al.

(2017b)³⁶ Strannegård et al.

(2017b)

³⁷<https://github.com/nils/animats>

³⁸ A bejelentett finanszírozás a Svéd Kutatási Tanácstól származik

Baidu kutatás

Fő weboldal: <http://research.baidu.com/learning-speak-via-interaction>

A Baidu Research egy mesterséges intelligencia kutatócsoport a Baidun belül. Irodái Pekingben, Sencsenben és a kaliforniai Sunnyvale-ben találhatóak.³⁹ Weboldalának egyik oldalán az áll, hogy küldetése "általános intelligencia létrehozása a gépek számára",⁴⁰ , bár úgy tűnik, hogy ez nem a csoport fő témája. Sikereket ért el a nyelvfeldolgozásban a "zéró lövéses tanulás" terén, amelynek során a mesterséges intelligencia "képes megérteni a még nem látott mondatokat".⁴¹ Egyes megfigyelők ezt jelentős áttörésnek értékelik.⁴²

Vezető intézmények: Baidu

Partnerintézmények: nincs

Az intézmény típusa: Állami vállalat

Nyílt forráskódú: nem

- Baidu kiadja néhány munka nyílt forráskódú,⁴³ de nem az AGI

Katonai kapcsolat: meghatározatlan

- A Baidu a kínai kormánytól mesterséges intelligencia finanszírozást kap a "számítógépes látás, biometrikus azonosítás, szellemi tulajdonjogok és ember-számítógép interakció" területén.⁴⁴

Vezető ország:

Partnerországok: Kína:

USA Kijelölt célok:

meghatározatlan

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

- Andrew Ng, a Baidu korábbi kutatási főmunkatársa szerint a Baidu komolyan veszi a biztonságot és az etikát, és kifejti személyes véleményét, miszerint az AI segíteni fogja az emberiséget, és hogy "az AI-val és a gyilkos robotokkal kapcsolatos félelmek túlzóak",⁴⁵ de ez nem az AGI-vel összefüggésben hangzott el. A Baidu Research részéről nem találtunk a biztonságról szóló közvetlen vitát.

Méret: közepes-kicsi

³⁹ http://bdl.baidu.com/contact_b.html

⁴⁰<http://research.baidu.com/learning-speak-via-interaction>

⁴¹<http://research.baidu.com/ai-agent-human-like-language-acquisition-virtual-environment>

⁴² Griffin (2017)

⁴³ <https://github.com/baidu>

⁴⁴ Gershgorn (2017)

⁴⁵ Maddox (2016)

Becca

Fő weboldal: <https://github.com/brohrrer/becca>

A Beccát Brandon Rohrer vezeti, aki jelenleg a Facebooknál dolgozik.⁴⁶ Honlapja szerint a Becca "egy általános tanulási program, amely bármilyen robotban vagy testet öltött rendszerben használható"; "arra törekszik, hogy egy legyen bármilyen robot számára, amely bármit csinál".⁴⁷ Rohrer nyílt forráskódú projektként írja le,⁴⁸ és a forráskódja elérhető a weboldalán.

Vezető intézmények:

Intézmény típusa: nincs.

- Becca számára nem találtak intézményi otthont; úgy tűnik, hogy Rohrer szabadidős projektje.

Nyílt forráskódú: igen⁴⁹

Katonai kapcsolat: meghatározatlan

- Becca akkor kezdte, amikor Rohrer a Sandia National Laboratories-nál dolgozott,⁵⁰ de úgy tűnik, hogy ez a kapcsolat már nem aktív.

Vezető ország:

Meghatározott célok:

meghatározatlan

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

⁴⁶ <https://www.linkedin.com/in/brohrrer>

⁴⁷ <https://github.com/brohrrer/becca>

⁴⁸ <https://www.linkedin.com/in/brohrrer>

⁴⁹ <https://github.com/brohrrer/becca>

⁵⁰ <https://www.linkedin.com/in/brohrrer>

Kék agy

Fő weboldal: <http://bluebrain.epfl.ch>

A Blue Brain egy Henry Markram által vezetett kutatási projekt. A projekt 2005 óta működik. Honlapja szerint célja "a rágcsálók, és végső soron az emberi agy biológiailag részletes digitális rekonstrukcióinak és szimulációinak elkészítése".⁵¹ Markram alapította a Human Brain Projectet is, amelynek kutatási stratégiája megegyezik a Blue Brainnel.⁵²

Vezető intézmény: École Polytechnique Fédérale de Lausanne

Partnerintézmények: Egyetem (Izrael); Universidad Politecnica de Madrid és Consejo Superior de Investigaciones Científicas (Spanyol Nemzeti Kutatási Tanács) (Spanyolország); University of London (Egyesült Királyság); és IBM, St. Elizabeth's Medical Center, University of Nevada-Reno, és Yale University (USA).⁵³

Az intézmény típusa: egyetemi

- A Blue Brain az IBM közvállalatot nevezi meg közreműködőként, amely "kutatóit rendelkezésre bocsátotta, hogy segítsenek a BlueGene szuperszámítógép telepítésében és a neuronhálózat ~~szimulációjában~~ megfelelő áramkörök létrehozásában".⁵⁴ Ezt a hozzájárulást túl csekélynek ítélték ahhoz, hogy a Blue Brain-t közvállalat részeként kódolják.

Nyílt forráskódú: igen⁵⁵

Katonai kapcsolat: meghatározatlan

Vezető ország: Svájc

Partnerországok: Egyesült Királyság, Izrael, Spanyolország, USA

Kijelölt célok: humanitárius, intellektualizmus

- A Blue Brain honlapja szerint "az agy megértése létfontosságú, nemcsak a gondolatainkat és érzelmeinket adó biológiai mechanizmusok megértéséhez, amelyek emberré tesznek minket, hanem gyakorlati okokból is", ez utóbbiak közé tartozik a számítástechnika, a robotika és az orvostudomány.⁵⁶ Ezek az alkalmazások nagyjából humanitárius jellegűek, bár ez egy visszafogottabb humanitárius, mint ami számos más projekt esetében megtalálható.

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-nagy

⁵¹<http://bluebrain.epfl.ch/page-56882-en.html>

⁵²<http://bluebrain.epfl.ch/page-52741-en.html>

⁵³<http://bluebrain.epfl.ch/page-56897-en.html>

⁵⁴<http://bluebrain.epfl.ch/page-56897-en.html>

⁵⁵<https://github.com/BlueBrain>

⁵⁶<http://bluebrain.epfl.ch/page-125344-en.html>

Kína Agy projekt

Fő weboldal: nem található

A China Brain Project a Kínai Tudományos Akadémia kutatási projektje, amely az alapvető és klinikai idegtudományokra és az agy által inspirált számítástechnikára összpontosít. A Kínai Tudományos Akadémia 2016 júliusában jelentette be a projektet, és közölte, hogy hamarosan elindul,⁵⁷ de 2017 augusztusában nem találtak a projekt weboldalát. A projektet vezető Mu-Ming Poo a *Neuron című* folyóiratban 2016-ban megjelent cikkében a következőket írta le a projektről: "Az agy információfeldolgozási mechanizmusából való tanulás egyértelműen ígéretes út az erősebb és általánosabb gépi intelligencia kiépítéséhez", és "A China Brain Project a kognitív robotika fejlesztésére összpontosítja majd erőfeszítéseit, mint az agy által inspirált számítási modellek és eszközök integrálásának platformjára".⁵⁸

Vezető intézmények: Kínai Tudományos Akadémia

Partnerintézmények: nincs

Az intézmény típusa: kormányzat

- A Kínai Tudományos Akadémia a kínai kormány alá tartozó közintézmény.⁵⁹
- Mu-Ming Poo a Kínai Természettudományi Alapítványt és a Tudományos és Technológiai Minisztériumot nevezi meg a China Brain Projectet irányító szervezetként.⁶⁰

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

Vezető ország: Kína

Partnerországok: nincs

Kijelölt célok: humanitárius, intellektuális

- Mu-Ming Poo a projekt céljai között említi "az emberi megismerés idegi alapjainak megértését" és "a súlyos agyi rendellenességek társadalmi terheinek csökkentését".⁶¹

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

⁵⁷ http://english.cas.cn/newsroom/news/201606/t20160617_164529.shtml

⁵⁸ Poo et al. (2016)

⁵⁹ http://english.cas.cn/about_us/introduction/201501/t20150114_135284.shtml

⁶⁰ Poo et al. (2016)

⁶¹ Poo et al. (2016)

CLARION

Fő honlap: <http://www.clarioncognitivearchitecture.com>,
<http://www.cogsci.rpi.edu/~rsun/clarion.html>

A CLARION-t Ron Sun, a Rensselaer Polytechnic Institute munkatársa vezeti az 1994 körüli kutatásokban.⁶² Ez az implicit és explicit tudás hibridje, amelyet Goertzel (2014, 23. o.) erőteljes, de hiányos és nem jól integrált megközelítésként ír le.

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

Vezető intézmények: Rensselaer Polytechnic Institute

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen⁶³

Katonai kapcsolat: igen⁶⁴

Vezető ország:

Partnerországok: nincs

Kijelölt célok: értelmiségi

- A CLARION honlapja szerint "célja az emberi elme alapvető struktúráinak vizsgálata", "végső célja pedig az, hogy egységes magyarázatot adjon a kognitív jelenségek széles körére".⁶⁵

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-kicsi

⁶² <http://www.cogsci.rpi.edu/~rsun/hybrid-resource.html>

⁶³ <http://www.cogsci.rpi.edu/~rsun/clarion.html>

⁶⁴ Az USA Haditengerészeti Kutatási Hivatalának és a Hadsereg Kutatóintézetének finanszírozásáról a

<http://www.cogsci.rpi.edu/~rsun/clarion.html> oldalon olvasható jelentés.

⁶⁵ <http://www.clarioncognitivearchitecture.com/home>

CogPrime

Fő weboldal: http://wiki.opencog.org/w/CogPrime_Overview

A CogPrime egy nyílt forráskódú projekt, amelyet Ben Goertzel vezet. Goertzel vezeti a SingularityNET-et is. Ez "több, különböző memóriatípusokhoz kapcsolódó tanulási algoritmust integrál", amelyeket jelenleg videojáték-karakterek vezérlésére használnak, és amelyeket humanoid robotok vezérlésére terveznek.⁶⁶ A CogPrime honlapja szerint ez egy olyan megközelítés "az általános intelligencia eléréséhez, amely minőségileg emberi szintű és sok tekintetben emberhez hasonló".⁶⁷

Vezető intézmény: OpenCog Alapítvány

Partnerintézmények: nincs

Az intézmény típusa: nonprofit

- Az OpenCog 501(c)(3) státuszú nonprofit szervezetként van leírva, és folyamatban van a bejelentése.⁶⁸

Nyílt forráskódú: igen⁶⁹

Katonai kapcsolat: meghatározatlan

- Az OpenCog aggodalmát fejezte ki a katonai AGI-vel kapcsolatban.⁷⁰

Vezető ország: USA

- Bár az OpenCog rendkívül nemzetközi, a honlapja az 501(c)(3) nonprofit státuszra vonatkozó kérelemre utal, a⁷¹ azt jelenti, hogy az adminisztrációs székhelye az Egyesült Államokban van.

Partnerországok: Kína, Etiópia⁷²

- Goertzel (2012b) azt írja, hogy "az OpenCog fejlesztői több kontinensen dolgoznak".

Kijelölt célok: meghatározatlan

- A CogPrime honlapja szerint a CogPrime képes számos etikai nézet megvalósítására.⁷³ Máshol a CogPrime vezetője, Goertzel a kozmisztika mellett érvel (Goertzel 2010).

A biztonsággal kapcsolatos elkötelezettség: aktív

- A CogPrime azt állítja, hogy az AGI "képesnek kell lennie arra, hogy megbízhatóan sokkal magasabb szintű, köznapi etikus viselkedést érjen el, mint bármely emberi lény", hozzátevé, hogy "A CogPrime célrendszerének részletes tervezésével kapcsolatos kutatásaink nem rontották ezt a meggyőződést".⁷⁴

Méret: közepes-nagy

⁶⁶ Goertzel (2014, 23. o.)

⁶⁷ http://wiki.opencog.org/w/CogPrime_Overview

⁶⁸ <http://opencog.org/about>

⁶⁹ <https://github.com/opencog>

⁷⁰ <http://blog.opencog.org/2014/01/17/what-is-consciousness>

⁷¹ <http://opencog.org/about>

⁷² Goertzel (2008); <http://opencog.org/2013/04/new-opencog-ai-lab-opens-in-addis-ababa-ethiopia>

⁷³http://wiki.opencog.org/w/CogPrime_Overview

⁷⁴http://wiki.opencog.org/w/CogPrime_Overview

CommAI

Fő weboldal: <https://research.fb.com/projects/commai>

A CommAI a FAIR (Facebook AI Research) projektje. A FAIR-t Yann LeCun vezeti, aki a New York-i Egyetem tanára is. A FAIR honlapján az áll, hogy "az emberi szintű intelligenciával rendelkező rendszerek megértésére és fejlesztésére törekszik a mesterséges intelligenciával kapcsolatos hosszabb távú tudományos problémák megoldásának előmozdításával".⁷⁵ A CommAI a FAIR projektje, amely leginkább az AGI-ra összpontosít. A CommAI csoport egy első dokumentumot tett közzé "CommAI: Evaluating the First Steps Towards a Useful General AI" címmel.⁷⁶ A CommAI-t a GoodAI által szponzorált általános mesterséges intelligencia versenyen használják.⁷⁷

Vezető intézmény: Facebook

Partnerintézmények: nincs

Az intézmény típusa: állami vállalat

Nyílt forráskódú: igen⁷⁸

Katonai kapcsolat: meghatározatlan

- Úgy tűnik, hogy a Facebooknak nincsenek amerikai védelmi szerződésai.⁷⁹

Vezető ország: USA

Partnerországok:

- A FAIR székhelye New York Cityben található, de irodái vannak a kaliforniai Menlo Parkban és Párizsban is.⁸⁰

Kijelölt célok: humanitárius szellem

- A CommAI honlapján az áll, hogy "célja olyan általános célú mesterséges ágensek kifejlesztése, amelyek *hasznosak* az emberek számára a mindennapi munkájuk során" (kiemelés az eredetiben).⁸¹ A Facebook alapító partnere a Partnership on AI to Benefit People & Society nevű szervezetnek is, amelynek humanitárius céljai vannak.⁸²

A biztonsággal kapcsolatos elkötelezettség: mérsékelt

- A Facebook alapító partnere a Partnership on AI to Benefit People & Society nevű szervezetnek, amely aggodalmát fejezi ki a mesterséges intelligencia biztonságával kapcsolatban,⁸³ de a CommAI nem mutat közvetlen részvételt a biztonsággal kapcsolatban.

Méret: közepes

⁷⁵ <https://research.fb.com/category/facebook-ai-research-fair>

⁷⁶ <https://research.fb.com/publications/commai-evaluating-the-first-steps-towards-a-useful-general-ai>

⁸³ <https://www.partnershiponai.org/tenets>

⁷⁷ <https://research.fb.com/projects/commai>

⁷⁸ <https://github.com/facebookresearch/CommAI-env>

⁷⁹ [https://www.fpds.gov/ezsearch/fpdsportal?q=facebook+DEPARTMENT_FULL_NAME%3A+\"DEPT+OF+DEFENSE\"](https://www.fpds.gov/ezsearch/fpdsportal?q=facebook+DEPARTMENT_FULL_NAME%3A+\)

⁸⁰ <https://code.facebook.com/posts/447727685394738/facebook-expands-ai-research-team-to-paris>

⁸¹ <https://research.fb.com/projects/commai>

⁸² <https://www.partnershiponai.org/tenets>

⁸³ <https://www.partnershiponai.org/tenets>

Cyc

Fő weboldal: <http://www.cyc.com>

A Cyc vezetője Doug Lenat, aki 1984-ben kezdte a Cyc-et. A Cyc a Cycorp projektje, egy texasi székhelyű vállalaté, amely a Cyc-et más vállalatoknak és kormányzati ügynökségeknek nyújtott tanácsadásra és egyéb szolgáltatásokra használja.⁸⁴ A Cycorp honlapja a Cyc-et "hosszú távú törekvésként írja le, hogy valódi mesterséges intelligenciát fejlesszen ki".⁸⁵ A Cyc egyedülálló, több millió, kézzel kódolt, józan emberi tudásból álló adatbázissal rendelkezik, amelyet emberi szintű AGI létrehozására kíván felhasználni.⁸⁶ Egy interjúban Lenat azt mondja: "A Cycorp célja az *általános emberi tudás és józan ész kódolása*, hogy a számítógépek is felhasználhassák" (kiemelés az eredetiben).⁸⁷

Vezető intézmény: Cycorp

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: korlátozott

- A Cycorp kérésre költségmentes licencet kínál a kutatók számára.⁸⁸
- A Cyc egy részét rövid ideig OpenCyc néven tették elérhetővé, de ez megszűnt.⁸⁹
- A Cycorp "a Cyc ontológia magját közkinccsé tette".⁹⁰

Katonai kapcsolat: igen

- A Cycorp 25 millió dolláros szerződést kapott a terrorizmus elemzésére az amerikai hadsereg számára.⁹¹

Vezető ország:

Meghatározott célok:

meghatározatlan

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes

⁸⁴ <http://www.cyc.com/enterprise-solutions>

⁸⁵ <http://www.cyc.com/about/company-profile>

⁸⁶ Goertzel 2014, 16. o.

⁸⁷ Szelem (2014)

⁹¹ <http://www.cyc.com/about/media-coverage/computer-save-world>; lásd még Deaton et al. (2005).

⁸⁸<http://www.cyc.com/platform/researchcyc>

⁸⁹ <http://www.cyc.com/platform>

⁹⁰<http://www.cyc.com/about/company-profile>

⁹¹ <http://www.cyc.com/about/media-coverage/computer-save-world>; lásd még Deaton et al. (2005).

DeepMind

Fő weboldal: <http://deepmind.com>

A DeepMind egy AI vállalat, amelyet Demis Hassabis, Shane Legg és Mustafa Suleyman vezet. A 2010-ben alapított vállalatot 2014-ben vásárolta fel a Google 400 millió fontért (650 millió dollárért; Gibbs 2014). Olyan rendszerek kifejlesztésére törekszik, amelyek képesek megtanulni bármilyen összetett probléma megoldását anélkül, hogy megtanítanák, hogyan kell megoldani", és abból a feltevésből indul ki, hogy a mesterséges intelligenciának általánosnak kell lennie.⁹² A DeepMind publikációkat tesz közzé az AGI-ről, pl. "PathNet: Evolution Channels Gradient Descent in Super Neural Networks" (Fernando et al. 2017).

Vezető intézmény: Google

Partnerintézmények: nincs

Az intézmény típusa: állami vállalat

Nyílt forráskódú: igen⁹³

Vezető ország: UK

Katonai kapcsolat: meghatározatlan

- A Google kiterjedt védelmi szerződésekkel rendelkezik az Egyesült Államokban,⁹⁴ de úgy tűnik, hogy ezek nem kapcsolódnak a DeepMindhoz.

Partnerországok: USA, Kanada

- A DeepMind londoni székhelyű, és a Google kaliforniai Mountain View-ban található központjában is van egy csapata (Shead 2017). Nemrégiben irodát nyitott az albertai Edmontonban.⁹⁵

Kijelölt célok: humanitárius szellem

- Honlapjukon a következő szlogent mutatják be: "Solve intelligence. Használjuk, hogy jobba tegyük a világot."; leírja, hogy "az AI az emberi találékonyság multiplikátora", hogy olyan problémákat oldjon meg, mint az éghajlatváltozás és az egészségügy; és kijelenti, hogy "hiszünk abban, hogy az AI-nak végső soron a világnak kell lennie, hogy sokak és ne kevesek javát szolgálja".⁹⁶ Hasonlóképpen azt írja, hogy a mesterséges intelligencia "segíteni fogja az emberiséget a legnagyobb kihívások leküzdésében, az éghajlatváltozástól kezdve a fejlett egészségügyi ellátás biztosításáig".⁹⁷

A biztonsággal kapcsolatos elkötelezettség: aktív

- A DeepMind a felvásárlás során ragaszkodott ahhoz, hogy a Google-nál egy mesterséges intelligencia etikai bizottságot hozzanak létre (Gibbs 2014). Együttműködik az OpenAI-val hosszú távú AI biztonsági projekteken.⁹⁸ A Google-tól függetlenül részt vesz a Partnership on AI to Benefit People & Society (Partnerség az AI-ról az emberek és a társadalom javára) elnevezésű szervezetben is.

⁹⁸ <https://blog.openai.com/deep-reinforcement-learning-from-human-preferences>

Méret: nagy

⁹² <https://deepmind.com/blog/open-sourcing-deepmind-lab>

⁹³ <https://github.com/deepmind>

⁹⁴ [https://www.fpbs.gov/ezsearch/fpbsportal?q=google+DEPARTMENT_FULL_NAME%3A "DEPT+OF+DEFENSE"](https://www.fpbs.gov/ezsearch/fpbsportal?q=google+DEPARTMENT_FULL_NAME%3A+\)

⁹⁵ <https://deepmind.com/blog/deepmind-office-canada-edmonton>

⁹⁶ <https://deepmind.com/about>

⁹⁷ <https://deepmind.com/blog/learning-through-human-feedback>

⁹⁸ <https://blog.openai.com/deep-reinforcement-learning-from-human-preferences>

DeSTIN (mély tér-időbeli következtetési hálózat)

Fő honlap: <http://web.eecs.utk.edu/~itamar/Papers/BICA2009.pdf> és <http://wiki.opencog.org/w/DeSTIN>.

A DeSTIN-t eredetileg Itamar Arel és munkatársai fejlesztették ki a Tennessee Egyetemen. Az OpenCog nyílt forráskódú mesterséges intelligencia projekt is fejleszti. A DeSTIN mély tanulást használ a mintafelismeréshez. Az OpenCog honlapja szerint az OpenCog "átvette ezt az akadémiai projektet, hogy előkészítse a nyílt forráskódú kiadásra".⁹⁹ Goertzel (2014, 17. o.) megjegyzi, hogy a DeSTIN "beépült a CogPrime architektúrájába... de elsősorban azért fejlesztik, hogy saját AGI-tervezésének központjaként szolgáljon".

Vezető intézmény: Tennessee Egyetem

Partnerintézmény: OpenCog Foundation

Intézmény típusa: egyetemi, nonprofit

intézmény **Nyílt forráskódú:** igen¹⁰⁰

Katonai kapcsolat: meghatározatlan

Vezető ország:

Meghatározott célok:

meghatározatlan

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

⁹⁹<http://wiki.opencog.org/w/DeSTIN>

¹⁰⁰ <https://github.com/opencog/destin>

DSO-CA

Fő weboldal: nem található

A DSO-CA Gee Wah Ng és munkatársai projektje a DSO Nemzeti Laboratóriumokban, amely Szingapúr elsődleges nemzeti védelmi kutatási ügynöksége. Ez "egy olyan felső szintű kognitív architektúra, amely az emberi agy információfeldolgozását modellezi", és hasonlóságot mutat a LIDA, a CogPrime és más AGI kognitív architektúrákhoz.¹⁰¹

Vezető intézmény: DSO National Laboratories

Partnerintézmény: nincs

Az intézmény típusa: kormányzat

- A DSO-t 1997-ben "testületté alakították".¹⁰² de a LinkedIn oldalán kormányzati ügynökségként szerepel.¹⁰³

Nyílt forráskódú: nem

Katonai kapcsolat: igen

Vezető ország:

Partnerországok: nincs

Meghatározott célok:

meghatározatlan.

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

¹⁰³ <https://www.linkedin.com/company-beta/15618>

¹⁰¹ Ng et al. (2017)

¹⁰² <https://www.dso.org.sg/about/history>

¹⁰³ <https://www.linkedin.com/company-beta/15618>

FLOWERS (FLOWing Epigenetikus robotok és rendszerek)

Fő weboldal: <https://flowers.inria.fr>

A FLOWERS-t Pierre-Yves Oudeyer, az Inria (Institut National de Recherche en Informatique et en Automatique, azaz a francia Nemzeti Számítástechnikai és Automatizálási Kutatóintézet) és David Filliat, az ENSTA ParisTech munkatársa vezeti. A projekt "olyan mechanizmusokat tanulmányoz, amelyek lehetővé teszik, hogy a robotok és az emberek hosszabb időn keresztül autonóm módon és halmozottan sajátítsák el az új készségek repertoárját".¹⁰⁴

Vezető intézmények: Inria és ENSTA ParisTech

Partnerintézmények: nincs

Intézmény típusa: egyetemi, kormányzati

- Az Inria állami kutatóintézet; az ENSTA ParisTech állami főiskola.

Nyílt forráskódú: igen¹⁰⁵

Katonai kapcsolat: meghatározatlan¹⁰⁶

Vezető ország: Kijelölt

célok: értelmiségi

- A projekt honlapja kizárólag a mesterséges intelligencia kutatásának szellemi aspektusaira összpontosít, és a mesterséges intelligencia mellett a kognitív tudományt is a két kutatási irányzat egyikeként mutatja be.¹⁰⁷

A biztonsággal kapcsolatos elkötelezettség: aktív

- A FLOWERS a biztonságot vizsgálta az ember-robot interakciók összefüggésében.¹⁰⁸

Méret: közepes-kicsi

¹⁰⁴ <https://flowers.inria.fr>

¹⁰⁵ <https://flowers.inria.fr/software>

¹⁰⁶ A közelmúltban megjelent publikációkban említett finanszírozás főként kormányzati tudományos alapítványoktól származik.

¹⁰⁷ <https://flowers.inria.fr>

¹⁰⁸ Oudeyer et al. (2011)

GoodAI

Fő weboldal: <https://www.goodai.com>

A GoodAI egy magántulajdonban lévő vállalat, amelyet Marek Rosa számítógépes játékokkal foglalkozó vállalkozó vezet. Székhelye Prágában található. A céget Rosa finanszírozza, aki legalább 10 millió dollárt fektetett bele. Rosa azt állítja, hogy "a GoodAI az általános mesterséges intelligencia létrehozására irányuló életre szóló álmom felé építkezik. Már 15 éves korom óta erre a célra összpontosítok".¹⁰⁹

Vezető intézmény: GoodAI

Partnerintézmények: nincs

- A GoodAI több partnerszervezetet is felsorol, köztük a szlovákiai székhelyű SlovakStartupot.¹¹⁰ Ebben a tanulmányban azonban egyik partner sem szerepel, mivel nem járulnak hozzá az AGI K+F-hez.

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: igen¹¹¹

Katonai kapcsolat:

meghatározatlan **Vezető ország:**

Partnerországok: nincs

Kijelölt célok: humanitárius, értelmiségi

- A GoodAI honlapja szerint "küldetése az általános mesterséges intelligencia fejlesztése - a lehető leggyorsabban - az emberiség megsegítésére és az univerzum megértésére", és célja, hogy "olyan általános mesterséges intelligenciát hozzon létre, amely képes gyógymódot találni a betegségekre, olyan dolgokat talál fel az emberek számára, amelyek feltalálása sokkal tovább tartana a mesterséges intelligencia együttműködése nélkül, és sokkal többet tanít nekünk az univerzumból, mint amit jelenleg tudunk".¹¹² Hangsúlyozza, hogy az AGI megalkotása "nem verseny. Nem a versengésről szól, és nem a pénzkeresésről".¹¹³

A biztonsággal kapcsolatos elkötelezettség: aktív

- A GoodAI arról számol be, hogy külön AI-biztonsági csapattal rendelkezik¹¹⁴ és Nick Bostrom *Superintelligence* című művét idézi, mint az AI-biztonsággal kapcsolatos kutatási inspirációt.¹¹⁵

Méret: közepes

¹⁰⁹ <https://www.goodai.com/about>

¹¹⁰ <https://www.goodai.com/partners>

¹¹¹ <https://github.com/GoodAI>

¹¹² <https://www.goodai.com/about>

¹¹³ <https://www.goodai.com/about>

¹¹⁴ <https://www.goodai.com/about>

¹¹⁵ <https://www.goodai.com/research-inspirations>

HTM (hierarchikus időbeli memória)

Fő weboldal: <https://numenta.com>

A HTM-et a kaliforniai Redwood Cityben működő Numenta vállalat és egy nyílt forráskódú közösség fejlesztette, amelynek a vállalat ad otthont. A HTM-et Jeffrey Hawkins vezeti, aki korábban a Palm Computingot alapította. A HTM az emberi neokortex modelljén alapul. Honlapjukon ez áll: "Hiszünk abban, hogy a neokortex működésének megértése a leggyorsabb út a gépi intelligenciához... A Numenta messze megelőz minden más csapatot a valódi gépi intelligencia létrehozására irányuló törekvésben".¹¹⁶

Vezető intézmény: Numenta

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: igen¹¹⁷

- A Numenta díjköteles kereskedelmi licencet és nyílt forráskódú licencet kínál.¹¹⁸

Katonai kapcsolat: meghatározatlan

- A HTM-et egy 2008-as Air Force Institute of Technology hallgatói szakdolgozatban használták (Bonhoff 2008).

Vezető ország: USA

Partnerországok: nincs

Kijelölt célok: humanitárius, értelmiségi

- A Numenta vállalati honlapján két napirend szerepel: (Például: "A Numenta minden idők egyik legfontosabb tudományos kihívásával foglalkozik: a neokortex visszafejtésével", és (2) a gépi intelligencia technológiájával, humanitárius témával, pl. kijelentve, hogy "fontos az emberiség további sikere szempontjából".¹¹⁹
- Hawkins azt írja, hogy "az emberiség jövőbeli sikere, sőt túlélése függhet" attól, hogy az emberiség "valóban intelligens gépeket épít-e", és az energia, az orvostudomány és az űrutazás területén történő alkalmazásokra hivatkozik.¹²⁰
- A Numenta Twitteren azt írja, hogy "csak akkor tudjuk valóban megváltoztatni a világot, ha a mesterséges intelligenciát közjővé tesszük, nem pedig néhány kiváltságos tulajdonává".¹²¹

A biztonsággal kapcsolatos elkötelezettség: elutasító

- Hawkins elutasította az AGI-vel kapcsolatos aggodalmakat, mint katasztrofális kockázatot, és kijelentette: "Nem látom, hogy a gépi intelligencia fenyegetést jelentene az emberiségre" (Hawkins 2015).

Méret: közepes

¹¹⁶<https://numenta.com>

¹¹⁷ <https://numenta.org>

¹¹⁸ <https://numenta.com/assets/pdf/apps/licensing-guide.pdf>

¹¹⁹ <https://numenta.com>

¹²⁰ Hawkins (2017) és a Numenta blog: <https://numenta.com/blog/2017/06/IEEE-special-edition-article-by-Jeff>

¹²¹ <https://twitter.com/Numenta/status/892445032736849920>

Emberi Agy Projekt (HBP)

Fő weboldal: <http://www.humanbrainproject.eu>

A HBP az idegtudományi kutatás és az agyszimuláció projektje. Az Európai Bizottság támogatja, és 2013-tól kezdődően tíz éven keresztül összesen 1 milliárd dollárt különítettek el rá.¹²² Eredetileg Henry Makram vezette, de a hosszas kritikákat követően átszervezték (Theil 2015). Központja az École Polytechnique Fédérale de Lausanne-ban található, együttműködő intézményei pedig Európa-szerte és Izraelben működnek.¹²³ Az agyszimuláció, a neuromorfikus számítástechnika és a neurorobotika platformjainak ad otthont.¹²⁴ Markram alapította a Blue Brain is, amely a HBP-vel közös kutatási stratégiát követ.¹²⁵

A megfontolás oka: Nagyszabású agykutatási projekt, hasonlóan a Blue Brainhez.

Vezető intézmények: École Polytechnique Fédérale de Lausanne

Partnerintézmények: A HBP weboldalán felsorolt 116¹²⁶

Az intézmény típusa: egyetemi

Nyílt forráskódú: korlátozott

- A számla megszerzéséhez egy kérelemre és az útlevel másolatának megosztására van szükség.¹²⁷

Katonai kapcsolat: nincs

- A HBP-politika tiltja a katonai alkalmazásokat¹²⁸

Vezető ország: Svájc

Partnerországok: Ausztria, Belgium, Dánia, Egyesült Királyság, Finnország, Franciaország, Görögország, Hollandia, Izrael, Magyarország, Izrael, Németország, Olaszország, Norvégia, Portugália, Spanyolország, Svédország, Szlovénia, Törökország és az Egyesült Királyság.

Kijelölt célok: állatvédelem, humanitárius, értelmiségi

- A HBP agyszimulációval igyekszik "csökkenteni az állatkísérletek szükségességét" és "betegségeket tanulmányozni".¹²⁹ A "megismerés megértése" is központi témaként szerepel.¹³⁰

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

- A HBP etikai programja a kutatási eljárásra, nem pedig a társadalmi hatásokra összpontosít.¹³¹

Méret: nagy

¹²² <http://www.humanbrainproject.eu/en/science/overview/>

¹²³ <http://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners>

¹²⁴ <http://www.humanbrainproject.eu/en/brain-simulation/brain-simulation-platform>; <http://www.humanbrainproject.eu/en/silicon-brains>; <http://www.humanbrainproject.eu/en/robots>¹²⁵
<http://bluebrain.epfl.ch/page-52741-en.html>

¹²⁶ <http://www.humanbrainproject.eu/en/open-ethical-engaged/contributors/partners>

¹²⁷ <http://www.humanbrainproject.eu/en/silicon-brains/neuromorphic-computing-platform>

¹²⁸ <https://nlp.humanbrainproject.eu/documentation>, <https://nlp.humanbrainproject.eu/documentation>

¹²⁹ <http://www.humanbrainproject.eu/en/brain-simulation>

¹³⁰ <http://www.humanbrainproject.eu/en/understanding-cognition>

¹³¹ <https://www.humanbrainproject.eu/en/open-ethical-engaged/ethics/ethics-management>

Icarus

Fő weboldal: <http://csl.stanford.edu/research/ongoing/icarus>

Az Icarust Pat Langley, a Stanford Egyetem munkatársa vezeti. A másik aktív közreműködő Dongkyu Choi a Kansasi Egyetemről, Langley korábbi doktorandusza. Az Icarus az ACT-R-hez és a Soarhoz hasonló kognitív építészeti projekt, amely a fizikai környezetben történő észlelésre és cselekvésre helyezi a hangsúlyt.¹³² A honlapja már elavult, de számos újabban megjelent tanulmányt publikáltak.¹³³

Vezető intézmény: Stanford Egyetem

Partnerintézmények: University of

Kansas **Intézménytípus:** egyetemi

Nyílt forráskódú: nem

Katonai kapcsolat: igen¹³⁴

Vezető ország:

Partnerországok: nincs

Meghatározott célok:

intellektualista

- Choi és Langley (2017) azt írják, hogy "fő célunk" "a megismerési funkciók széles körű lefedettségének elérése" "az intelligens ágensek építésében".

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

¹³² Goertzel (2014); Choi and Langley (2017)

¹³³ Lásd pl. Menager és Choi (2016); Choi és Langley (2017); Langley (2017).

¹³⁴ Az Icarus csoport Choi és Langley (2017) beszámolójában az amerikai Haditengerészeti Kutatási Hivatal, a Navy Research Lab és a DARPA finanszírozásáról számol be.

Leabra

Fő weboldal: <https://grey.colorado.edu/emergent/index.php/Leabra>

A Leabrát Randall O'Reilly, a Coloradói Egyetem munkatársa vezeti. Ez egy kognitív architektúra projekt, amely az idegi tevékenység modellezésére helyezi a hangsúlyt. Egy nemrégiben megjelent tanulmány szerint a Leabra "hosszú távú erőfeszítés az emberi megismerés neurális alapjainak belsőleg konzisztens elméletének megalkotására", és hogy "talán minden más javasolt kognitív architektúránál jobban, a Leabra közvetlenül az agy mögöttes biológiájára épül, és biológiai szempontból reális neurális feldolgozási mechanizmusok alkotják a magját".¹³⁵ Rövid ideig a SAL-projekten keresztül kapcsolódott az ACT-R-hez.

Vezető intézmény: University of Colorado

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen¹³⁶

Katonai kapcsolat: igen¹³⁷

Vezető ország:

Partnerországok: nincs

Meghatározott célok:

intellektualista

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-kicsi

¹³⁵ O'Reilly et al. (2016)

¹³⁶ https://grey.colorado.edu/emergent/index.php/Main_Page

¹³⁷ A Leabra-csoport az Egyesült Államok Haditengerészeti Kutatási Hivatalának és a Hadsereg Kutatási Laboratóriumának finanszírozásáról számol be a <https://grey.colorado.edu/CompCogNeuro/index.php/CCNLab/funding> címen.

LIDA (Learning Intelligent Distribution Agent)

Fő weboldal: <http://ccrg.cs.memphis.edu>

A LIDA-t Stan Franklin, a Memphisi Egyetem munkatársa vezeti. Bernard Baars globális munkaterület-elméletén alapul, "a memória és az intelligens feldolgozás különböző formáit integrálja egyetlen feldolgozási ciklusba" (Goertzel 2014, 24. o.). Goertzel (2014, 24. o.) szerint az LIDA jó alapokkal rendelkezik az idegtudományokban, de csak "alacsonyabb szintű" intelligenciára képes, fejlettebb gondolkodásra nem. A LIDA-t az amerikai Haditengerészeti Kutatási Hivatal támogatja; egy egyszerűbb, IDA nevű változatát arra használják, hogy automatizálják "a tengerészek új beosztásukhoz való hozzárendelésének döntéshozatali folyamatát".¹³⁸

Vezető intézmények: Egyetem: Memphisi Egyetem

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forrás: korlátozott¹³⁹

- A kód letöltéséhez regisztráció szükséges; a kereskedelmi felhasználáshoz kereskedelmi licenc szükséges.

Katonai kapcsolat: igen

Vezető ország:

Partnerországok: nincs

Kijelölt célok: humanitárius, értelmiségi

- A LIDA honlapja szerint "az elme működésének teljes kognitív modelljére" törekszik, és elsősorban intellektuális kutatási célokra összpontosít.¹⁴⁰
- A LIDA kutatóinak, Madl Tamásnak és Stan Franklinnek a tanulmánya szerint a robotoknak etikára van szükségük, hogy "az ember számára előnyös cselekedetekre korlátozzák őket".¹⁴¹
- Franklin utal az AGI erkölcsi státuszának támogatására, megemlítve a mesterséges intelligenciában a "szintetikus érzelmek" tanulmányozását, ami arra utalhat, hogy egy AGI-nak "erkölcsi státuszt kellene kapnia".¹⁴²

A biztonsággal kapcsolatos elkötelezettség: aktív

- Madl és Franklin tanulmánya olyan AGI biztonsági kihívásokkal foglalkozik, mint például az emberi etika finom meghatározása a programozáshoz szükséges pontossággal.¹⁴³
- Franklin Colin Allen és Wendell Wallach mesterséges intelligencia-etikussal is együttműködött azon a kihíváson, hogy az AGI-ket helyes erkölcsi döntések meghozatalára bírják.¹⁴⁴

Méret: közepes

¹³⁸ <http://ccrg.cs.memphis.edu>

¹³⁹ <http://ccrg.cs.memphis.edu/framework.html>

¹⁴⁰ <http://ccrg.cs.memphis.edu>

¹⁴¹ Madl és Franklin (2015)

¹⁴² Wallach et al. (2011, 181. o.)

¹⁴³ Madl és Franklin (2015)

¹⁴⁴ Wallach et al. (2010)

Maluuba

Fő weboldal: <http://www.maluuba.com>

A Maluuba egy montreali székhelyű, mesterséges intelligenciával foglalkozó vállalat, amelyet nemrégiben vásárolt fel a Microsoft. A Maluuba azt írja: "vizióink a mesterséges általános intelligencia megoldása volt, olyan írástudó gépek létrehozásával, amelyek képesek úgy gondolkodni, érvelni és kommunikálni, mint az emberek".¹⁴⁵

Vezető intézmény: Microsoft

Partnerintézmények: nincs

Az intézmény típusa: állami vállalat

Nyílt forráskódú: igen¹⁴⁶

Katonai kapcsolat: meghatározatlan

Vezető ország: Kanada

Partnerországok:

Kijelölt célok: intellektualista, profit

- Maluuba azt írja, hogy "az emberi nyelv megértése rendkívül összetett, és végső soron a mesterséges intelligencia szent grálja", és hogy céljuk "a nyelvi megértés alapvető problémáinak megoldása, azzal a céllal, hogy egy valóban írástudó gépet hozzanak létre".¹⁴⁷
- A Maluuba termékért felelős alelnöke, Mo Musbah szerint a Maluuba általános AI-t szeretne, hogy "skalázható legyen a különböző iparágakban alkalmazott AI-mód tekintetében".¹⁴⁸
- A Microsoft alapító partnere a Partnership on AI to Benefit People & Society nevű szervezetnek, amelynek humanitárius céljai vannak,¹⁴⁹ de úgy tűnik, hogy ez nem terjedt át a Maluuba céljaira.

A biztonsággal kapcsolatos elkötelezettség: mérsékelt

- Maluuba a biztonságot fontos kutatási kihívásnak nevezi.¹⁵⁰ Harm van Seijen, a Maluuba kutatója azt írja, hogy "szerintem az ilyen [a mesterséges intelligencia biztonságáról szóló] viták jót tesznek, bár óvatosnak kell lennünk a félelemkeltéssel".¹⁵¹ A Microsoft is alapító partnere a Partnership on AI to Benefit People & Society nevű szervezetnek, amely aggodalmát fejezi ki az AI biztonságával kapcsolatban.¹⁵² A Maluuba közvetlen biztonsági tevékenységét nem azonosították.

Méret: közepes

¹⁴⁵<http://www.maluuba.com/blog/2017/1/13/maluuba-microsoft>

¹⁴⁶ <https://github.com/Maluuba>

¹⁴⁷<http://www.maluuba.com/blog/2017/1/13/maluuba-microsoft>

¹⁴⁸ Townsend (2016)

¹⁴⁹<https://www.partnershiponai.org/tenets>

¹⁵⁰ <http://www.maluuba.com/blog/2017/3/14/the-next-challenges-for-reinforcement-learning>

¹⁵¹ Townsend (2016)

¹⁵² <https://www.partnershiponai.org/tenets>

MicroPsi

Fő weboldal: <http://cognitive-ai.com>

A MicroPsi vezetője Joscha Bach, a Harvard Program for Evolutionary Dynamics munkatársa. Bach küldetése állítólag "az elme modelljének megalkotása az erős mesterséges intelligencia, azaz az emberi megismeréssel egyenrangú megismerés megteremtésének alapkutatása".¹⁵³ Goertzel (2014, 24. o.) szerint a MicroPsi jó alapokkal rendelkezik az idegtudományokban, és csak alacsonyabb szintű intelligenciára képes. A MicroPsi-ről egy nemrégiben megjelent tanulmány szerint "a mesterséges általános intelligencia architektúrája, amely a kognitív ágensek létrehozásának és szimulációjának keretrendszerén alapul", és 2003-ban kezdődött.¹⁵⁴

Vezető intézmények: Harvard Egyetem

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen¹⁵⁵

Katonai kapcsolat: meghatározatlan

Vezető ország: Kijelölt

célok: értelmiségi

- A MicroPsi honlapja szerint "a MicroPsi egy kis lépés az elme működésének megértése felé".¹⁵⁶

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

¹⁵³<http://bigthink.com/experts/joscha-bach>

¹⁵⁴ Bach (2015)

¹⁵⁵<https://github.com/joschabach/micropsi2>

¹⁵⁶ <http://cognitive-ai.com>

Microsoft Research AI (MSR AI)

Fő weboldal: <https://www.microsoft.com/en-us/research/lab/microsoft-research-ai>

Az MSR AI a Microsoft 2017 júliusában bejelentett AI "kutatási és inkubációs központja".¹⁵⁷ A projekt célja "az intelligencia alapelveinek vizsgálata, beleértve az emberi értelem rejtélyeinek megfejtésére irányuló erőfeszítéseket, és e tudás felhasználása egy általánosabb, rugalmasabb mesterséges intelligencia kifejlesztéséhez".¹⁵⁸ A projekt a Microsoft redmondi központjában a mesterséges intelligencia különböző ágainak több mint 100 kutatóját fogja össze.¹⁵⁹ Az MSR AI a különböző ágak összefogásával olyan kifinomultabb mesterséges intelligenciát szeretne létrehozni, mint például "olyan rendszerek, amelyek megértik a nyelvet, és ennek alapján cselekednek".¹⁶⁰ Ugyanakkor kritikák is érik a potenciálisan nehézkes szervezeti struktúra miatt.¹⁶¹

Vezető intézmények: Microsoft

Partnerintézmények: nincs

Az intézmény típusa: állami vállalat

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

- A Microsoftnak van egy katonai ügyekkel foglalkozó programja,¹⁶² de ennek kapcsolata az MSR AI-val nem világos.

Vezető ország: USA

Partnerországok: nincs

Kijelölt célok: humanitárius, értelmiségi

- Satya Nadella, a Microsoft vezérigazgatója széleskörű humanitárius célokat fogalmaz meg, például azt, hogy "a mesterséges intelligenciát úgy kell megtervezni, hogy az emberiséget segítse".¹⁶³
- Az MSR AI honlapja szerint célja, hogy "megoldja a mesterséges intelligencia legnehezebb kihívásait" és "megvizsgálja az intelligencia alapelveit".¹⁶⁴

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

- Az MSR légi informatikával és robotikával foglalkozó mesterséges intelligencia csoportja nagy figyelmet fordít a biztonságra, de ez a szűk értelemben vett repülőgépekre vonatkozik, nem pedig az AGI-re.¹⁶⁵

Méret: közepes-nagy

¹⁵⁷ <https://blogs.microsoft.com/blog/2017/07/12/microsofts-role-intersection-ai-people-society>

¹⁵⁸ <https://www.microsoft.com/en-us/research/lab/microsoft-research-ai>

¹⁵⁹ Etherington (2017)

¹⁶⁰ <https://blogs.microsoft.com/blog/2017/07/12/microsofts-role-intersection-ai-people-society>

¹⁶¹ Architekt (dátum nélkül)

¹⁶² <https://military.microsoft.com/about>

¹⁶³ Nadella (2016)

¹⁶⁴ <https://www.microsoft.com/en-us/research/lab/microsoft-research-ai>

¹⁶⁵ <https://www.microsoft.com/en-us/research/group/air>

MLECOG

Fő weboldal: nincs

Az MLECOG egy kognitív építészeti projekt, amelyet Janusz Starzyk (Ohio University) vezet. Az MLECOG-ról szóló tanulmány a NARS-hoz és a Soarhoz hasonlónak írja le.¹⁶⁶ Az MLECOG a Motivated Learning Embodied Cognitive Architecture rövidítése.

Vezető intézmény: Ohio University

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: nincs

Katonai kapcsolat: meghatározatlan

Vezető ország:

Partnerországok: nincs

Meghatározott célok:

meghatározatlan

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: kicsi

¹⁶⁶ Starzyk és Graham (2015)

NARS

Fő weboldal: <https://sites.google.com/site/narswang>

A NARS egy AGI kutatási projekt, amelyet Pei Wang (Temple University) vezet. A NARS a Non-Axiomatic Reasoning System rövidítése, utalva arra, hogy a mesterséges intelligencia nem axiomatikus logikán, hanem előzetes tapasztalatokon alapul, összhangban a "nem elegendő tudás és erőforrások feltételezésével".¹⁶⁷ Egy 2011-es interjúban Wang azt sugallja, hogy a NARS 2021-re elérheti az emberi szintű mesterséges intelligenciát.¹⁶⁸

Vezető intézmények: Temple University

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen¹⁶⁹

Katonai kapcsolat: meghatározatlan

Vezető ország: USA

Partnerországok: nincs

Kijelölt célok: humanitárius, értelmiségi

- A NARS honlapja elmagyarázza, hogy a NARS "erkölcsileg semleges" abban az értelemben, hogy bármilyen erkölcsi rendszerrel programozható.¹⁷⁰
- A NARS honlapja hangsúlyozza, hogy a NARS-nak pozitív hatást kell gyakorolnia "az emberi társadalomra" és "emberbarátnak" kell lennie.¹⁷¹
- A NARS honlapja szerint "a kutatás végső célja az elme teljes megértése, valamint gondolkodó gépek építése".¹⁷²

A biztonsággal kapcsolatos elkötelezettség: aktív

- Wang írt a NARS biztonsági kérdéseiről, például a "motivációkezelésről", amely tényezője annak, hogy a NARS képes-e megbízhatóan követni céljait, és nem veszik-e el az irányítást.¹⁷³

Méret: közepes

¹⁶⁷ <https://sites.google.com/site/narswang/home/nars-introduction>

¹⁶⁸ Goertzel (2011)

¹⁶⁹ <https://github.com/opennars>

¹⁷⁰ <https://sites.google.com/site/narswang/EBook/Chapter5/section-5-5-education>

¹⁷¹ <https://sites.google.com/site/narswang/EBook/Chapter5/section-5-5-education>

¹⁷² <https://sites.google.com/site/narswang>

¹⁷³ Wang (2012)

Nigel

Fő weboldal: <http://kimera.ai>

Nigel a Kimera, egy Oregon állambeli Portlandben működő mesterséges intelligencia vállalat AGI-projektje. A Kimera-t 2005-ben alapította Mounir Shita és Nicholas Gilman. A cég a "The AGI Company" (Az AGI vállalat) elnevezést viseli.¹⁷⁴ A Kimera 2016-ban mutatta be a Nigelt, amely állítása szerint "az első kereskedelmi forgalomban alkalmazható mesterséges általános intelligencia technológia".¹⁷⁵ A Nigeltől azonban 2016-tól kezdve nyilvánosan kevés információ áll rendelkezésre, és a kritikusok szkeptikusak az AGI állításával kapcsolatban.¹⁷⁶ A Nigelt az Apple SIRI-jéhez és az Amazon Alexájához hasonló személyi asszisztens robotként írták le.¹⁷⁷ Kimera azt is elképzelte, hogy Nigelt különféle online tevékenységekre használják, "nagyobb átalakulást hozva a globális üzleti életben, mint akár az internet", és "az internetet az összeköttetések passzív rendszeréből proaktív, intelligens globális hálózattá alakítva".¹⁷⁸

Vezető intézmény: Kimera

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

Vezető ország: USA

Partnerországok: nincs

Kijelölt célok: humanitárius szellem

- Kimera az AGI humanitárius vízióját mutatja be, és azt írja, hogy "a mesterséges általános intelligencia képes megoldani az emberiség néhány - vagy akár az összes - legnagyobb problémáját, például a rák gyógyítását vagy a globális szegénység felszámolását".¹⁷⁹

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-kicsi

¹⁷⁴<http://kimera.ai>

¹⁷⁵ <http://kimera.ai>

¹⁷⁶ Boyle (2016); Jee (2016)

¹⁷⁷ Boyle (2016)

¹⁷⁸ <http://kimera.ai/nigel>

¹⁷⁹ <http://kimera.ai/company>

NNAISENSE

Fő weboldal: <https://nnaisense.com>

Az NNAISENSE egy Luganóban, Svájcban székhellyel rendelkező magáncég. A csapat több tagja, köztük Jürgen Schmidhuber társalapító és vezető tudós, kapcsolatban áll a Dalle Molle Mesterséges Intelligencia Intézettel (IDSIA, egy svájci nonprofit kutatóintézet). Honlapján az áll, hogy "nagyméretű neurális hálózati megoldások létrehozására törekszik az emberfeletti érzékelés és az intelligens automatizálás érdekében, végső célja pedig az általános célú mesterséges intelligenciák forgalmazása".¹⁸⁰

Vezető intézmények: NNAISENSE

Partnerintézmények: nincs

Az intézmény típusa: magáncég

Nyílt forráskódú: nem

Katonai kapcsolat:

meghatározatlan **Vezető ország:**

Partnerországok: nincs

Kijelölt célok: intellektualizmus, profit

- Schmidhuber a leírás szerint "tökéletes akadémikus", aki azért alapította a céget, hogy megakadályozza, hogy más cégek elorozzák a kutatási tehetségét; az NNAISENSE állítólag "a projekteket az alapján választja ki, hogy azok a gép tudásának javát szolgálják-e, nem pedig az alapján, hogy melyik hozza a legmagasabb díjat".¹⁸¹
- Az NNAISENSE honlapja szerint "a *marketing* végső célja" az AGI (kiemelés hozzáadva).¹⁸²

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-kicsi

¹⁸⁰ <https://nnaisense.com>

¹⁸¹ Webb (2017)

¹⁸² <https://nnaisense.com>

OpenAI

Fő weboldal: <https://openai.com>

Az OpenAI egy nonprofit AI-kutató szervezet, amelyet több neves technológiai befektető alapított. Székhelye San Franciscóban található. Támogatói 1 milliárd dollárt ajánlottak fel a projektre. Honlapja szerint: "A mesterséges általános intelligencia (AGI) lesz a legjelentősebb technológia, amelyet az ember valaha is létrehozott. Az OpenAI küldetése, hogy biztonságos AGI-t építsen, és biztosítsa, hogy az AGI előnyei a lehető legszélesebb körben és egyenletesen legyenek elosztva".¹⁸³ A szervezet a Partnership on AI to Benefit People & Society (Partnerség az emberek és a társadalom javát szolgáló mesterséges intelligenciáért) részeként működik.¹⁸⁴

Vezető intézmény: Az

intézmény típusa: nonprofit

Nyílt forráskódú: igen¹⁸⁵

Katonai kapcsolat: meghatározatlan

Vezető ország: USA

Partnerországok: nincs

Kijelölt célok: humanitárius szellem

- Az OpenAI arra törekszik, hogy az AGI "jó eredményt hozzon az emberek számára",¹⁸⁶ és hogy "az AGI előnyei a lehető legszélesebb körben és egyenletesen legyenek elosztva".¹⁸⁷

A biztonsággal kapcsolatos elkötelezettség: aktív

- A biztonságos AGI az OpenAI küldetésének része. Míg munkájának nagy részét nyíltan teszi közzé, a weboldalán az áll, hogy "hosszú távon várhatóan hivatalos eljárásokat hozunk létre a technológiák titokban tartására, ha biztonsági aggályok merülnek fel".¹⁸⁸ A DeepMinddel is együttműködik a hosszú távú AI-biztonság terén.¹⁸⁹

Méret: nagy

¹⁸³ <https://openai.com/about>

¹⁸⁴ <https://www.partnershiponai.org/partners>

¹⁸⁵ <https://github.com/openai>

¹⁸⁶ <https://openai.com/jobs>

¹⁸⁷ <https://openai.com/about>

¹⁸⁸ <https://openai.com/about>

¹⁸⁹ <https://blog.openai.com/deep-reinforcement-learning-from-human-preferences>

Valódi mesterséges intelligencia

Fő weboldal: <http://realai.org>

A Real AI egy hongkongi magáncég, amelyet Jonathan Yan vezet. Egyszemélyes társaság.¹⁹⁰ Küldetése, hogy "biztosítsa az emberiség fényes jövőjét biztonságos AGI-vel".¹⁹¹ A vállalat a biztonságos AGI stratégiáján és a mélytanulás technikai kutatásán dolgozik, ez utóbbi azon a feltételezésen alapul, hogy a mélytanulás képes az AGI-re is kiterjedni.¹⁹² Honlapja szerint "a hatékony altruizmushoz igazodunk, és célunk, hogy másoknak a lehető legnagyobb hasznára legyünk".¹⁹³

Vezető intézmény: Real AI

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

Vezető ország:

Partnerországok: nincs

Meghatározott célok:

humanitárius jellegű

A biztonsággal kapcsolatos elkötelezettség: aktív

- A Real AI egy külön oldalt szentel az AGI biztonságával kapcsolatos elképzelések felmérésének.¹⁹⁴ és saját gondolkodásának részletes megvitatását.¹⁹⁵

Méret: kicsi

¹⁹⁰<http://realai.org/about/admin>

¹⁹¹ <http://realai.org/about>

¹⁹² <http://realai.org/prosaic>

¹⁹³ <http://realai.org/about>

¹⁹⁴ <http://realai.org/safety>

¹⁹⁵ <http://realai.org/blog/towards-safe-and-beneficial-intelligence>

Agy által inspirált intelligencia kutatóközpont (RCBII)

Fő weboldal: <http://bii.ia.ac.cn>

Az RCBII a Kínai Tudományos Akadémia Automatizálási Intézete által javasolt hosszú távú stratégiai tudományos program.¹⁹⁶ A csoport székhelye Pekingben található.¹⁹⁷ Az RCBII az alapvető idegtudományok, az agyszimuláció és a mesterséges intelligencia területén végez kutatásokat. Állítása szerint "az agy által inspirált intelligencia az emberi szintű mesterséges intelligencia elérésének nagy kihívása".¹⁹⁸

Vezető intézmény: Kínai Tudományos Akadémia

Partnerintézmények: nincs

Az intézmény típusa: kormányzat

- A Kínai Tudományos Akadémia a kínai kormány alá tartozó közintézmény.¹⁹⁹

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

Vezető ország: Kijelölt

célok: értelmiségiek

- Az RCBII honlapja csak a szellemi motivációkat sorolja fel: "Az Agy által inspirált intelligenciával kapcsolatos erőfeszítések a kognitív agy megértésére és szimulálására összpontosítanak több léptékben, valamint az agy által inspirált intelligens rendszerekben való alkalmazására". Ezen alkalmazások társadalmi vagy etikai szempontjait nem tárgyalják.

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-kicsi

¹⁹⁶ <http://bii.ia.ac.cn/about.htm>

¹⁹⁷ <http://english.ia.cas.cn/au/fu/>

¹⁹⁸ <http://bii.ia.ac.cn/about.htm>

¹⁹⁹ http://english.cas.cn/about_us/introduction/201501/t20150114_135284.shtml

Sigma

Fő weboldal: <http://cogarch.ict.usc.edu>

A Sigmát Paul Rosenbloom, a Dél-kaliforniai Egyetem munkatársa vezeti. A csoport 2009 óta publikál²⁰⁰ és díjat nyert a 2011-es és 2012-es AGI-konferencián.²⁰¹ Rosenbloom korábban a Soar társproducere volt.²⁰²

Vezető intézmény: Dél-kaliforniai Egyetem

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen²⁰³

Katonai kapcsolat: igen²⁰⁴

Vezető ország:

Partnerországok: nincs

Kijelölt célok: értelmiségi

- A Sigma honlapja szerint a cél "egy kellően hatékony, funkcionálisan elegáns, általánosan kognitív, nagy, egységes kognitív architektúra kifejlesztése a virtuális emberek (és remélhetőleg intelligens ágensek/robotok - és akár az emberi megismerés egységes elméletének új formája - támogatására".²⁰⁵
- Rosenbloom egy 2013-as interjúban kozmikus nézetekre is utal: "Nem látok hosszú távon más választási lehetőséget, mint hogy meghatározzuk és vállaljuk az etikai magaslatot, még akkor is, ha ez megnyitja annak lehetőségét, hogy végül valamilyen alapvető módon felváltanak minket - vagy kiolvadunk a puszta létezésből -".²⁰⁶

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

- Egy 2013-as interjúban Rosenbloom utal arra, hogy elutasító, és megkérdőjelezi, hogy "lehetséges-e egyáltalán az emberfeletti általános intelligencia", de feltár néhány következményt is, ha lehetséges, miközben megjegyzi, hogy nincs "semmilyen különleges szakértelme" a kérdésben.²⁰⁷ A Sigma biztonsági munkájára utaló jeleket nem találtunk.

Méret: közepes

²⁰⁰ <http://cogarch.ict.usc.edu/publications-new>

²⁰¹ <http://cs.usc.edu/~rosenblo>

²⁰² <http://cs.usc.edu/~rosenblo>

²⁰³ <https://bitbucket.org/sigma-development/sigma-release/wiki/Home>

²⁰⁴ Az amerikai hadsereg, a légierő tudományos kutatási hivatala és a haditengerészeti kutatás hivatala által nyújtott finanszírozásról az R osenbloom (2013) számol be.

²⁰⁵ <http://cogarch.ict.usc.edu>

²⁰⁶ <https://intelligence.org/2013/09/25/paul-rosenbloom-interview>

²⁰⁷ <https://intelligence.org/2013/09/25/paul-rosenbloom-interview>

SiMA

Fő weboldal: <http://sima.ict.tuwien.ac.at/description>

A SiMA a Bécsi Műszaki Egyetem Dietmar Dietrich által vezetett projekt. A SiMA a Simulation of the Mental Apparatus & Applications rövidítése. A projekt célja "egy olyan széleskörű, emberhez hasonló intelligens rendszer kifejlesztése, amely nem szűken és jól körülhatárolt területeken, hanem komplex és dinamikus problémákkal képes megbirkózni".²⁰⁸ Kiterjedt figyelmet szentel a pszichoanalízisnek, különösen Freudnak és más német nyelvű tudósoknak. Dietrich 1999-ben indította el.²⁰⁹

Vezető intézmények: Bécsi Műszaki Egyetem

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen²¹⁰

- A projekt nyílt forráskódú része elavultnak tűnik.

Katonai kapcsolat: meghatározatlan

Vezető ország: Ausztria

Partnerországok: nincs

- Egy projektdokumentum kanadai, portugál, dél-afrikai és spanyolországi munkatársakat említ, de ezen együttműködések részleteit nem sikerült azonosítani.²¹¹

Kijelölt célok: értelmiségi

- Egy projektdokumentum szerint a SiMA-t azért alapították, hogy "megértsük, hogyan működik az agy egésze".²¹²
- A dokumentum kitér az automatizálásban való alkalmazásokra és arra is, hogy "a gépeknek lesznek érzései",²¹³ de ebből nem lehetett konkrét célokat azonosítani.

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes

²⁰⁸<http://sima.ict.tuwien.ac.at/description>

²⁰⁹ Brandstätter et al. (2015, V. o.)

²¹⁰http://sima.ict.tuwien.ac.at/wiki/index.php/Main_Page

²¹¹ Brandstätter et al. (2015, IV. o.)

²¹² Brandstätter et al. (2015, V. o.)

²¹³ Brandstätter et al. (2015, XI. o.)

SingularityNET

Fő weboldal: <https://singularitynet.io>

A SingularityNET egy AGI projekt, amelyet Ben Goertzel vezet. Nyilvánosan 2017-ben indult el.²¹⁴ Célja, hogy egy olyan platform legyen, amelyen bárki közzétehet mesterséges intelligencia kódot, vagy felhasználhatja a mások által közzétett mesterséges intelligencia kódot. Azt tervezi, hogy kriptovalutát használ majd az AI használatáért való fizetésre az oldalán. Ez a felállás célja, hogy az AI demokratikusabbá váljon, mint ami a kormányokon vagy vállalatokon keresztül történhetne (Goertzel 2017b). A SingularityNET azt tervezi, hogy a döntéshozatal a felhasználói közösségen belüli szavazással történik (Goertzel et al. 2017).

Vezető intézmények: SingularityNET Alapítvány

Partnerintézmények: OpenCog Alapítvány

- A SingularityNET weboldalán számos más partner is szerepel, de az OpenCog Foundation az egyetlen, amely AGI K+F-hez járul hozzá.

Az intézmény típusa: nonprofit²¹⁵

Nyílt forráskódú: igen²¹⁶

Katonai kapcsolat: meghatározatlan

Vezető ország: Kína²¹⁷

Partnerországok: USA, Ausztrália, Brazília, Kanada, Németország, Portugália, Oroszország, USA.²¹⁸

Kijelölt célok: állatjólét, ökocentrikus, humanitárius, transzhumanista

- A SingularityNET-et úgy írják le, mint "az emberekért (és a robotokért!)" és "az érző lények boldogságáért", "minden ember és minden élet számára előnyös" (Goertzel 2017b; Goertzel et al. 2017).
- A SingularityNET szintén a profitot tűzi ki célul, de ezt más célok elérésének eszközeként írja le, például kijelentve, hogy "a SingularityNET óriási profitra képes... [és] az így keletkező profitot a mesterséges intelligencia globális jóra való alkalmazására irányítja" (Goertzel 2017b).

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-kicsi

²¹⁴ A <https://blog.singularitynet.io> oldalon található blogbejegyzések 2017 októberére nyúlnak vissza.

²¹⁵ A SingularityNET Alapítványt a Goertzel et al. (2017) 8. oldalán nonprofit szervezetként írják le.

²¹⁶ <https://github.com/singnet/singnet>

²¹⁷ Goertzel vezérigazgató és David Hanson elnök egyaránt Hongkongban (<https://twitter.com/bengoertzel>; <https://twitter.com/hansonrobotics>) székel, ahol egy fejlesztői találkozó is zajlott: <https://blog.singularitynet.io/singularitynet-tech-development-update-v1-0-4de5f87b4f42>.

²¹⁸ Ezek azok az országok, amelyeket a SingularityNET munkatársai számára a <https://singularitynet.io> oldalon lehetett azonosítani: Ausztrália (Sergei Sergienko, <https://www.linkedin.com/in/sergeisergienko>); Brazília (Cassio Pennachin, <http://www.pennachin.com>); Kanada (Tal Ball, <https://www.linkedin.com/in/talgball>); Németország (Trent McConaghy,

<https://www.linkedin.com/in/trentmc>); Portugália (Mitchell Loureiro, <https://www.linkedin.com/in/mitchellloureiro>); Oroszország (Alexey Potapov, Potapov et al. 2016; Anton Kolonin, Oroszország: <https://www.linkedin.com/in/anton-kolonin-81a47836>); és USA (Eddie Monroe, <https://www.linkedin.com/in/eddiemonroe>; Linas Vepstas, <https://linas.org/resume.html>; Jim Rutt, https://en.wikipedia.org/wiki/Jim_Rutt).

SNePS (szemantikus hálózatfeldolgozó rendszer)

Fő weboldal: <http://www.cse.buffalo.edu/sneps>

A SNePS-t Stuart Shapiro vezeti a New York-i Állami Egyetemen Buffalóban, és 1969 óta publikál.²¹⁹ Honlapja szerint hosszú távú célja "az intelligens kognitív folyamatok természetének megértése olyan számítógépes kognitív ágensek kifejlesztésével és kísérletezésével, amelyek képesek természetes nyelvet használni és megérteni, érvelni, cselekedni és problémákat megoldani a legkülönbözőbb területeken".²²⁰

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

Vezető intézmények: State University of New York at Buffalo

Partnerintézmények: nincs

Intézmény típusa: egyetemi

Nyílt forráskódú: igen²²¹

Katonai kapcsolat: igen²²²

Vezető ország:

Partnerországok: nincs

Meghatározott célok:

intellektualista

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes

²¹⁹ <http://www.cse.buffalo.edu/sneps/Bibliography>

²²⁰ <http://www.cse.buffalo.edu/sneps>

²²¹ <https://github.com/SNePS/CSNePS>, <https://www.cse.buffalo.edu/sneps/Downloads>

²²² Az amerikai hadsereg kutatási hivatalának finanszírozásáról a közelmúltban megjelent tanulmányok, köztük Shapiro és Schlegel (2016) számolnak be.

Soar

Fő weboldal: <http://soar.eecs.umich.edu> és <https://soartech.com>

A Soar vezetője John Laird a Michigani Egyetemről és a szintén a michigani Ann Arborban működő SoarTech spin-off vállalat. Laird és munkatársai 1981-ben kezdték el a Soar-t.²²³ A SOAR a State, Operator Apply Result rövidítése.

Vezető intézmény: Michigan Egyetem, SoarTech

Partnerintézmények: Pace University, Pennsylvania State University, University of Portland és University of Southern California az Egyesült Államokban, University of Portsmouth az Egyesült Királyságban, valamint Bar Ilan University és Cogniteam (magántulajdonban lévő vállalat) Izraelben.²²⁴

Intézmény típusa: egyetemi, magántársaság

Nyílt forráskódú: igen²²⁵

Katonai kapcsolat: igen

- A SoarTech olyan ügyfeleket sorol fel, mint az amerikai légierő, hadsereg és haditengerészet kutatólaboratóriumai, valamint az amerikai közlekedési minisztérium.²²⁶

Vezető ország: USA

Partnerországok: Egyesült Királyság

Kijelölt célok: értelmiségi

- A Soar honlapja úgy írja le, hogy "a teljes racionalitás egy közelítőjét" vizsgálja, amelynek célja, hogy "az ember által használt kognitív képességek teljes skálájának megvalósításához szükséges összes primitív képességgel rendelkezzen".²²⁷

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

Méret: közepes-nagy

²²³ <http://ai.eecs.umich.edu/people/laird>

²²⁴ <https://soar.eecs.umich.edu/groups>

²²⁵ <https://github.com/SoarGroup>, <https://soar.eecs.umich.edu/Downloads>

²²⁶ <http://soartech.com/about>

Susaro

Fő weboldal: <http://www.susaro.com>

A Susaro egy mesterséges intelligenciával foglalkozó vállalat, amelynek székhelye Cambridge-ben, az Egyesült Királyságban található. Honlapján azt állítja, hogy lopakodó üzemmódban működik, és hogy "a világ legfejlettebb *mesterséges általános intelligencia* rendszereit tervezi" (kiemelés eredeti), olyan megközelítéssel, amely "radikálisan eltér a hagyományos mesterséges intelligenciától".²²⁸ A Susaro honlapja nem sorolja fel a személyzetet, de külső weboldalak szerint Richard Loosemore AGI-kutató vezeti.²²⁹

Vezető intézmény: Susaro

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: nem

- Susaro rendelkezik egy GitHub oldallal tartalom nélkül²³⁰

Katonai kapcsolat: meghatározatlan

Vezető ország: UK

Partnerországok: nincs

Kijelölt célok: ökocentrikus, humanitárius

- A Susaro honlapja szerint célja az "emberi és planetáris jólét előmozdítása... anélkül, hogy az embereket feleslegessé tenné".²³¹

A biztonsággal kapcsolatos elkötelezettség: aktív

- A Susaro honlapján az áll, hogy "az általunk épített rendszerekbe példátlan mértékű biztonságot építenek be... ami gyakorlatilag lehetetlenné teszi, hogy barátságtalanná váljanak".²³²

Méret: meghatározatlan

²²⁸ Minden információ és idézet a <http://www.susaro.com> honlapról származik.

²²⁹ <https://www.linkedin.com/in/richard-loosemore-47a2164>, https://www.researchgate.net/profile/Richard_Loosemore, <https://cofounderslab.com/profile/richard-loosemore>

²³⁰ <https://github.com/susarold>

²³¹ Minden információ és idézet a <http://www.susaro.com> honlapról származik.

²³² Minden információ és idézet a <http://www.susaro.com> honlapról származik.

Tencent AI Lab (TAIL)

Fő weboldal: <http://ai.tencent.com/ailab>

A TAIL a Tencent, a sencseni székhelyű kínai technológiai vállalat AI-csoportja. Honlapján számos kutatási területet sorol fel, amelyek közül az egyik a gépi tanulás, amely a saját állítása szerint az "általános mesterséges intelligenciát" is magában foglalja.²³³ Tong Zhang, a TAIL igazgatója azt írja, hogy a TAIL "nemcsak a mesterséges általános intelligencia területén fejleszti a tudomány jelenlegi állását, hanem a vállalat termékeit is támogatja".²³⁴

Vezető intézmény: Tencent

Partnerintézmények: nincs

Az intézmény típusa: állami vállalat

- A Tencent honlapján az áll, hogy "a Tencent a Tencent Cloudon keresztül nyílt forráskódúvá teszi a kép, a hang és a biztonság területén működő AI-megoldásait partnerei számára", de azt nem állítja, hogy az AGI-kutatásai nyílt forráskódúak lennének.²³⁵

Nyílt forráskódú: nem

- Tencent kiadja néhány munka nyílt forráskódú,²³⁶ de nem az AGI

Katonai kapcsolat: meghatározatlan

Vezető ország: Kína

Partnerországok:

- A TAIL nemrégiben nyitott irodát Seattle-ben.²³⁷

Kijelölt célok: meghatározatlan A

biztonsággal kapcsolatos

elkötelezettség: meghatározatlan

Méret: közepes-kicsi

²³³ <http://ai.tencent.com/ailab>

²³⁴ <http://tongzhang-ml.org/research.html>

²³⁵ <http://ai.tencent.com/ailab>
²³⁶ <https://github.com/Tencent>
²³⁷ Mannes (2017)

Uber AI Labs (UAIL)

Fő weboldal: <https://www.uber.com/info/ailabs>

Az UAIL az Uber mesterséges intelligenciával foglalkozó kutatási részlege. Az UAIL 2016-ban indult a Geometric Intelligence (Temperton 2016) felvásárlásával, egy magánvállalkozással, amelyet Gary Marcus, Kenneth Stanley és Zoubin Ghahramani alapított 2014-ben a NYU inkubációs támogatásával.²³⁸ A Geometric Intelligence Marcus AGI-vel kapcsolatos elképzelésein alapult, különösen azon, hogy hogyan lehet "kevesebb képzési adattal tanulni", mint amennyire a mélytanuláshoz szükség van (Chen 2017). Marcus azóta elhagyta az UAIL-t (Chen 2017). Az Uber általi felvásárláskor a Geometric Intelligence munkatársai San Franciscóba költöztek, kivéve Ghahramanit, aki az Egyesült Királyságban, Cambridge-ben maradt (Metz 2016). Az UAIL a hírek szerint része az Uber azon kísérletének, hogy jelentősen túlterjeszkedjen a magántaxi-piacon, hasonlóan ahhoz, ahogyan az Amazon terjeszkedett a könyveken túl (Metz 2016). Ghahramani szerint a mesterséges intelligencia egyesíti "a szabályalapú tanulás néhány ötletét a statisztikai tanulás és a mélytanulás ötleteivel" (Metz 2016).

Vezető intézmények: Uber

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: nem

- Az Uber rendelkezik némi nyílt forráskódú mesterséges intelligenciával,²³⁹ , de úgy tűnik, hogy ez nem tartalmazza az AGI-t.

Katonai kapcsolat: meghatározatlan

- Úgy tűnik, hogy az Ubernek nincsenek védelmi szerződésai.²⁴⁰

Vezető ország:

Partnerországok: USA:

Kijelölt célok:

humanitárius

- Az UAIL honlapja szerint "világszerte emberek millióinak életét kívánja javítani".²⁴¹

A biztonsággal kapcsolatos kötelezettségvállalás: nem meghatározott

- Az UAIL biztonságával kapcsolatban nem találtunk vitát, kivéve az Uber járműveinek biztonságával kapcsolatban.²⁴²
- Marcus jelezte, hogy támogatja a mesterséges intelligencia etikai kutatását mindaddig, amíg megértik, hogy a fejlett AGI nem áll a küszöbön.²⁴³

Méret: közepes

- ²³⁸ <https://www.nyu.edu/about/news-publications/news/2016/december/nyu-incubated-start-up-geometric-intelligence-acquired-by-uber.html>
- ²³⁹ <https://github.com/uber/tensorflow>
- ²⁴⁰ [https://www.fpds.gov/ezsearch/fpdsportal?q=uber+DEPARTMENT_FULL_NAME%3A "DEPT+OF+DEFENSE"](https://www.fpds.gov/ezsearch/fpdsportal?q=uber+DEPARTMENT_FULL_NAME%3A%20DEPT+OF+DEFENSE)
- ²⁴¹ <https://www.uber.com/info/ailabs>
- ²⁴² Chamberlain (2016)
- ²⁴³ <https://techcrunch.com/2017/04/01/discussing-the-limits-of-artificial-intelligence>

Vicarious

Fő weboldal: <https://www.vicarious.com>

A Vicarious egy magántulajdonban lévő mesterséges intelligenciával foglalkozó vállalat, amelyet 2010-ben Scott Phoenix és Dileep George alapított San Franciscóban. Több tízmillió dolláros befektetést szerzett több neves befektetőtől.²⁴⁴ Állítása szerint "olyan rendszereket épít, amelyek emberhez hasonló intelligenciát hoznak a robotok világába".²⁴⁵ Egy interjúban Phoenix azt mondja, hogy a Vicarious az AGI felé dolgozik, és "rengeteg értéket teremt közben",²⁴⁶ és az AGI "gyakorlatilag az utolsó találmány lenne, amit az emberiség valaha is fog csinálni".²⁴⁷

Vezető intézmény: Vicarious

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: igen²⁴⁸

Katonai kapcsolat: meghatározatlan

Vezető ország: USA

Partnerországok: nincs

Kijelölt célok: humanitárius szellem

- A Vicarious egy rugalmas célú társaság, állítólag azért, hogy "a profit helyett a társadalmi haszon maximalizálására törekedhessen".²⁴⁹ Scott Phoenix szerint a Vicarious célja, hogy a mesterséges intelligenciát "az emberiség boldogulásának elősegítésére" fejlessze.²⁵⁰

A biztonsággal kapcsolatos elkötelezettség: mérsékelt

- Phoenix szerint a mesterséges intelligencia biztonságára szükség lehet "valamikor a jövőben... de a kutatás most még nagyon korai stádiumban van",²⁵¹ és hogy ez nem lesz nehéz, mert a mesterséges intelligencia valószínűleg "elég okos lesz ahhoz, hogy rájöjjön, mit is akartunk, hogy csináljon".²⁵²

Méret: közepes-nagy

²⁴⁴ Cutler (2014); High (2016)

²⁴⁵ <https://www.vicarious.com/research.html>

²⁴⁶ Magas (2016)

²⁴⁷ TWiStartups (2016)

²⁴⁸ <https://github.com/vicariousinc>

²⁴⁹ Magas (2016)

²⁵⁰ Magas (2016)

²⁵¹ Legjobb (2016)

Victor

Fő weboldal: <http://2ai.org/victor>

A Victor a 2AI fő projektje, amely a Cifer Inc. magáncég leányvállalata.²⁵³ A 2AI-t Timothy Barber és Mark Changizi vezeti, és az idahói Boise-ban, valamint az Amerikai Virgin-szigeteken található címetek jegyzi.²⁵⁴ A 2AI a Victort AGI-projektként írja le.²⁵⁵ Honlapja szerint "úgy gondoljuk, hogy a mesterséges intelligencia jövője végső soron a versenyképes interakcióra való képességén múlik majd".²⁵⁶

Vezető intézmények: Cifer

Partnerintézmények: nincs

Az intézmény típusa: magánvállalkozás

Nyílt forráskódú: nem

Katonai kapcsolat: meghatározatlan

Vezető ország: USA

Partnerországok: nincs

Meghatározott célok:

ökocentrikus

- A 2AI honlapja szerint "a 2AI határozottan támogatja a vízi ökoszisztémák, különösen a zátonyok és a tenger alatti hegyek védelmét és megőrzését célzó megoldásokat, amelyek a Föld biomasszájának és biológiai sokféleségének kulcsfontosságú kapcsolódási pontjai".²⁵⁷

A biztonsággal kapcsolatos elkötelezettség: elutasító

- A projekt honlapja szerint a mesterséges intelligencia katasztrófa-forgatókönyvek "örültségek", mivel a mesterséges intelligenciának szüksége lesz az emberekre, hogy fenntartsa a fizikai eszközöket, amelyekeken léteznek, és így az emberek gondoskodni fognak ezen eszközök fenntartásáról és bővítéséről.²⁵⁸

Méret: kicsi

²⁵³<http://2ai.org/legal>

²⁵⁴ <http://2ai.org/legal>

²⁵⁵ <http://2ai.org/landscape>

²⁵⁶ <http://2ai.org/victor>

²⁵⁷ <http://2ai.org/legal>

²⁵⁸ <http://www.2ai.org/killera>

A teljes agyi architektúra kezdeményezés (WBAI)

Fő weboldal: <https://wba-initiative.org>

A WBAI egy tokiói nonprofit szervezet, amelyet Hiroshi Yamakawa vezet. Yamakawa a Dwango AI igazgatója, emellett a Tamagawa Egyetem és a Japán Mesterséges Intelligencia Társaság tagja. A WBAI küldetése "egy emberhez hasonló mesterséges általános intelligencia (AGI) létrehozása (megtervezése) a teljes agy felépítéséből való tanulással".²⁵⁹ Céljuk, hogy "a mesterséges általános intelligencia (AGI) felépítése 2030 körül meghaladja az emberi agy képességeit".²⁶⁰ A WBAI-t többek között a Panasonic, a Toshiba és a Toyota támogatja.²⁶¹

Vezető intézmény: Whole Brain Architecture Initiative

Partnerintézmények: nincs

Intézmény típusa: nonprofit

Nyílt forráskódú: igen²⁶²

Katonai kapcsolat: meghatározatlan

Vezető ország: Japán

Partnerországok: nincs

Kijelölt célok: humanitárius szellem

- A WBAI olyan mesterséges intelligencia-fejlesztést támogat, amely "a legjobb az emberi társadalom számára".²⁶³ A WBAI-ról szóló diavetítésben továbbá Yamakawát idézik, aki szerint "az EcSIA [a megosztott intelligens ágensek ökoszisztémája] által biztosított kegyet és gazdagságot megfelelően el kell osztani mindenki között".²⁶⁴

A biztonsággal kapcsolatos elkötelezettség: aktív

- A biztonság a WBAI egyik fontos témája. A honlapjukon például ez áll: "Az intelligencia mennyiségi vonatkozásaiban az AGI felül fogja múlni az embereket. Ha az AGI heterogén lesz számunkra, akkor nehéz lehet számunkra megérteni, a társadalom tagjaként etikailag beágyazódni és az irányítást fenntartani. Így azt mondhatjuk, hogy viszonylag biztonságos választás, ha az első AGI-t hozzánk hasonló formában építjük meg".²⁶⁵ Ebből az következik, hogy a WBAI részben azért igyekszik agyhoz hasonló AGI-t építeni, mert az biztonságosabb lenne.

Méret: közepes-kicsi

²⁵⁹ <https://wba-initiative.org/en/about/greetings>

²⁶⁰ <https://wba-initiative.org/en/wba>

²⁶¹ <https://wba-initiative.org/en/supporting-members>

²⁶² <https://github.com/wbap>

²⁶³ <https://wba-initiative.org/en/about/vision>

²⁶⁴ <https://www.slideshare.net/HiroshiYamakawa/2017-0512gatsby-wsv10b-75941913>

²⁶⁵ <https://wba-initiative.org/en/2071>

2. függelék. Egyéb figyelemre méltó projektek

4CAPS

Fő weboldal: http://www.ccbi.cmu.edu/projects_4caps.html

A 4CAPS-t Marcel Just pszichológus vezette a Carnegie Mellon Egyetemről. Ez "egy számítógépes idegtudományi modell és egy szimbolikus mesterséges intelligencia rendszer hibridje" (Goertzel 2014, 19. o.). "Mind a hagyományos viselkedési adatokat, mind pedig - ami még érdekesebb - a neuroimaging vizsgálatok eredményeit képes figyelembe venni".²⁶⁶ A projekt a tengerészeti kutatási hivatal és az egyetemi kutatási kezdeményezés multidiszciplináris kutatási programjának finanszírozásáról számol be.²⁶⁷

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

4D/RCS (Valós idejű vezérlőrendszerek architektúrája)

Fő weboldal: <https://www.nist.gov/intelligent-systems-division/rcs-real-time-control-systems-építészet>

A 4D/RCS-t James Albus vezette az amerikai Nemzeti Szabványügyi és Technológiai Intézetben. "Keményen bedrótozott architektúrából és algoritmusokból áll... tanulással kiegészítve" (Goertzel 2014, 24. o.).

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Achler

Fő weboldal:

A Los Alamos National Labs Tsvi Achler által vezetett, meg nem nevezett projekt a neurális hálózatokat használta "a szimbolikus-szubszimbolikus szakadék áthidalásának újszerű megközelítésében" Goertzel (2014, 17. o.).

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

AGINAO

Fő weboldal: <http://aginao.com>

²⁶⁶ http://www.ccbi.cmu.edu/projects_4caps.html

²⁶⁷ http://www.ccbi.cmu.edu/projects_4caps.html

Az AGINAO a lengyelországi Gdanskban élő Wojciech Skaba projektje volt. A projekt 2011-2013 között működött,²⁶⁸ és az utóbbi időben nem mutat aktivitást.

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

Alibaba

Fő weboldal:

Az Alibaba aktívan tevékenykedik a mesterséges intelligencia területén, de az AGI-re utaló jeleket nem találtunk.

A megfontolás oka: Az Alibaba egy jelentős számítástechnikai vállalat.

A kizárás oka: Nem találtak AGI-projektekre utaló jeleket

Amazon

Fő weboldal: <https://aws.amazon.com/amazon-ai>

Az Amazon rendelkezik egy AI-csoporttal az Amazon Web Services (AWS) részlegén belül, de úgy tűnik, hogy ez nem foglalkozik AGI-vel. Az Amazon AWS-erőforrásokat adományozott az OpenAI-nak.²⁶⁹

A megfontolás oka: Az Amazon egy jelentős számítástechnikai vállalat

A kizárás oka: Nem találtak AGI-projektekre utaló jeleket

Apple

Fő weboldal: <https://machinelearning.apple.com>

Az Apple-nek van egy AI-csoportja, amely a jelek szerint nem dolgozik az AGI-n. Az Apple azonban a titkolózás hírében áll, és csak minimális weboldala van. Az Apple állítólag azért rendelkezik kevésbé alkalmas AI-vel, mint az olyan vállalatok, mint a Google és a Microsoft, mert az Apple szigorúbb adatvédelmi szabályokat alkalmaz, és megtagadja magától az AI képzéséhez használt adatokat.²⁷⁰ Hasonlóképpen, a mesterséges intelligencia kutatásainak legalábbis egy része korlátozott adatokból vagy szintetikus adatokból való tanulásra irányulhat.²⁷¹ A nemrégiben felvásárolt mesterséges intelligenciával foglalkozó vállalatai szűk értelemben vett mesterséges intelligenciára irányulnak.²⁷² Bár lehetséges, hogy az Apple AGI-n dolgozik, az online keresések során nem találtunk erre utaló jeleket.

A megfontolás oka: Az Apple egy jelentős számítástechnikai vállalat

A kizárás oka: Nem találtak AGI-projektekre utaló jeleket

²⁶⁸ Például Skaba (2012a; 2012b); <http://aginao.com/page2.php>.

²⁶⁹ <https://blog.openai.com/infrastructure-for-deep-learning>

²⁷⁰ Vanian (2017a)

²⁷¹ Vanian (2017b)

²⁷² Tamturk (2017)

Mesterséges agy laboratórium

Fő weboldal:

A Mesterséges Agy Laboratóriumot Hugo de Garis vezette a Xiamen Egyetemen. A projekt a jelek szerint befejeződött, amikor de Garis 2010 körül nyugdíjba vonult. A Xiamen Egyetemen most egy egyszerű intelligens robotrendszerekkel foglalkozó csoport működik,²⁷³ de ez nem feltétlenül kapcsolódik az ABL-hez.

A megfontolás oka: A szerző előzetes ismeretei

A kizárás oka: Nyilvánvalóan inaktív

Agyi képalkotó és modellező szekció (BIMS)

Fő weboldal: <https://www.nidcd.nih.gov/research/labs/brain-imaging-and-modeling-section>

A BIMS egy kutatási projekt, amelyet Barry Horwitz, az Egyesült Államok Nemzeti Egészségügyi Intézeteinek munkatársa vezet. A projekt az agyi képalkotást számítógépes modellezéssel kombinálja az alapvető idegtudományok és az agyi rendellenességek kezelésének előmozdítása érdekében.

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: A projekt az alapvető idegtudományokra összpontosított, nem pedig az AGI kifejlesztésére.

BRAIN kezdeményezés

Fő weboldal: <https://www.braininitiative.nih.gov>

A BRAIN Initiative egy kutatási projekt, amelynek célja az emberi agy megértése. A BRAIN a Brain Research through Advancing Innovative Neurotechnologies (Agykutatás az innovatív neurotechnológiák előmozdításával) rövidítése. A projekt székhelye az amerikai Nemzeti Egészségügyi Intézeteknél található, és számos más amerikai kormányzati szervezettel és magánszervezettel működik együtt.²⁷⁴ A honlapja szerint "az innovatív technológiák fejlesztésének és alkalmazásának felgyorsításával a kutatók képesek lesznek egy forradalmian új, dinamikus képet alkotni az agyról, amely először mutatja meg, hogy az egyes sejtek és az összetett idegpályák hogyan hatnak egymásra térben és időben".²⁷⁵

A megfontolás oka: Nagyszabású agykutatási projekt, hasonlóan a Blue Brainhez.

A kizárás oka: A projekt az alapvető idegtudományokra összpontosított, nem pedig az AGI kifejlesztésére.

Agy/MINDS

²⁷³ <http://information.xmu.edu.cn/en/?mod=departments&id=31>

²⁷⁴ <https://www.braininitiative.nih.gov/about/index.htm>

²⁷⁵ <https://www.braininitiative.nih.gov>

Fő weboldal: <http://brainminds.jp/en>

A Brain/MINDS egy idegtudományi kutatási projekt. A Brain/MINDS a Brain Mapping by Integrated Neurotechnologies for Disease Studies rövidítése. A projektet a japán Oktatási, Kulturális, Sport, Tudományos és Technológiai Minisztérium (MEXT) támogatja. A projekt a főemlősök agyának vizsgálatára, az agyi rendellenességek idegi hálózataira, valamint az alap- és klinikai idegtudományok közötti együttműködés javítására összpontosít.²⁷⁶

A megfontolás oka: Nagyszabású agykutatási projekt, hasonlóan a Blue Brainhez.

A kizárás oka: A projekt az alapvető idegtudományokra összpontosított, nem pedig az AGI kifejlesztésére.

Szénmásolatok

Fő weboldal: <https://www.carboncopies.org>

A Carboncopies egy San Francisco-i székhelyű nonprofit szervezet, amely "támogatást nyújt a Whole brain emulációval kapcsolatos területeken dolgozó tudósoknak".²⁷⁷

A megfontolás oka: Az AGI technikai részleteire összpontosító kutatási projekt.

A kizárás oka: A projekt az AGI K+F támogatására összpontosított, nem pedig a K+F végrehajtására.

CERA-CRANIUM

Fő weboldal: <https://www.carboncopies.org>

A CERA-CRANIUM egy kognitív architektúra projekt volt, amelyet néha az AGI kontextusában vitattak meg.²⁷⁸ Vezetője Raúl Arrabales volt a Madridi Egyetemről. Számítógépes játékokhoz használták, 2010-ben megnyert egy versenyt.²⁷⁹ A CERA-CRANIUMmal kapcsolatos legfrissebb azonosított munka 2013-ból származik.²⁸⁰

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

CHAI (Center for Human-Compatible AI)

Fő weboldal: <http://humancompatible.ai>

A CHAI a Kaliforniai Egyetemen (Berkeley) működő kutatócsoport. Honlapján azt írja, hogy célja "olyan koncepcionális és technikai eszközök kifejlesztése, amelyek segítségével az AI-kutatás általános irányvonalát a bizonyíthatóan előnyös rendszerek felé lehet átirányítani", különösen az olyan gépek vonatkozásában, amelyek "nagyobb mértékben képesek

²⁷⁶ <http://brainminds.jp/en/overview/greeting>

²⁷⁷ <https://www.carboncopies.org/mission>

²⁷⁸ Például Ng et al. (2017)

²⁷⁹ Arrabales és Muñoz (2010)

²⁸⁰ Arrabales et al. (2013)

mint az emberek, a célok és környezetek széles skáláján", amit valószínűnek tart, hogy előbb-utóbb létezni fog.²⁸¹

A megfontolás oka: Az AGI technikai részleteire összpontosító kutatási projekt.

A kizárás oka: A projekt az AGI biztonsági szempontjaira összpontosított, nem pedig az AGI fejlesztésére.

CHREST

Fő weboldal: <http://chrest.info>

A CHREST vezetője Fernand Gobet a Liverpooli Egyetemről. Gobet 1992-ben kezdte el, és az 1959-es EPAM rendszerre vezethető vissza.²⁸² A CHREST a Chunk Hierarchy and REtrieval STRuctures rövidítése. Ez "egy olyan kognitív architektúra, amely modellezi az emberi észlelést, tanulást, memóriát és problémamegoldást".²⁸³ A CHREST-ről szóló tanulmány úgy írja le a kategorizálás és a megértés terén mutatkozó erősségeit, mint amelyek kiegészítik más projektek (pl. ACT-R, Soar) problémamegoldásban mutatkozó erősségeit.²⁸⁴ A honlapon 2010 óta nem szerepelnek publikációk.²⁸⁵ és 2012 óta nincs frissítés.²⁸⁶

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

Kognitív számítástechnikai projekt

Fő weboldal: <http://research.ibm.com/cognitive-computing>, különösen a <https://www.research.ibm.com/cognitive-computing/neurosynaptic-chips.shtml>

A CCP az IBM mesterséges intelligencia projektjeinek része, amelyek közé tartozik a híres Watson rendszer is. Goertzel (2014) egy Dharmendra Modha által vezetett projektet tárgyal, amelynek célja az emberi agy mintájára számítógépes hardver- és szoftverrendszerek létrehozása. A projekt egy új programozási nyelvet és egy TrueNorth nevű új számítógépes chipet hozott létre, amelyet Modha "a számítástechnika történetének fordulópontjaként" aposztrofál.²⁸⁷ A chipet a *Science* 2014-es cikkében mutatták be.²⁸⁸ A chip fejlesztését a DARPA SyNAPSE programja támogatta, amelynek célja "biológiai szintre skálázható, alacsony fogyasztású elektronikus neuromorfikus számítógépek" előállítása volt.²⁸⁹

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: A projekt az AGI-hez kapcsolódó hardverfejlesztésre összpontosított, nem pedig magára az AGI-re.

²⁸¹ <http://humancompatible.ai/about>

²⁸² <http://www.chrest.info/history.html>

²⁸³ <http://www.chrest.info>

²⁸⁴ Lane és Gobet (2012)

²⁸⁵ <http://chrest.info/publications.html>

²⁸⁶ <http://chrest.info>

²⁸⁷ <http://www.research.ibm.com/articles/brain-chip.shtml>

²⁸⁸ Merolla et al. (2014)

²⁸⁹ <http://www.darpa.mil/program/systems-of-neuromorphic-adaptive-plastic-scalable-electronics>

Kognitív rendszerek eszköztára (CST)

Fő weboldal: <http://cst.fee.unicamp.br>

A CST a braziliai Campinas-i Egyetem Ricardo Gudwin által vezetett projekt. Ez "egy Java-alapú eszközkészlet, amely lehetővé teszi a kognitív architektúrák létrehozását".²⁹⁰

A megfontolás oka: Az AGI technikai vonatkozásaihoz kapcsolódó projekt

A kizárás oka: A projekt az AGI fejlesztésére használható eszközökre összpontosított, nem pedig az AGI létrehozására.

Comirit

Fő weboldal: <http://www.comirit.com>

A Comirit a Sydneyi Műszaki Egyetem Benjamin Johnston projektje volt. Célja "józan intelligenciával rendelkező robot- és szoftverrendszerek" létrehozása volt, rövid távú, "inkább hetekre vagy hónapokra, mint évtizedekre" koncentrálva, de "olyan rendszerek létrehozásának hosszú távú céljai inspirálták, amelyek mélyen emberi módon értik meg a valós világot", és így olyan terveket követett, amelyek "fokozatosan továbbfejleszthetők, és így egyre alkalmasabb rendszerekké válhatnak".²⁹¹ Főleg 2010-2011 körül volt aktív.

A megfontolás oka: K+F projekt AGI-törekvésekkel

A kizárás oka: Nyilvánvalóan inaktív

Dav & SAIL

Fő weboldal: <http://www.cse.msu.edu/~weng/research/LM.html>

A Dav és a SAIL a Michigani Állami Egyetem Juyang Weng projektje volt. A két robotot úgy tervezték, hogy úgy tanuljanak, mint az emberi gyerekek. A projekt célja "a gép emberi szintű teljesítményének elérése autonóm fejlesztéssel" volt.²⁹² A munka a Nemzeti Tudományos Alapítvány és a DARPA támogatását élvezte.²⁹³

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

DUAL

Fő weboldal: <http://alexpetrov.com/proj/dual>

²⁹⁰ <http://cst.fee.unicamp.br>

²⁹¹ <http://www.comirit.com>

²⁹² <http://www.cse.msu.edu/~weng/research/LM.html>

A DUAL-t Boicho Kokinov vezette az Új Bolgár Egyetemen. Körülbelül 1999 és 2005 között működött. A program Marvin Minsky *Society of Mind* című *elmetársadalmán* alapult, amelyben az elmék egymással kölcsönhatásban álló al-elme "ágensekből" állnak. A DUAL integrálja a szimbolikus és az emergencialista megközelítéseket, és integrálja a deklaratív tanulást (az olyan információkról való tanulás, amelyekről könnyen beszélhetünk) és a procedurális tanulást (a tanulás, amely inkább szokásszerű és nehezebb róla beszélni).²⁹⁴

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Entropica

Fő weboldal: <http://entropica.com>

Az Entropica egy AGI-projekt és egy magánvállalkozás, amelyet Alexander Wissner-Gross vezet. Úgy tűnik, hogy 2013-ban alapították.²⁹⁵ Az AGI-projektek 2017-es listáján szerepelt.²⁹⁶ Azonban 2013 óta nem találtunk semmilyen tevékenységet, és 2017 augusztusában Wissner-Gross nem szerepeltette az Entropicát a²⁹⁷ weboldalán vagy az önéletrajzában található cégek listáján.²⁹⁸ Az Entropica Wissner-Gross ugyanabban az évben egy kutatási tanulmányban közzétett ötletein alapul.²⁹⁹ A projektet bemutató videó szerint a projekt "széles körben alkalmazható számos területen", és több, látszólag különböző területen mutatja be működését.³⁰⁰ A videó az egyetlen tartalom a projekt weboldalán. A médiatudósítások az AGI és a szuperintelligencia felé való áttörésként írták le,³⁰¹ de más AI-kutatók kritikusan nyilatkoztak³⁰² és néhány megfigyelő átverésre gyanakszik.³⁰³

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

Einstein

Fő weboldal: <https://www.salesforce.com/products/einstein>

Az Einstein a Salesforce egyik projektje, amely a mesterséges intelligenciát alkalmazza az ügyfélszolgálati üzletágban. Az Einstein a MetaMind magáncégből nőtt ki, amelyet a Salesforce 2016-ban vásárolt fel.³⁰⁴ Richard Socher vezető tudós állítólag arra törekszik, hogy AGI-t építsen.³⁰⁵ Nem találtunk azonban arra utaló jeleket, hogy az Einstein AGI-n dolgozna.

²⁹⁴<http://alexpetrov.com/proj/dual>

²⁹⁵ <https://www.ventureradar.com/organisation/Entropica/ff1cc5d7-f82a-4a9c-b951-cfd9747ff310>

²⁹⁶<http://2ai.org/landscape>

²⁹⁷ <http://www.alexwg.org/companies>

²⁹⁸<http://www.alexwg.org/AWG-CV.pdf>

²⁹⁹ Wissner-Gross és Freer (2013)

³⁰⁰<https://www.youtube.com/watch?v=cT8ZqChv8P0>

³⁰¹ Dvorsky (2013)

³⁰² Marcus és Davis (2013)

³⁰³ <https://www.quora.com/How-can-we-prove-that-Entropica-is-a-hoax>

³⁰⁴ Novet (2016)

³⁰⁵ The Economist (2016)

A megfontolás oka: Egy olyan személy által vezetett mesterséges intelligencia K+F projekt, aki az AGI létrehozására törekszik.

A kizárás oka: Nem találtak AGI K+F-et

EPIC

Fő weboldal:

Az EPIC-et David Kieras informatikus vezette a Michigani Egyetemen. Goertzel (2014, 16. o.) azt írja, hogy "a SOAR-hoz kapcsolódott a problémamegoldás, a tervezés és a tanulás terén".

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

EvoGrid

Fő weboldal: <http://www.evogrid.org>

Az EvoGrid egy nyílt forráskódú mesterséges élet projekt volt, amelyet Bruce Damer kezdeményezett. A mesterséges élet számítástechnikai kihívását a SETI@home projekthez hasonló, elosztott számítógépes hardverhálózathoz való hozzáféréssel próbálta megoldani. A projekt honlapja 2010 óta nem tartalmaz frissítéseket, bár Damer munkája a mesterséges étellel kapcsolatban folytatódik.³⁰⁶

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

GLAIR

Fő weboldal: nem található

A GLAIR a buffalói New York Állami Egyetem Stuart Shapiro projektje volt. Körülbelül 1993-2013 között működött.³⁰⁷ A GLAIR célja "az emberi szintű intelligens viselkedés számítógépes megértése és megvalósítása" volt.³⁰⁸

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

GMU BICA

³⁰⁶ <http://www.damer.com>

³⁰⁷ <https://www.cse.buffalo.edu/~shapiro/Papers>

³⁰⁸ Shapiro és Bona (2010, 307. o.)

Fő weboldal: <http://mason.gmu.edu/~asamsono/bica.html>

A GMU Bica a George Mason Egyetem Alekszej Szamszonovics projektje volt. A projekt 2006-2007 körül volt aktív. Weboldalán "általános kognitív architektúráként" írják le, amely az öntudatosságon alapul.³⁰⁹

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

Goedel gép

Fő weboldal: <http://people.idsia.ch/~juergen/goedelmachine.html>

A Goedel Machine a svájci Dalle Molle Mesterséges Intelligencia Kutatóintézet Jürgen Schmidhuber projektje volt. A Goedel-gép úgy halad előre, hogy minden egyes lépésben a legjobbnak bizonyuló cselekvést hajtja végre, ami végtelen számítási teljesítményt igényel (Goertzel 2014, 25. o.). Schmidhuber azt írja a honlapján, hogy "nagyjából 15 éves kora óta Jürgen Schmidhuber professzor fő célja az volt, hogy egy önmagánál okosabb, önfejlesztő mesterséges intelligenciát (AI) építsen, majd visszavonuljon".³¹⁰

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: A projekt az AGI elméleti aspektusaira összpontosított, nem pedig az AGI kifejlesztésére.

Google Brain

Fő weboldal: <https://research.google.com/teams/brain>

A Google Brain a Google mesterséges intelligencia kutatócsoportja. Kutatói együttműködtek a DeepMinddel az AGI-kutatásban,³¹¹ de munkája főként a mélytanulásra összpontosít.³¹²

A megfontolás oka: Az AGI-val kapcsolatban álló kutatócsoport

A kizárás oka: Nem eléggé az AGI-ra összpontosít

HUMANOBS

Fő weboldal: <http://www.humanobs.org>

A HUMANOBS egy olyan robotok fejlesztésére irányuló projekt volt, amelyek "képesek megtanulni a szociális interakciót", ami a közlemény szerint "nagy lépés a végső cél felé, hogy olyan intelligenciát hozzunk létre, amely egyszerre önellátó és alkalmazkodó a legkülönbözőbb környezetekben". A projekt 2009 és 2014 között volt aktív, és több AGI-projektben is részt vett.

³⁰⁹ <http://mason.gmu.edu/~asamsono/bica.html>

³¹⁰ <http://people.idsia.ch/~juergen>

³¹¹ Fernando et al. (2017)

³¹² <https://research.google.com/teams/brain/about.html>

konferenciák. A programot izlandi, olaszországi, spanyolországi, spanyolországi, svájci és egyesült királyságbeli egyetemek és kutatóintézetek konzorciuma működtette.³¹³

A megfontolás oka: Egy korábbi AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

IM-CLEVER

Fő weboldal: <http://www.im-clever.eu>

Az IM-CLEVER egy olyan projekt volt, amelynek célja olyan robotok tervezése volt, amelyek képesek önállóan tanulni és tudásukat különböző kontextusokban alkalmazni. Körülbelül 2013 óta nem működik. A projektet Gianluca Baldassarre vezette az olaszországi Istituto di Scienze e Technologie della Cognizione (Kognitív Tudományok és Technológiák Intézete) munkatársai Európa-szerte és egy amerikai csoporttal együtt.³¹⁴ A csoport az "iCub" nevű robotikai platformon dolgozott olyan robotok számára, amelyek képesek új készségeket megtanulni és azokat különböző feladatokra alkalmazni.³¹⁵

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Intel

Fő weboldal: <https://intel.com> (nem találtak külön AI-weboldalt)

Az Intel aktívan részt vesz a mesterséges intelligencia kutatásában.³¹⁶ Számos AI-vállalatot vásárolt fel.³¹⁷ Azonban nem látható, hogy az AGI-re összpontosítana.

A megfontolás oka: Az Intel egy jelentős számítástechnikai vállalat

A kizárás oka: Nem találtak AGI-projektekre utaló jeleket

Israel Brain Technologies (IBT)

Fő weboldal: <http://israelbrain.org>

Az IBT egy idegtudományi kutatási projekt. Ez egy izraeli nonprofit szervezet, amelynek székhelye Ramat HaSharonban van. Az izraeli kormánytól, filantrópoktól és vállalatoktól kap finanszírozást.³¹⁸ Küldetése, hogy "felgyorsítsa az agyi betegségek innovatív kezeléseinek és gyógymódjainak kifejlesztését".³¹⁹

³¹³ <http://www.humanobs.org>

³¹⁴ <http://www.im-clever.eu/homepage/project/partners/partners>

³¹⁵ <http://www.im-clever.eu>

³¹⁶ <https://www.intel.com/content/www/us/en/analytics/artificial-intelligence/overview.html>

³¹⁷ Tamturk (2017)

³¹⁸ <http://israelbrain.org/donate>

³¹⁹ <http://israelbrain.org/about-us/mission>

A megfontolás oka: Nagyszabású agykutatási projekt, hasonlóan a Blue Brainhez.

A kizárás oka: A projekt az alapvető idegtudományokra összpontosított, nem pedig az AGI kifejlesztésére.

Kuipers csoport

Fő weboldal: <https://www.cs.utexas.edu/users/qr/robotics/bootstrap-learning.html>

Benjamin Kuipers, a Texasi Egyetem munkatársa egy olyan csoportot vezetett, amely olyan robotokat fejlesztett ki, amelyek a világról szóló információk széles skáláját saját maguk tanulják meg saját tapasztalataikból. Az 1997-2007 közötti időszakban publikáltak.

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Az emlősök talamokortikális rendszereinek nagyszabású modellje (LSMMTS)

Fő weboldal: https://www.izhikevich.org/publications/large-scale_model_of_human_brain.htm

Az LSMMTS Eugene Izhikevich és Gerald Edelman (The Neurosciences Institute) projektje volt. Az LSMMTS arról nevezetes, hogy egy agyszimuláció "magához a teljes emberi agyhoz hasonló méretű" (Goertzel 2014, 19. o.).

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Gépi Intelligencia Kutatóintézet (MIRI)

Fő weboldal: <https://intelligence.org>

A MIRI egy független nonprofit kutatócsoport, amely "az embernél okosabb mesterséges intelligencia pozitív hatását biztosító matematikai alapkutatásokra" összpontosít.³²⁰ Állítása szerint küldetése, hogy "formális eszközöket fejlesszen ki az általános célú mesterséges intelligencia rendszerek tiszta tervezéséhez és elemzéséhez, azzal a szándékkal, hogy az ilyen rendszerek biztonságosabbá és megbízhatóbbá váljanak, amikor kifejlesztésre kerülnek".³²¹

A megfontolás oka: Az AGI technikai részleteire összpontosító kutatási projekt.

A kizárás oka: A projekt az AGI biztonsági szempontjaira összpontosított, nem pedig az AGI fejlesztésére.

Neurogrid

Fő weboldal: <https://web.stanford.edu/group/brainsinsilicon/neurogrid.html>

³²⁰ <https://intelligence.org>

³²¹ <https://intelligence.org/about>

A Neurogrid olyan számítógépes hardver, amelyet alacsony költségű agyszimulációk futtatására terveztek. Vezetője Kwabena Boahen, a Stanford Egyetem biomérnöki tanszékének munkatársa. A Neurogrid 2014-es változata azt állítja, hogy "9000-szer gyorsabb és lényegesen kevesebb energiát használ, mint egy átlagos PC", de még mindig sokkal kevésbé energiatékony, mint az emberi agy.³²²

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: A projekt az AGI-hez kapcsolódó hardverfejlesztésre összpontosított, nem pedig magára az AGI-re.

NOMAD (Neurálisan szervezett mobil adaptív eszköz)

Fő weboldal: <http://www.nsi.edu/~nomad>

A NOMAD a kaliforniai La Jollában működő, a Nobel-díjas Gerald Edelman által vezetett The Neurosciences Institute nonprofit kutatóintézet projektje volt. A NOMAD "nagy számú, természetes szelekció által fejlődő szimulált neuronokat" használt különböző feladatok elvégzésére Goertzel (2014, 17. o.).

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

OSCAR

Fő weboldal: <http://johnpollock.us/ftp/OSCAR-web-page/oscar.html>

Az OSCAR az Arizonai Egyetemen dolgozó John Pollock projektje volt, amely 1995 és 2005 között működött. (Pollock 2009-ben hunyt el.³²³) Pollock írja: "A mesterséges intelligencia "nagy problémája" mindig is az volt, hogy emberhez hasonló intelligenciájú mesterséges ágenseket hozzunk létre... amelyek képesek valós, komplex környezetben működni... Az OSCAR egy kognitív architektúra a GIA-k [általánosan intelligens ágensek] számára, LISP nyelven megvalósítva."³²⁴

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

PAGI World

Fő weboldal: <http://rair.cogsci.rpi.edu/projects/pagi-world>

³²² <http://news.stanford.edu/pr/2014/pr-neurogrid-boahen-engineering-042814.html>, a Benjamin et al. (2014) sajtóközleménye.

³²³ <http://johnpollock.us>

³²⁴ Pollock (2008)

A PEGI World a Dél-Floridai Egyetem John Licato által vezetett projekt, amelynek székhelye a Rensselaer Polytechnic Institute, ahol Licato doktorandusz hallgató volt. A PEGI World "egy Unity 2D-ben írt szimulációs környezet, amely lehetővé teszi a mesterséges intelligencia és az AGI kutatóinak, hogy kipróbálják ötleteiket".³²⁵

A megfontolás oka: Az AGI K+F valamely aspektusával kapcsolatos projekt

A kizárás oka: A projekt az AGI értékelésére szolgáló eszközökre összpontosított, nem pedig az AGI kifejlesztésére.

PolyScheme

Fő weboldal: <https://dspace.mit.edu/handle/1721.1/8325>

A PolyScheme-t Nicholas Cassimatis doktori disszertációjához fejlesztették ki a Massachusetts Institute of Technology-n. "Többféle reprezentációs módszert, következtetési és következtetési sémát integrál az általános problémamegoldáshoz" (Goertzel 2014, 24. o.). A projekt már nem aktív.

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Kvantum Mesterséges Intelligencia Labor (QAIL)

Fő weboldal: <https://research.googleblog.com/2013/05/launching-quantum-artificial.html>

A QAIL a Google projektje, amely a kvantumszámítást kívánja felhasználni a mesterséges intelligencia fejlesztésére. A 2AI a QAIL-t AGI-projektként tartja nyilván,³²⁶ és egyes kommentátorok szerint a kvantumszámítástechnika fontos lehet az AGI fejlesztésében.³²⁷ A QAIL azonban nem utal arra, hogy az AGI-t célozza meg.

A megfontolás oka: A 2AI AGI-projektként szerepel a listán.

A kizárás oka: Nincs nyilvánvaló AGI-fókusz

SAL (ACT-R és Leabra szintézise)

Fő weboldal:

A SAL-t David Jilk, Christian Lebiere és kollégái fejlesztették ki a coloradói Boulderben működő eCortex vállalatnál, a Coloradói Egyetemen és a Carnegie Mellon Egyetemen. A projekt egy rövid publikációs rekordot hozott létre³²⁸ és 2008 óta inaktívnak tűnik. Ahogy a neve is mutatja, a SAL az ACT-R-en (lásd a dedikált ACT-R bejegyzést) és a Leabrán, egy neurális szimuláción alapul.

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

³²⁵ <http://rair.cogsci.rpi.edu/projects/pagi-world>

³²⁶ <http://2ai.org/landscape>

³²⁷ Például, Wang (2014), DeAngelis (2014)

³²⁸ Pl. Jilk et al. (2008)

A kizárás oka: Nyilvánvalóan inaktív

Jelenetalapú érvelés (SBR)

Fő weboldal: http://agi-conf.org/2015/wp-content/uploads/2015/07/agi15_bergmann.pdf

Az SBR egy AGI K+F projekt, amelyet Frank Bergmann és Brian Fenton mutatott be a 2015-ös AGI-konferencián, de azóta láthatóan nem aktív.

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

Shruti

Fő weboldal: <http://www1.icsi.berkeley.edu/~shastri/shruti>

A Shruti a Berkeley-i Kaliforniai Egyetem Lokendra Shastri projektje volt. Körülbelül 1996-2007 között működött. A projekt gyors, reflexszerű emberi következtetésen alapuló számítási eszközöket fejlesztett ki. A projektet többek között az USA Nemzeti Tudományos Alapítványa, a Haditengerészeti Kutatási Hivatal és a Hadsereg Kutatóintézete finanszírozta.³²⁹

A megfontolás oka: Goertzel (2014) AGI felülvizsgálati tanulmányában szerepel.

A kizárás oka: Nyilvánvalóan inaktív

Twitter

Fő weboldal: <https://cortex.twitter.com>

A Twitternek van egy Cortex nevű AI-csoportja, de nem találtunk utalást AGI-

munkára. **A megfontolás oka: Kizáró ok:** Nem találtak AGI-projektekre utaló

jeleket

Xapagy

Fő weboldal: <http://www.xapagy.com>

A Xapagy egy olyan kognitív architektúra, amelyet "narratív gondolkodásra terveztek, azaz az emberek által a történetekkel kapcsolatban végzett mentális folyamatok modellezésére/utánzására".³³⁰

³²⁹ <http://www1.icsi.berkeley.edu/~shastri/shruti>

³³⁰ http://www.xapagy.com/?page_id=26

A megfontolás oka: Egy AGI-konferencián bemutatott kognitív architektúra.³³¹

A kizárás oka: Az elbeszélésre való összpontosítás miatt a mesterséges intelligencia végső soron szűkös.

Ymir

Fő honlap: <http://alumni.media.mit.edu/~kris/ymir.html>

Az Ymir Kristinn Thórisson projektje volt. Körülbelül 1999-2009 között tevékenykedett.³³² "Azzal a céllal hozták létre, hogy mesterséges ágenseket emberhez hasonló kommunikációs és manipulációs képességekkel ruházzanak fel, megtestesült multimodális, feladatorientált párbeszédkészség formájában".³³³

A megfontolás oka: AGI K+F projekt

A kizárás oka: Nyilvánvalóan inaktív

³³¹ Bölöni (2014)

³³² <http://alumni.media.mit.edu/~kris/ymir.html>

³³³ Thórisson és Helgasson (2012), 8. o.

Hivatkozások

- Allen G, Kania EB, 2017. Kína felhasználja Amerika saját tervét, hogy uralja a mesterséges intelligencia jövőjét. Foreign Policy, szeptember 8., <http://foreignpolicy.com/2017/09/08/china-is-using-america-own-plan-to-dominate-the-future-of-artificial-intelligence>.
- Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D, 2016. Konkrét problémák a mesterséges intelligencia biztonságában. arXiv:1606.06565.
- Architekt, dátum nélkül. A Microsoft mesterséges intelligencia bemutatkozó partit tart. Architekt, 111. szám, <http://news.architekt.io/issues/microsoft-has-an-ai-coming-out-party-65494>.
- Armstrong S, Bostrom N, Shulman C, 2016. Verseny a szakadék felé: A mesterséges intelligencia fejlődésének modellje. AI & Society, 31(2), 201-206.
- Arrabales R, Muñoz J, 2010. A tudatos botok ébredése: A 2K BotPrize 2010 győztesének elméjében. AiGameDev, október 21., <https://aigamedev.com/open/articles/conscious-bot>.
- Arrabales R, Muñoz J, Ledezma A, Gutierrez G, Sanchis A, 2013. Egy gépi tudatossági megközelítés az emberhez hasonló robotok tervezéséhez. In Hingston P (Ed), Hihető botok. Berlin: Springer, pp. 171- 191.
- Auerbach C, Silverstein LB, 2003. Minőségi adatok: Bevezetés a kódolásba és elemzésbe. New York: NYU Press.
- Bach J, 2015. Modeling motivation in MicroPsi 2. In Bieger J, Goertzel B, Potapov A (Eds), Proceedings of AGI 2015, 8th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 3-13.
- Barrett AM, Baum SD, 2017a. A mesterséges szuperintelligens katasztrófához vezető utak modellje a kockázat- és döntéselemzéshez. Journal of Experimental & Theoretical Artificial Intelligence, 29(2), 397- 414.
- Barrett AM, Baum SD, 2017b. Kockázatelemzés és kockázatkezelés a mesterséges szuperintelligencia kutatási és fejlesztési folyamatában. In Callaghan V, Miller J, Yampolskiy R, Armstrong S (szerk.), A technológiai szingularitás. Berlin: Springer, pp. 127-140.
- Baum SD, 2017a. A közeli és hosszú távú mesterséges intelligenciára összpontosító frakciók közötti megbékélés. AI & Society, in press, doi 10.1007/s00146-017-0734-3.
- Baum SD, 2017b. Társadalmi választás etikája a mesterséges intelligenciában. AI & Society, in press, doi 10.1007/s00146-017-0760-1.
- Baum SD, 2017c. A biztonságos és társadalmilag hasznos mesterséges intelligencia előmozdításáról. AI & Society, 32(4), 543-551.
- Baum SD, Goertzel B, Goertzel TG, 2011. Mennyi idő van még az emberi szintű mesterséges intelligenciáig? Egy szakértői értékelés eredményei. Technological Forecasting and Social Change 78(1), 185-195.
- Benjamin BV, Gao P, McQuinn E, Choudhary S, Chandrasekaran AR, Bussat JM, et al., 2014. Neurogrid: Egy vegyes analóg-digitális multichip rendszer nagyméretű neurális szimulációkhoz. Proceedings of the IEEE, 102(5), 699-716.
- Best J, 2016. A Vicarious startup célja, hogy megépítse az első általános mesterséges intelligencia rendszert - csak ne számíton rá egyhamar. ZDNet, március 2., <http://www.zdnet.com/article/the-ghost-in-the-machine-vicarious-and-the-search-for-ai-that-can-rival-the-human-brain>
- Bieger J, Thórisson KR, Wang P, 2015. Biztonságos baba AGI. In Bieger J, Goertzel B, Potapov A (Eds), Proceedings of AGI 2015, 8th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 46-49.

Bölöni L, 2014. Önéletrajz alapú előrejelzés egy szituált AGI-ügynökben. In Goertzel B, Orseau L, Snider J (Eds), Proceedings of AGI 2014, 7th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 11-20.

- Bonhoff GM, 2008. A hierarchikus időbeli memória használata az anomális hálózati tevékenység észlelésére. Mesterdiplomamunka, Villamos- és Számítástechnikai Tanszék, Amerikai Légierő Technológiai Intézet.
- Boyle A, 2016. Megszületett a bot: A Kimera Systems "Nigel"-el bővíti a mesterséges intelligencia asszisztensek tömegét. GeekWire, augusztus 9., <https://www.geekwire.com/2016/kimera-systems-nigel-ai-agi>.
- Brandstätter C, Dietrich D, Doblhammer K, Fittner M, Fodor G, Gelbard F, et al., 2015. A psziché természettudományos, pszichoanalitikus modellje a szimulációhoz és az emulációhoz. Tudományos jelentés III, Institute of Computer Technology, Vienna University of Technology, Version eV005.
- Bringsjord S, 2012. A szingularitásba vetett hit logikailag törékeny. Journal of Consciousness Studies, 19(7), 14-20.
- Bostrom N, 2014. Szuperintelligencia: Útvonalak, veszélyek, stratégiák. Oxford: Oxford University Press. Bostrom N, 2017. A nyitottság stratégiai következményei a mesterséges intelligencia fejlesztésében. Global Policy, 8(2), 135-148. Chamberlain L, 2016. Az Uber nagy összegeket fektet be a mesterséges intelligenciába az Uber AI Labs elindításával. GeoMarketing, december 5., <http://www.geomarketing.com/uber-invests-in-artificial-intelligence-uber-ai-labs>
- Chen X, 2017. Gary Marcus: A mesterséges általános intelligencia felé vezető út. Medium, július 1., <https://medium.com/@Synced/gary-marcus-the-road-to-artificial-general-intelligence- ce5b1371aa02>.
- Choi D, Langley P, 2017. Az Icarus kognitív architektúra evolúciója. Cognitive Systems Research, in press, doi 10.1016/j.cogsys.2017.05.005.
- Cutler K-M, 2014. A Vicarious hatalmas, új, 40 millió dolláros növekedési kört kaparint meg a mesterséges intelligencia fejlesztésére. TechCrunch, <https://techcrunch.com/2014/03/22/vicarious-grabs-a-huge-new-40m-growth-round-to-advance-artificial-intelligence>.
- Chong HQ, Tan AH, Ng GW, 2007. Integrált kognitív architektúrák: A survey. Artificial Intelligence Review, 28(2), 103.
- Christiano P, Leike J, Brown TB, Martic M, Legg S, Amodei D, 2017. Deep reinforcement learning from human preferences. arXiv:1706.03741.
- Coffey A, Atkinson P, 1996. Making Sense of Qualitative Data: Complementary Research Strategies. Thousand Oaks, CA: Sage.
- Conn A, 2016. A Fehér Ház a mesterséges intelligencia jövőjét vizsgálja. Future of Life Institute, június 15., <https://futureoflife.org/2016/06/15/white-house-future-ai-1>.
- de Garis H, 2004. A művészeti háború. Pittsburgh: Etc Press.
- de Garis H, Shuo C, Goertzel B, Ruiting L, 2010. A mesterséges agyi projektek világméretű áttekintése, I. rész: Nagyszabású agyszimulációk. Neurocomputing, 74(1), 3-29.
- DeAngelis SF, 2014. A kvantumszámítógépek át fogják alakítani a világot - reméljük. Enterra Insights Blog, december 24., <http://www.enterrasolutions.com/quantum-computers-will-transform-world-hope>.
- Deaton C, Shepard B, Klein C, Mayans C, Summers B, Brusseau A, et al., 2005. Az átfogó terrorizmus tudásbázis a Cyc. In Proceedings of the 2005 International Conference on Intelligence Analysis.
- Dewey D, 2015. Hosszú távú stratégiák a gyors felszállásból eredő egzisztenciális kockázat megszüntetésére. In Müller VC (szerk.), A mesterséges intelligencia kockázatai. Boca Raton: CRC, pp. 243-266.
- Dong D, Franklin S, 2014. A különböző kognitív architektúrákban megvalósított cselekvésvégrehajtási folyamat: A review. Journal of Artificial General Intelligence, 5(1), 49-68.
- Duch W, Oentaryo RJ, Pasquier M, 2008. Kognitív architektúrák: Merre tovább? In Wang P, Goertzel

B, Franklin S (Eds), *Frontiers in Artificial Intelligence and Applications*, Vol. 171. Amsterdam: IOS Press, pp. 122-136.

Dvorsky G, 2013. How Skynet may emerge from simple physics. io9, 26 April, <http://io9.gizmodo.com/how-skynet-might-emerge-from-simple-physics-482402911>.

- Etherington D, 2017. A Microsoft mesterséges intelligencia-kutató laboratóriumot hoz létre a Google és a DeepMind kihívására.
TechCrunch, július 12., <https://techcrunch.com/2017/07/12/microsoft-creates-an-ai-research-lab-to-challenge-google-and-deepmind>
- Etzioni O, 2016. Nem, a szakértők szerint a szuperintelligens mesterséges intelligencia nem jelent veszélyt az emberiségre. MIT Technology Review, szeptember 20.,
<https://www.technologyreview.com/s/602410/no-the-experts-dont-think-superintelligent-ai-is-a-threat-to-humanity>.
- Fernando C, Banarse D, Blundell C, Zwols Y, Ha D, Rusu AA, et al., 2017. PathNet: Evolution channels gradient descent in super neural networks. arXiv:1701.08734.
- Gershgorn D, 2017. Kína finanszírozza a Baidut, hogy felvegye a versenyt az Egyesült Államokkal a mélytanulási kutatásban. Quartz, február 22., <https://qz.com/916738/china-is-funding-baidu-to-take-on-the-united-states-in-deep-tanulas-kutatas>.
- Gibbs S, 2014. A Google 400 millió fontért megveszi a Deepmind nevű brit mesterséges intelligencia startupot. The Guardian, január 27.,
<https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind>.
- Goertzel B, 2008. Ben Goertzel beszámolója Xiamenből, Kínából. Institute for Ethics and Emerging Technologies, június 20., <https://ieet.org/index.php/IEET2/more/goertzel20080620>.
- Goertzel B, 2009. Kína fogja először megépíteni az AGI-t? The Multiverse According to Ben, július 26., <http://multiverseaccordingtoben.blogspot.com/2009/07/will-china-build-agi-first.html>.
- Goertzel B, 2010. Kozmikus kiáltvány: Gyakorlati filozófia a poszthumán korszak számára. Humanity+ Press.
- Goertzel B, 2011. Pei Wang a mesterséges általános intelligencia felé vezető úton. h+ Magazine, január 27., <http://hplusmagazine.com/2011/01/27/pei-wang-path-artificial-general-intelligence>.
- Goertzel B, 2012a. Készítsen-e az emberiség egy globális AI dadust, hogy késleltesse a szingularitást, amíg azt jobban megértjük? Journal of Consciousness Studies, 19(1-2), 96-111.
- Goertzel B, 2012b. Mesterséges intelligencia és Etiópia: Egy váratlan szinergia. KurzweilAI, október 25., <http://www.kurzweilai.net/ai-and-ethiopia-an-unexpected-synergy>.
- Goertzel B, 2014. Mesterséges általános intelligencia: Az intelligens mesterséges intelligencia: Fogalom, a technika jelenlegi állása és jövőbeli kilátások. Journal of Artificial General Intelligence, 5(1), 1-48.
- Goertzel B, 2015. Szuperintelligencia: Félelmek, ígéretetek és lehetőségek. Journal of Evolution and Technology, 25(2), 55-87.
- Goertzel B, 2016. A fejlett AGI-k emberhez hasonló értékrenddel való ellátása: Két tézis. Journal of Evolution and Technology, 26(1), 50-72.
- Goertzel B, 2017a. A mesterséges intelligencia korporatizálódása komoly veszélyt jelent az emberiségre. H+ Magazin, július 21., <http://hplusmagazine.com/2017/07/21/corporatization-ai-major-threat-humanity>.
- Goertzel B, 2017b. SingularityNET-AGI az emberektől, az emberek által és az emberekért (és a robotokért!). Medium, október 26., <https://medium.com/ben-goertzel-on-singularitynet/singularitynet-agi-of-the-people-by-the-people-and-for-the-people-and-the-robots-7246a6886e90>.
- Goertzel B, Pitt J, 2012. Kilenc módja annak, hogy a nyílt forráskódú AGI-t a barátságosság irányába tereljük. Journal of Evolution & Technology, 22(1), 116-131.
- Goertzel B, Lian R, Arel I, De Garis H, Chen S, 2010. A mesterséges agyi projektek világméretű áttekintése, II. rész: Biológiai inspirált kognitív architektúrák. Neurocomputing, 74(1), 30-49.

- Goertzel B, Giacomelli S, Hanson D, Pennachin C, Argentieri M, the SingularityNET team, 2017. SingularityNET: A mesterséges intelligenciák decentralizált, nyílt piaca és hálózatközi hálózata. SingularityNET, november 1., <https://public.singularitynet.io/whitepaper.pdf>.
- Jó JJ, 1965. Spekulációk az első ultraintelligens géppel kapcsolatban. Advances in Computers, 6, 31-88.
- Griffin M, 2017. A Baidu mesterséges intelligenciája épp most érte el a zéró lövéses tanulást. Fanatical Futurist, április 14., <http://www.fanaticalfuturist.com/2017/04/baidus-ai-just-achieved-zero-shot-learning>.

- Hawkins J, 2015. A Terminátor nem jön. A jövő majd megköszöni nekünk. recode, március 2., <https://www.recode.net/2015/3/2/11559576/the-terminator-is-not-coming-the-future-will-thank-us>.
- Hawkins J, 2017. Mit kell tanulniuk az intelligens gépeknek a neokortextól. IEEE Spectrum, június 2., <http://spectrum.ieee.org/computing/software/what-intelligent-machines-need-to-learn-from-the-neocortex>.
- Heath A, 2017. Mark Zuckerberg terve, hogy szavazati jogot nem biztosító Facebook-részvényeket hozzon létre, szeptemberben bíróság elé kerül. Business Insider, május 4., <http://www.businessinsider.com/trial-challenge-facebook-non-voting-shares-set-september-2017-2017-5>.
- Hellman D, Hellman M, 2016. A kapcsolatok új térképe: Creating True Love At Home and Peace On The Planet. New Map Publishing.
- Hibbard B, 2002. Szuperintelligens gépek. New York: Springer.
- Magas P, 2016. A Vicarious az a mesterséges intelligenciával foglalkozó vállalat, amelynek Zuckerberg, Bezos, Musk és Thiel a befektetői között szerepel. Forbes, április 11., <https://www.forbes.com/sites/peterhigh/2016/04/11/vicarious-is-the-ai-company-that-includes-zuckerberg-bezos-musk-and-thiel-as-investors>.
- Hughes JJ, 2007. Globális technológiai szabályozás és potenciálisan apokaliptikus technológiai fenyegetések. In Allhoff F (szerk.), Nanoetika: The Ethical and Social Implications of Nanotechnology. Hoboken, NJ: John Wiley, pp. 201-214.
- Ingram M, 2017. Az Alphabetnél csak két részvényes számít. Fortune, június 7., <http://fortune.com/2017/06/07/alphabet-shareholders-meeting>
- Jee C, 2016. Mi a mesterséges általános intelligencia? És vajon a Kimera Systems "Nigel"-je megfelel ennek?
TechWorld, augusztus 26., <https://www.techworld.com/data/what-is-artificial-general-intelligence-3645268>
- Jilk DJ, Lebiere C, O'Reilly RC, Anderson JR, 2008. SAL: Egy kifejezetten pluralista kognitív architektúra. Journal of Experimental and Theoretical Artificial Intelligence, 20(3), 197-218.
- Joy B, 2000. Miért nincs ránk szüksége a jövőnek. Wired, 8(4), 238-263.
- Kaczynski T, 1995. Az ipari társadalom és jövője.
- Kotseruba I, Gonzalez OJA, Tsotsos JK, 2016. A kognitív architektúra kutatás 40 évének áttekintése: Focus on perception, attention, learning and applications. arXiv:1610.08602.
- Lane PC, Gobet F, 2012. Az implicit tanulás és a társasjátékok értelmezésének CHREST modelljei. In Bach J, Goertzel B, Ikle M (Eds), Proceedings of AGI 2012, 5th International Conference on Artificial General Intelligence. Berlin: Springer, pp. 148-157.
- Langley P, 2017. A kognitív architektúrák kutatásának előrehaladása és kihívásai. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17), pp. 4870-4876.
- Langley P, Laird J, Rogers S, 2008. Kognitív architektúrák: Kutatási kérdések és kihívások. Cognitive Systems Research 10, 141-160.
- LeCun Y, Bengio Y, Hinton G, 2015. Mélytanulás. Nature, 521(7553), 436-444.
- Lee HS, Betts S, Anderson JR, 2017. A problémamegoldó példák mélyszerkezeti információkkal való kidomborítása megkönnyíti az átadást. Journal of Experimental Education, 85(2), 309-333.
- Lentzos F, 2011. Nehéz bizonyítani: A biológiai fegyverekről szóló egyezmény ellenőrzési dilemmája. Nonproliferation Review, 18(3), 571-582.
- Lipsitch M, Inglesby TV, 2014. Moratórium az új, potenciális pandémiás kórokozók létrehozására irányuló kutatásokra. mBio, 5(6), e02366-14.
- Love D, 2014. A világ legambiciózusabb mesterséges intelligencia projektje 30 éve szinte titokban működik. Business Insider, július 2., <http://www.businessinsider.com/cycorp-ai-2014-7>
- Maddox T, 2016. A Baidu legyen az Ön AI és gépi tanulási platformja? ZDNet, december 1., <http://www.zdnet.com/article/should-baidu-be-your-ai-and-machine-learning-platform>

Madl T, Franklin S, 2015. Korlátozott inkrementalista morális döntéshozatal egy biológiailag inspirált kognitív architektúra számára. In Trapp R (szerk.), A robotok etikai rendszereinek építési kézikönyve: Követelmények, módszerek, megvalósítások. Cham, Svájc: Springer, pp. 137-153.

- Mannes J, 2017. A Tencent mesterséges intelligencia kutatóközpontot nyit Seattle-ben. TechCrunch, április 28., <https://techcrunch.com/2017/04/28/tencent-to-open-ai-research-center-in-seattle>.
- Marcus G, 2017. A mesterséges intelligencia megrekedt. Íme, hogyan lehet előrébb vinni. New York Times, július 29. <https://nytimes.com/2017/07/29/opinion/sunday/artificial-intelligence-is-stuck-heres-how-to-move-it-forward.html>
- Marcus G, Davis E, 2013. A mindenek nagy egyesített elmélete. New Yorker, május 6., <http://www.newyorker.com/tech/elements/a-grand-unified-theory-of-everything>
- Marquis C, Toffel, MW, Zhou Y, 2016. Ellenőrzés, normák és szelektív közzététel: A global study of greenwashing. Organization Science, 27(2), 483-504.
- McDermott D, 2012. David Chalmers válasza a szingularitásra. Journal of Consciousness Studies, 19(1-2), 167-172.
- Menager D, Choi D, 2016. Az epizodikus memória robusztus megvalósítása egy kognitív architektúra számára.
In Papafragou A, Grodner D, Mirman D, Trueswell JC (Eds), Proceedings of the 38th Annual Conference of the Cognitive Science Society. Austin: Cognitive Science Society, pp. 620-625.
- Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, Akopyan F, et al., 2014. Egy millió tüskés neuron integrált áramkör skálázható kommunikációs hálózattal és interfésszel. Science, 345(6197), 668-673.
- Metz C, 2016. Az Uber megvesz egy titokzatos startupot, hogy mesterséges intelligenciával foglalkozó céggé váljon. Wired, december 5., <https://www.wired.com/2016/12/uber-buys-mysterious-startup-make-ai-company>
- Muehlhauser L, Helm L, 2012. Intelligenciarobbanás és gépi etika. In Eden A, Søraaker J, Moor JH, Steinhart E (szerk.), Szingularitás hipotézisek: A Scientific and Philosophical Assessment. Berlin: Springer, pp. 101-126.
- Nadella S, 2016. A jövő partnersége. Slate, június 28., http://www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_és_a_i_mindig_együtt_működhet_a_társadalom_megoldásán.html
- Ng KH, Du Z, Ng GW, 2017. DSO kognitív architektúra: Egységes gondolkodás integratív memóriával a globális munkaterület-elmélet segítségével. In Everitt T, Goertzel B, Potapov A (Eds), Proceedings of AGI 2017, 10th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 44-53.
- Nivel E, Thórisson KR, Steunebrink BR, Dindo H, Pezzulo G, Rodriguez M, et al., 2013. Korlátozott rekurzív önfejlesztés. Reykjavik University School of Computer Science Technical Report RUTR-SCS13006.
- Novet J, 2016. A Salesforce kutatócsoportot alakít, elindítja az Einstein A.I. platformot, amely a Sales Clouddal, Marketing Clouddal működik együtt. Venture Beat, szeptember 18., <https://venturebeat.com/2016/09/18/salesforce-forms-research-group-launches-einstein-a-i-platform-that-at-works-with-sales-cloud-marketing-cloud>.
- O'Reilly RC, Hazy TE, Herd SA, 2016. A Leabra kognitív architektúra: Hogyan játsszunk 20 elvet a természettel. In Chipman SEF (szerk.), The Oxford Handbook of Cognitive Science. Oxford: Oxford University Press, pp. 91-116.
- Oreskes N, Conway EM, 2010. A kétely kereskedői: Hogyan homályosította el egy maroknyi tudós az igazságot a dohányfüsttől a globális felmelegedésig. New York: Bloomsbury.
- Oudeyer PY, Ly O, Rouanet P, 2011. Robusztus, intuitív és emergens fizikai ember-robot interakció vizsgálata a humanoid Acroban segítségével. IEEE-RAS International Conference on Humanoid Robots, Bled, Szlovénia.

- Pollock JL, 2008. OSCAR: Egy legyőzhető érvelésen alapuló ágensarchitektúra. Proceedings of the 2008 AAAI Spring Symposium on Architectures for Intelligent Theory-Based Agents, pp. 55-60.
- Poo MM, Du JL, Ip NY, Xiong ZQ, Xu B, Tan T, 2016. China Brain Project: Alapvető idegtudományok, agyi betegségek és az agy által inspirált számítástechnika. Neuron, 92(3), 591-596.
- Posner EA, 2014, Az emberi jogi jog alkonya. Oxford: Oxford University Press.

- Potapov A, Rodionov S, 2014. Univerzális empátia és etikai elfogultság a mesterséges általános intelligencia számára.
Journal of Experimental & Theoretical Artificial Intelligence, 26(3), 405-416.
- Potapov A, Rodionov S, Potapova V, 2016. Valós idejű GA-alapú valószínűségi programozás alkalmazása robotvezérlésben. In Steunebrink B, Wang P, Goertzel B (Eds), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 95-105.
- Valódi mesterséges intelligencia, 2017. Próza mesterséges intelligencia. <http://realai.org/prosaic>
- Rosenbloom PS, 2013. A Sigma kognitív architektúra és rendszer. AISB Quarterly, 136, 4-13. RT, 2017. 'Aki a mesterséges intelligenciában vezet, az fogja uralni a világot': Putyin az orosz gyerekeknek a tudás napján.
RT, szeptember 1., <https://www.rt.com/news/401731-ai-rule-world-putin>
- Samsonovich AV, 2010. A megvalósított kognitív architektúrák egységes katalógusa felé. Biológiai inspirált kognitív architektúrák, 221, 195-244.
- Scherer MU, 2016. A mesterséges intelligencia rendszerek szabályozása: Kockázatok, kihívások, kompetenciák és stratégiák. Harvard Journal of Law & Technology, 29(2), 353-400.
- Schienze EW, Tuana N, Brown DA, Davis KJ, Keller K, Shortle JS, et al., 2009. Az NSF Broader Impacts Criterion szerepe a kutatási etika pedagógiájának javításában. Social Epistemology, 23(3-4), 317-336.
- Shapiro SC, Bona JP, 2010. A GLAIR kognitív architektúra. International Journal of Machine Consciousness, 2(2), 307-332.
- Shapiro SC, Schlegel DR, 2016. Természetes nyelvi megértés az információfúzióhoz. In Rogova G, Scott P (szerk.), Fúziós módszertanok a válságkezelésben. Cham, Svájc: Springer, pp. 27- 45.
- Shed S, 2017. A DeepMind mesterséges intelligencia kutatóinak serege Mountain View-ban már több mint 20 fő.
Business Insider, május 31., <http://www.businessinsider.com/deepminds-small-army-of-ai-researchers-in-mountain-view-is-growing-2017-5>
- Shulman C, 2009. Fegyverzetellenőrzés és hírszerzési robbanások. In 7th European Conference on Computing and Philosophy (ECAP), Bellaterra, Spanyolország, július 2-4.
- Skaba W, 2012a. Bináris térfelosztás mint belső jutalom. In Bach J, Goertzel B, Ikle M (Eds), Proceedings of AGI 2012, 5th International Conference on Artificial General Intelligence. Berlin: Springer, pp. 282-291.
- Skaba W, 2012b. Az AGINAO önprogramozó motor. Journal of Artificial General Intelligence, 3(3), 74-100.
- Sotala K, Yampolskiy RV, 2014. A katasztrofális AGI kockázatra adott válaszok: egy felmérés. Physica Scripta, 90(1), 018001.
- Starzyk JA, Graham J, 2015. MLECOG: Motiváltan tanuló megtettesült kognitív architektúra. IEEE Systems Journal, 11(3), 1272-1283.
- Steunebrink BR, Thórisson KR, Schmidhuber J, 2016. Rekurzív önfejlesztők növekedése. In Steunebrink B, Wang P, Goertzel B (Eds), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 129-139.
- Stilgoe J, Maynard A, 2017. Itt az ideje néhány zűrös, demokratikus vitának a mesterséges intelligencia jövőjéről. The Guardian, február 1., <https://www.theguardian.com/science/political-science/2017/feb/01/ai-artificial-intelligence-its-time-for-some-messy-democratic-discussions-about-the-future>.
- Stranegård C, Nizamani AR, 2016. Szimbolikus és szubszimbolikus érvelés integrálása. In Steunebrink B, Wang P, Goertzel B (Eds), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 171-180.
- Stranegård C, Nizamani AR, Juel J, Persson U, 2016. Tanulás és következtetés ismeretlen

tartományokban.

Journal of Artificial General Intelligence, 7(1), 104-127.

- Stranegård C, Svangård N, Bach J, Steunebrink B, 2017a. Generikus animátumok. In Everitt T, Goertzel B, Potapov A (Eds), Proceedings of AGI 2017, 10th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 23-32.
- Stranegård C, Svangård N, Lindström D, Bach J, Steunebrink B, 2017b. Az animátori út a mesterséges általános intelligenciához. In Proceedings of IJCAI-17 Workshop on Architectures for Generality & Autonomy.
- Taatgen N, Anderson J, 2010. A kognitív architektúrák múltja, jelene és jövője. Topics in Cognitive Science 2(4), 693-704.
- Tamturk V, 2017. A Google, az Apple, a Facebook és az Intel harca a mesterséges intelligencia elsőségéért. CMS Connected, április 21. <http://www.cms-connected.com/News-Archive/April-2017/Google-Apple-Facebook-Intel-Microsoft-Salesforce-Twitter-Battle-AI-Supremacy>.
- Temperton J, 2016. Az Uber felvásárolja a mesterséges intelligenciával foglalkozó céget az önvezető autós projekt fellendítése érdekében. Wired, december 6., <http://www.wired.co.uk/article/uber-artificial-intelligence-lab-self-driving-cars>
- The Economist, 2016. A nem működőktől a neurális hálózatokig. The Economist, június 25., <https://www.economist.com/news/special-report/21700756-artificial-intelligence-boom-based-old-idea-modern-twist-not>.
- Theil S, 2015. Miért ment rosszul az emberi agy projektje - és hogyan lehet helyrehozni. Scientific American, október 1., <https://www.scientificamerican.com/article/why-the-human-brain-project-went-wrong-and-how-to-fix-it>.
- Thórisson K, Helgasson H, 2012. Kognitív architektúrák és autonómia: A comparative review. Journal of Artificial General Intelligence, 3(2), 1-30.
- Totschnig W, 2017. A szuperintelligencia problémája: Politikai, nem technológiai. AI & Society, in press, doi 10.1007/s00146-017-0753-0.
- Townsend JC, 2016. Ez a startup megtanítja a gépeket úgy gondolkodni, érvelni és kommunikálni, mint mi. Fast Company, november 22., <https://www.fastcompany.com/3065779/this-startup-is-teaching-machines-to-think-reason-and-communicate-like-us>
- TWiStartups, 2016. Scott Phoenix, a Vicarious társalapítója a mesterséges intelligencia területén történő innovációról és az emberi agy felszabadításáért folytatott versenyről, hogy létrehozzák a mesterséges általános intelligenciát, az utolsó technológiát, amelyet az emberek fognak feltalálni. Medium, május 13., <https://medium.com/@TWiStartups/vicarious-co-founder-scott-phoenix-on-innovating-in-ai-the-race-to-unlock-the-human-brain-to-9161264b0c09>.
- Vanian J, 2017a. Az Apple mesterséges intelligencia guruja a sci-fi jövőről beszél. Fortune, március 28., <http://fortune.com/2017/03/28/apple-artificial-intelligence>.
- Vanian J, 2017b. Az Apple most még nyilvánvalóbbá tette mesterséges intelligencia terveit. Fortune, július 19., <http://fortune.com/2017/07/19/apple-artificial-intelligence-research-journal>.
- Veness J, Ng KS, Hutter M, Uther W, Silver D, 2011. Monte-Carlo AIXI közelítés. Journal of Artificial Intelligence Research, 40(1), 95-142.
- Wang P, 2012. Motivációkezelés AGI-rendszerekben. In Bach J, Goertzel B, Ikle M (Eds.), Proceedings of AGI 2012, 5th International Conference on Artificial General Intelligence. Berlin: Springer, pp. 352-361.
- Wallach W, Franklin S, Allen C, 2010. Az emberi és mesterséges ágensek erkölcsi döntéshozatalának fogalmi és számítási modellje. Topics in Cognitive Science, 2(3), 454-485.
- Wallach W, Allen C, Franklin S, 2011. Tudat és etika: Mesterségesen tudatos erkölcsi ágensek. International Journal of Machine Consciousness, 3(1), 177-192.
- Wang B, 2014. A kvantumszámítás és a mesterséges intelligencia új megközelítései megkaphatják az erőforrásokat a számítástechnika valódi áttöréséhez. Next Big Future, március 31., <https://www.nextbigfuture.com/2014/03/quantum-computing-and-new->

[approaches-to.html](#).

Wang P, Li X, 2016. A tanulás különböző felfogásai: Funkcióközelítés vs. önszerveződés.
In Steunebrink B, Wang P, Goertzel B (Eds.), Proceedings of AGI 2016, 9th International
Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 140-149.

- Wang P, Talanov M, Hammer P, 2016. Az érzelmi mechanizmusok a NARS-ban. In Steunebrink B, Wang P, Goertzel B (Eds.), Proceedings of AGI 2016, 9th International Conference on Artificial General Intelligence. Cham, Svájc: Springer, pp. 150-159.
- Webb A, 2017. A mesterséges intelligencia úttörője a jövő reneszánsz gépét akarja megépíteni. Bloomberg, január 16., <https://www.bloomberg.com/news/articles/2017-01-16/ai-pioneer-wants-to-build-the-renaissance-machine-of-the-future>.
- Webster G, Creemers R, Triolo P, Kania E, 2017. Kína terve a mesterséges intelligencia terén való "vezető szerepvállalásra": cél, kilátások és problémák. New America, augusztus 1., <https://www.newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-lead-ai-purpose-prospects-and-problems>
- Weng J, Evans CH, Hwang WS, Lee Y-B, 1999. A mesterséges intelligencia fejlődési megközelítése: Fogalmak, fejlesztési algoritmusok és kísérleti eredmények. In Proceedings of NSF Design and Manufacturing Grantees Conference, Long Beach, CA.
- Wilson G, 2013. Az új technológiákból eredő globális katasztrófa- és egzisztenciális kockázatok minimalizálása a nemzetközi jog segítségével. Virginia Environmental Law Journal, 31,307-364.
- Wissner-Gross AD, Freer CE, 2013. Kauzális entrópikus erők. Physical Review Letters, 110(16), 168702.
- Yang W, Liu W, Viña A, Tuanmu MN, He G, Dietz T, Liu J, 2013. A csoportméret nemlineáris hatásai a kollektív cselekvésre és az erőforrás-eredményekre. Proceedings of the National Academy of Sciences, 110(27), 10916-10921.
- Yampolskiy RV, 2013. Mesterséges intelligencia biztonságtechnika: Miért rossz megközelítés a gépi etika. In Müller VC (szerk.), A mesterséges intelligencia filozófiája és elmélete. Berlin: Springer, pp. 389-396.
- Yampolskiy R, Fox J, 2013. Biztonságtechnika a mesterséges általános intelligencia számára. Topoi, 32(2), 217- 226.
- Yudkowsky E, 2004. Koherens extrapolált akarat. San Francisco: The Singularity Institute.
- Zhang Q, Walsh MM, Anderson JR, 2016. A szonda hasonlóságának hatása az előhívási és összehasonlítási folyamatokra az asszociatív felismerésben. Journal of Cognitive Neuroscience, 29(2), 352-367.