

A kiadványhoz kapcsolódó viták, statisztikák és szerzői profilok a következő címen olvashatók: <https://www.researchgate.net/publication/328332406>.

A csápos művészi intelligencia és annak architektúrája, bemutatásra került

Preprint - október 2018

SZÍTÁCIÓK

0

READS

172

6 szerző, köztük:



Naveen Sundar Govindarajulu

48 PUBLIKÁCIÓ 232 IDÉZET

NÉZZE MEG AZ
ADATLAPOT



Atriya Sen

New Orleans-i Egyetem

10 PUBLIKÁCIÓ 62 IDÉZET

NÉZZE MEG AZ
ADATLAPOT



Matthew Peveler

Rensselaer Polytechnic Institute

31 PUBLIKÁCIÓ 55 IDÉZET

NÉZZE MEG AZ
ADATLAPOT



Biplav Srivastava

Dél-Karolinai Egyetem

213 PUBLIKÁCIÓ 2,379 IDÉZET

NÉZZE MEG AZ
ADATLAPOT

E kiadvány néhány szerzője szintén dolgozik ezeken a kapcsolódó projekteken:



Az AI View projekt bizalmi értékelése



Robosztus rendszer a nyilvános maszkviselési viselkedés felismerésére és magyarázatára [Projekt megtekintése](#)

Az ezen az oldalon található összes tartalmat [Naveen Sundar Govindarajulu](#) töltötte fel 2019. augusztus 14-én.

A felhasználó a letöltött fájlt továbbfejlesztését kérte.

Bemutatásra kerül a csápos mesterséges intelligencia és annak architektúrája

Selmer Bringsjord¹, Naveen Sundar G¹, Atriya Sen¹, Matthew Peveler¹, Biplav Srivastava² Kartik Talamadupula²

¹ Rensselaer Polytechnic Institute (RPI); RAIR Lab

² IBM Research

Selmer.Bringsjord@gmail.com, naveensundarg@gmail.com, atriya@atriyasen.com, matt.peveler@gmail.com,

biplavs@us.ibm.com, krtalamad@us.ibm.com

Absztrakt

A következőkben röviden bemutatjuk a disztributált, több ágensből álló mesterséges intelligencia egy új formáját, amelyet "tentacular"-nak nevezünk. A sátoros mesterséges intelligenciát hat tulajdonsága különbözteti meg, amelyek többek között magukban foglalják a nagy kifejezőerejű kalkuláción (logikákon) alapuló gondolkodási és tervezési képességet, és amely csak egy vagy több adott hálózat hatósugarával körülhatárolt távolságokon keresztül vesz igénybe segédügynököket.

¹Pontosan ezt a kifejezést használja Russell és Norvig [2009]. Más átfogó áttekintések is megfelelnek a Russell-Norvig orientációnak; pl. [Luger, 2008].

1 Bevezetés

A következőkben röviden bemutatjuk az elosztott, több ágensből álló mesterséges intelligencia egy új formáját. Az AI artifact S jelenleg nem értelmezhető olyan ágensként, amely előre meghatározott célokkal, egy rögzített bemenetek és kimenetek, valamint kötelezettségek és engedélyek halmaza. Az ágensnek nincs mozgásteret céljai elérésében, illetve a kötelezettségeinek, tilalmainak vagy más, őt kötő jogi/etikai elveknek a betartásában. Szükségünk van-e olyan ügynökökre, amelyek ezeken a korlátozásokon túl? Egy szerény példa következik: A napi munkába járás során egy ügynök a_c az autójában azt észleli, hogy a szokásosnál nagyobb a forgalom az út felé. helyi bolt. Ezután konzultál az időjárési szolgálattal, és megállapítja, hogy egy nagy vihar tart az Ön városa felé. a_c ezt közvetíti az információkat a_h , egy ügynöknek az Ön otthonában. a_h aztán kommunikál cates egy ügynökkel a_p a telefonján, és kiderül, hogy nem tud az Ön felé közeledő viharról, mivel Ön már nem készült fel rá; és tudatlanságának további bizonyítéka, hogy nem olvasta el a viharról szóló értesítéseket. a_h aztán a naptárából arra következtet, hogy nem lehet, hogy elég idő lesz, hogy beszerezd a készleteket, miután elővástad az értesítést... a_h utasítja a_c , hogy ajánljon Önnek egy listát a hazafelé vezető úton bevásárolni kívánt készletekről, amely legalább n terméket tartalmaz bizonyos kategóriákban (pl. 3 gallon palack víz).

A mai mesterséges intelligencia, ahogyan azt bármelyik ortodox, átfogó áttekintés (pl. [Russell és Norvig, 2009]) meghatározza, a mesterséges ágensek tervezéséből, létrehozásából, megvalósításából és elemzéséből áll.¹ Minden ilyen a -ügynök információt vesz fel a

az adott E környezetét (azaz befogadja E **érzékelseit**), számításokat végez, majd a számítások erejéig

a számítás, végrehajt egy műveletet/műveleteket abban a környezetben. (Természetesen egy olyan ágens esetében, amely tartósan fennmarad, ez a ciklus az időben ismétlődik.) E definíció szerint egy számítógépes program, amely

mondjuk, az $n!$ faktoriális függvényt valósítja meg, egy olyan művészi ágensnek minősül (nevezzük " a_{FAC} "), amely, az alapvető aritmetikai számítások E környezetében működik; és az ember, aki a számítási feladatokat

ceved és írt ez a program épített egy mesterséges ügynök. Míg a manapság reklámozott mesterséges ügynökök nagy része sokkal lenyűgözőbb, mint a_{FAC} ,

a mi célunk az, hogy a világra hozzuk, egy évtizeden belül egy forradalmi mesterséges intelligenciát, amely művészi

a radikálisan magasabb intelligenciaszinttel (beleértve az elég magas intelligenciát ahhoz, hogy az ágenseket *kognitív tudatossággal rendelkezőnek* minősítsék) és hatalommal rendelkező hivatalos ágensek. Ezt az elképzelt mesterséges intelligenciát nevezzük **sátoros mesterséges intelligenciának**, vagy röviden csak "TAI"-nak (rímel a "pie"-re). Mielőtt bemutatnánk a TAI-val kapcsolatos architektúra-szintű információkat, egy olyan példát hozunk, amely egy kicsit robusztusabb, mint az első bekezdésben említett.

Tegyük fel, hogy egy mesterséges intelligencia-ügynök a_{HOME} felügyel egy otthont, és azzal az egyetlen, szerény feladattal bízzák meg, hogy mozdítson el egy csészt.

az otthoni konyhaasztalon egy csészealjra, amely szintén az asztalon van. Hogyan valósítsa meg az ügynök ezt a célt? Ha a mesterséges intelligencia delegálni tud egy olyan robotot a házban, amely képes a szabványos asztali tárgyak manipulálására egy szűk asztali környezetben E_{TABLE} , és ez a robot az asztalnál van, vagy oda tud jutni egy

ésszerű időn belül, akkor természetesen a_{HOME} irányíthatja a robotot, hogy vegye fel a csészt, és tegye a csészealjra. Ez a

nincs semmi különös, hiszen a mai mesterséges intelligencia olyan ügynök-robot kombinációkat adott nekünk, amelyek laboratóriumokban (például a sajátunkban) és hamarosan a technológizált világ otthonaiban is képesek rutinszerűen és megbízhatóan elvégezni az ilyesmit. Valójában ez a fajta képesség arra, hogy terveket találjunk és asztali tárgyakat mozgassunk, hogy elérjük céljainkat asztali környezetben.² már évtizedek óta megoldott probléma a kutatás szempontjából [Gene- sereth és Nilsson, 1987], például. Nem csak ez, hanem régóta léteznek olyan tételek, amelyek szerint a különböző szabványos asztali tárgyak tetszőleges kiindulási konfigurációkban történő mozgatására vonatkozó tervek megtalálásának belső nehézsége algoritmikusan megoldható és általában követhető.³

²Az ilyen környezetek a hagyományosan "blokk-világoknak" nevezett környezetek változatai.

³Lásd például [Gupta, 1992].

A mai mesterséges intelligencia azonban, ha úgy tetszik, egy kicsit hazugságban él. Hogy miért? Mert a való életben az ügynök $\mathcal{A}_{\text{HOME}}$ nem csak az asztali környezetben $\mathbf{E}_{\text{TABLE}}$ működne; inkább a

Az elképzelés az, hogy ennek az ügynöknek képesnek kell lennie megérteni és kezelni az otthon teljes környezetét \mathbf{E}_{HOME} , amely bizonyára sokkal több mindemből áll, mint a konyhában szokásos dolgok.

asztal! Az otthonokban lehetnek szülők, gyerekek, kutyák, látogatók, ... *ad in definitum*.

Tegyük fel például, hogy $\mathcal{A}_{\text{HOME}}$ találja, hogy az asztali robot eltört, mert a házban élő, élénk beagle szétmancangolta. *Akkor hogyan* oldja meg $\mathcal{A}_{\text{HOME}}$, hogy a csészéaljat elmozdítsa? A mai mesterséges ágensek képesek a

a tervezés, ami a bonyodalom előtt működött, most akadályokba ütközik. De nem így egy TAI-ügynök. Ennek egyik oka, hogy a TAI-ügynökök képesek emberi szintű kommunikációra. A oldalon

Bizonyos körülmények között $\mathcal{A}_{\text{HOME}}$ weboldalon belül $\mathcal{A}_{\text{HOME}}$ ügynök számára a feladat elvégzésének leghatékonyabb módja az lehet, hogy egyszerűen udvariasan azt mondja az I/loT (Internet vagy Interent of Things) segítségével egy

hangszóró, okostelefon vagy okoszemüveg segítségével egy otthon tartózkodó emberhez (akinek az elméjéről a TAI-ügynöknek van egy modellje).

a kérdéses asztalnál: "Lenne szíves azt a csészét a csészéalj tetejére tenni?" Természetesen $\mathcal{A}_{\text{HOME}}$ nem biztos, hogy olyan szerencsés, hogy egy ember szolgálati rendelkezésre állnak: talán

nincs ember otthon, mégis el kell végezni a feladatot. Ebben az esetben egy TAI-ügynök még mindig képes elvégezni a dolgokat, kreatív gyorsasággal. Tegyük fel például, hogy az otthonban egy családtag előre kapott egy kis léghajót, amely képes körbepülni a lakásban és felvenni dolgokat.⁴ A TAI-ügynök ekkor aktiválhatja és használhatja ezt a léghajót az I/loT-n keresztül, hogy a csészét a csészéalj tetejére tegye. De mi is pontosan az a TAI-ügynök? Azt mondjuk, hogy egy TAI-ügynöknek lennie kell:

D1 *Problémamegoldásra képes.* Míg, mint már említettük, a standard mesterséges intelligencia az észlelések és cselekvések egyszerű leképezését tekinti *jóhiszemű* mesterséges intelligenciának, addig a TAI-ügynököknek képesnek kell lenniük a problémamegoldásra. Ez a TAI első, jelentéktelen tulajdonságának tűnhet, de meg kell jegyeznünk egy következményt, amely ebből a tulajdonságból ered: Mivel a problémamegoldás a mesterséges intelligencia fő részterületeinek képességeit foglalja magában, a TAI-ügynökök sokoldalú képességekkel rendelkeznek. A problémamegoldás a mesterséges intelligencia alábbi részterületein igényel képességeket: tervezés, érvelés, tanulás, kommunikáció, kreativitás (legalábbis annak viszonylag egyszerű formái), és - a fizikai környezetben történő fizikai változtatásokhoz - kognitív robotika.⁵ Ezért minden TAI-ügynök képes tervezni, gondolkodni, tanulni, kommunikálni; továbbá kreatívak és képesek fizikai cselekvések végrehajtására.

D2 *Képes megoldani legalább olyan problémák fontos eseteit, amelyek a Turing által megoldhatatlan problémák szintjén és/vagy azok felett vannak.* A mai mesterséges intelligencia, ha képes problémákat megoldani, ezt a sikert mindig olyan problémákon éri el, amelyek pusztán algoritmikusan megoldhatóak és megoldhatóak (pl. dáma, sakk, Go).

D3 *Képes indokolni, magyarázni és tanúsítani a szolgáltatott megoldásokat, hogy hogyan jutottak hozzájuk, és hogy ezek a megoldások biztonságosak/etikusak.* Azt mondjuk tehát, hogy

⁴Egy ilyen léghajó egyszerű adaptációja annak, ami könnyen elérhető, mint egy viszonylag olcsó játék.

⁵A kognitív robotikát [Levesque és Lakemeyer, 2007] a robotika olyan típusaként definiálja, amelyben a robotok által végrehajtott minden érdemi cselekvés a robotok kognitív állapotainak (pl. be-liefs és szándékok) függvénye.

egy TAI-ügynök **racionalista**. Ez a címke azt a követelményt tükrözi, hogy a TAI-ügynök által felfedezett problémára javasolt minden megoldási javaslatot olyan indoklásnak kell kísélnie, amely megvédi és elmagyarázza, hogy a javasolt megoldás megoldás, és adott esetben azt is, hogy a megoldásnak (és esetleg a megoldás eléréséhez használt eljárásnak) vannak bizonyos kívánatos tulajdonságai. Az indoklásnak minimálisan tartalmaznia kell egy érvt vagy bizonyítékot a vonatkozó következtetésekre. Ezen túlmenően az indoklást formálisan ellenőrizni kell; így azt mondjuk, hogy **az igazolásról** egy TAI-ügynök gondoskodik.

- D4** *Képes "elméleti szintű" gondolkodásra, tervezésre és kommunikációra.* Ennek az attribútumnak a tárgyalását helytakarékosági okokból elhagyjuk; lásd például [Arkoudas and Bringsjord, 2009] a laboratóriumunk első, ezen a szinten történő automatizált gondolkodásra tett kísérletét. (Az igazság az, hogy pontosabb azt mondani, hogy a negyedik követelmény az, hogy egy TAI-ügynöknek *kognitív tudatossággal* kell rendelkeznie, mivel ezt a jelenséget a [Bringsjord et al., 2018] elmagyarázza és axiomatizálja.)
- D5** *Képes kreativitásra, minimálisan az úgynevezett m-kreativitás szintjén.* A mesterséges ágensek kreativitását és annak mérnöki tervezését már több helyen tárgyalta Bringsjord [Bringsjord and Ferrucci, 2000]pl., de nemrégiben Bringsjord és Sen [2016] a kreativitás egy formáját szorgalmazta a mesterséges ágensekben az I/IoT segítségével.
- D6** *"Csápos" hatalom, amelyet az I/IoT, az Edge Computing és a kibertérben használ.* Ez a TAI-ügynökök legfontosabb tulajdonsága, amelyet a "TAI" betűvel írt "T" is tükröz. Ha azt mondjuk, hogy az ilyen ágensek csápos problémamegoldó erővel rendelkeznek, akkor azt mondjuk, hogy képesek érzékelni és cselekedni az I/IoT (vagy azzal egyenértékű hálózatokon) és a kibertéren keresztül, az egész világon. A TAI-ügynökök tehát egy bolygóméretű, heterogén környezetben működnek, amely túllép a hagyományos, mai mesterséges intelligencia finomítására használt szűkebb, rögzített környezeteken, mint például a [Russell and Norvig, 2009].

2 Kapcsolódó munka

Tekintettel a jelen dolgozat korlátozott terjedelmére, csak néhány rövid megjegyzést teszünk a kapcsolódó munkákról, amelyek a kényelem érdekében felhasználhatók olyan munkákra, amelyek a TAI által megkövetelt kifejezőképesség és a kapcsolódó automatizált következtetések szintje felé vezető útnak tekinthetők, valamint olyan korábbi munkákra, amelyek éles és tanulságos kontrasztot képeznek a TAI-val. Először is, ami az általunk a TAI felé vezetőnek tekintett munkát illeti, megjegyezzük, hogy nemrég Miller és társai [2018] egy olyan tervezési keretrendszert mutatnak be, amelyet ők *szociális tervezésnek* neveznek, amelyben a vizsgált ágens úgy tervezhet és cselekedhet, hogy figyelembe veszi más ágensek meggyőződéseit. A társadalmi tervezésben az ágens célja lehet a külső világ egy adott állapota, vagy más ágensek hiedelmeinek egy halmaza (vagy a kettő keveréke). A rendszer egy tételes modális logika egyszerűsített változatára épül (ellentétben a mi alább bemutatott rendszerünkkel, amely kifejezőbb és összetettebb célokat is képes figyelembe venni),

pl. korlátlan tartományok feletti célok vagy olyan célok, amelyek nu-merikus kvantifikációt tartalmaznak; az ilyen kijelentésekhez túl kell lépni a tételes modális logikán). Ezen túlmenően a Wang [2006] NARS rendszere bizonyosan tartalmaz olyan elemeket, amelyek arányosan a TAI-val rokonszenvesnek tekinthetők. A NARS például többrétegű és érvelés-centrikus. Másrészt a "N" a "NARS"-ben a "nem-axiomatikus" szót jelenti, és a TAI, sőt a logikai mesterséges intelligencia egész megközelítése, amelyet legalább Bringsjord és Govindarajulu követ, arra törekszik, hogy amikor csak lehetséges, az auto-matikus gondolkodást erős axiómarendszerekkel szemben használja ki, mint például a Peano.

Aritmetika.⁶ Ráadásul a TAI mélyen és redukálhatatlanul intencionális, míg a NARS tisztán extenzionálisnak tűnik. A számítási erőforrások okos kezelése a TAI-ban egyértelműen kulcsfontosságú lesz, és Thorisson és munkatársai (pl. [Helgason *et al.*, 2012]) munkáját igen fontosnak tartjuk a TAI és a megvalósítás kihívásai szempontjából. A TAI-hoz általánosságban igazodó munka utolsó példajaként felhívjuk az olvasó figyelmét az informatikában a bizonyításlapú munka nemrégiben megjelent átfogó feldolgozására: [Arkoudas and Musser, 2017]. Mivel a TAI rendületlenül bizonyító-alapú mesterséges intelligencia, ez a kötet nagyon szépen bemutatja a TAI megvalósításához szükséges munkát.

Másodszor, a szemléletes kontraszt kedvéért először is megjegyezzük, hogy egyesek a több ágensből, köztük gépekből álló vállalati intelligencia koncepcióját tekintik, ahol az inspiráció a biológiából származik. Ennek egyik példája a [Seidita *et al.*, 2016] lenyűgöző modellezése.⁷ A mi esetünkben a TAI a földi biológiától független, alaposan formális koncepció, amely a jelenleg a Földön élő ágenseknél nagyobb intelligenciával rendelkező ágens típusokat hivatott magában foglalni. Egy másik megvilágító kontraszt a tervezésre létrehozott, tisztán extenzionális jellegű nyelvek (pl. a PDDL, amely korai formájában a [Mcdermott *et al.*, 1998]-ban található meg), mint olyanok figyelembevételén keresztül jön létre, amelyek ezért egészen mások, mint a TAI-hoz szükséges típusú tervezés, amelynek intencionális jellegűnek kell lennie, és az is (mivel a kognitív kalkulusok intencionális számítási logikák). Az MA-PDDL a PDDL egy kiterjesztése a több, különböző cselekvésekkel és célokkal rendelkező ágensekkel történő do- mainok kezelésére [Kovacs, 2012], és mint ilyen, relevánsnak tűnik a TAI szempontjából. A fentebb tárgyalt társadalmi tervezéssel ellentétben azonban az MA-PDDL célja nem más ágensek hiedelmeinek (és egyébként más episztemikus attitűdöknek) a megváltoztatása. Bár az MA-PDDL használható lenne erre, a meggyőződések és más kognitív állapotok PDDL kiterjesztő nyelvén történő reprezentálása nemkívánatos következményekhez vezethet, ahogyan azt a [Bringsjord and Govindarajulu, 2012] is bemutatta. Az eredeti PDDL (PDDL1) kiterjesztései, például a PDDL3 [Gerevini és Long, 2004], még mindig extenzionális jellegűek.

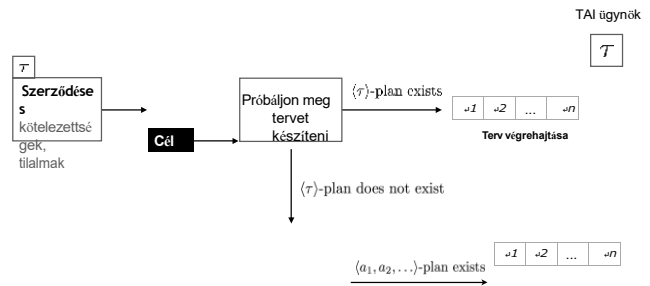
Ezzel zárul a kapcsolódó munkákról szóló rész. Megjegyezzük, hogy az alábbiakban a TAI-t a mesterséges intelligencia tervezése szempontjából írjuk le és definiáljuk.

3 Gyors áttekintés

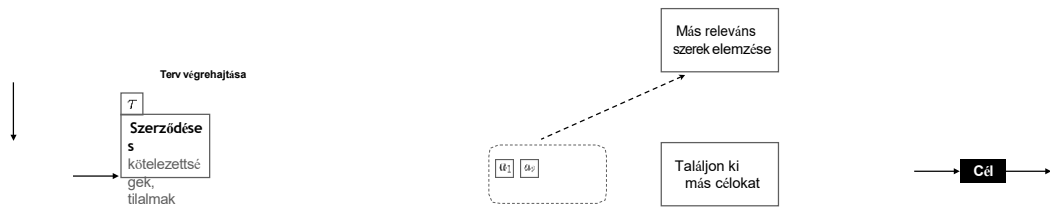
Gyors és informális áttekintést adunk a TAI-ról. Van egy a_1, \dots, a_n . Minden egyes ágenshez tartozik egy (implicit vagy explicit) szerződés, amelyet be kell tartania. Tekintsünk egy bizonyos \mathcal{T} ügynököt. Ennek az ágensnek az élete során,

⁶A TAI rétegződését valójában a [Bringsjord, 2015]-ben leírt, egyre erőteljesebb axióma-központú megismerés vetíti előre, amely a Peano Aritmetikát tekinti központi elemnek.

⁷Bár egyelőre elérhetetlen, mivel fő célunk csupán a TAI informatív bevezetése, a TAI-ügynökök számára központi szerepet játszó kognitív tudatosságról alkotott felfogásunk (fenti 4. attribútum) és a Chella által felfogott tudatosság közötti kapcsolat termékeny témája a jövőbeli vizsgálatoknak. A mesterséges tudatosság sokoldalú tárgyalása egyébként a [Chella and Manzotti, 2007] című könyvben található. A TAI-ügynökök számára központi szerepet játszó tudatosság márkájának első



axiomatizációját lásd [Bringsjord *et al.*, 2018].

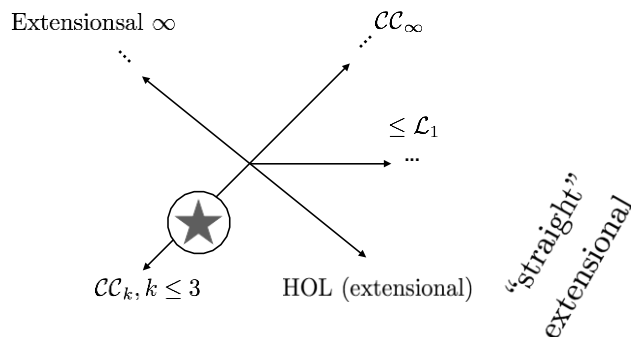


1. ábra: TAI informális áttekintés: A TAI-ügynök τ működésének felépítése. τ a szerződése alapján folyamatosan célokat tűz ki. Ha egy cél nem érhető el τ saját erőforrásaiból, akkor τ -nek más ágenseket kell igénybe vennie a cél eléréséhez. A sikeresen megtenni ezt τ -nek a D_1 - D_6 közül egy vagy több tényezővel kell rendelkeznie. tulajdonságok.

az ügynök olyan célokat tűz ki, amelyeket el kell érnie, hogy a szerződése ne sérüljön. E célok némelyike megkövetelheti, hogy az ágens a hat attribútum D_1 D_6 közül néhányat vagy mindegyiket gyakorolja.

ezt az 1. ábrán látható tervezéssel formalizáljuk. Ahogy az ábrán látható, ha egy cél egyedül nem érhető el, τ megpróbálhat más ágenseket toborozni, kihasználva azok erőforrásait, hiedelmek, kötelezettségek stb.

4 A hivatalos rendszer

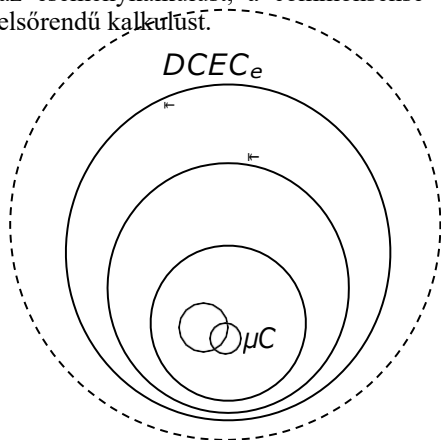


2. ábra: A logikai kalkulusok tere. A logikai kalkulusok teljes, hatalmas terét öt dimenzió fedi le. A nyugati dimenzióba tartoznak azok a kalkulusok, amelyek a szemantikus webet működtetik (és amelyek gen- Az első rendű logika = L_1), és magában foglalja az úgynevezett **leíró logikákat**. Mind az NW, mind az NE olyan logikai rendszereket tartalmaz, amelyekben a wff-ek végtelenül hosszúak lehetnek, és mondanom sem kell, hogy nehéz velük számolni, és nehezen számolhatók át. Az SE a magasabb rendű logika, amely köré egy robusztus, automatizált tételmegoldó közösség gyűlt össze. Ez az SW irányzat tartja a jelen dolgozatban leírt és a TAI-hoz kapcsolódó kognitív kalkulusokat; a csillag pedig azokra a konkrét kognitív kalkulusokra utal, amelyeket ezen oldalakon mi neveztünk meg.

A fenti fogalmak konkretizálása érdekében egy verzióörklési programozási nyelvet használunk. Az alábbiakban bemutatunk néhány egy számítási logika. Az általunk használt logika a **deontikus logika**. A **DCEC**-ben használt fontos fajták.

ive event calculus (DCEC). Ez a kalkulus egy elsőrendű modális logika. A 2. ábra azt a régiót mutatja, ahol a logikai kalkulusok teljes terében a **DCEC** a **DCEC** **lo-** **a** **cated**. logikai kalkulusok **kognitív kalkulusok** családjába tartozik (a 2. ábrán csillaggal jelölve, a 3. ábrán kibővítvé). jól definiált szintaxissal és következtetési rendszerrel rendelkezik; a teljes leírást lásd a [Govindarajulu and Bringsjord, 2017a] A. függelékében. A következtetési rendszer a természetes dedukción [Gentzen, 1935] alapul, és tartalmazza az elsőrendű logika összes bevezetési és megszüntetési szabályát, valamint a modális operátorok és a kapcsolódó struktúrák következtetési sémáit.

Ezt a rendszert korábban már használták [Govindarajulu and Bringsjord, 2017a; Govindarajulu et al., 2017] a kettős hatás tanának auto- mate változataihoz, amely egy deontológiai és konzekvencialista komponensekkel rendelkező etikai elv. Bár a kalkulus leírása meghaladja a dolgozat kereteit, az alábbiakban rövid áttekintést adunk a rendszerről. A dialektusait használták az erősen intencionális (azaz kognitív) érvelési folyamatok formalizálására és automatizálására is, mint például a hamis hit feladat [Arkoudas és Bringsjord, 2008] és az akrázia (engedni a kísértésnek, hogy megsértse az erkölcsi elveket) [Bringsjord et al., 2014]. Arkoudas és Bringsjord [2008] a **DCEC** hamisít-feladat formalizálásával vezette be a **kognitív eseménykalkulust** általános családját, amelyhez tartozik egy rendezett (tipizált) kvantifikált modális logika (más néven rendezett elsőrendű modális logika), amely magában foglalja az eseménykalkulust, a commonsense érveléshez használt elsőrendű kalkulust.



3. ábra: Kognitív számítások. A **kognitív kalkulák** családjába számos rokon kalkulusból áll. Arkoudas és Bringsjord a család első tagját, a **CEC**-et a hamis és a tévhit modellezésére vezették be. a **CEC** feladat. A család legkisebb tagja, a **muC**, már a kvantifikált meggyőződések bizonytalanságának modellezésére használják [Govindarajulu és Bringsjord, 2017b]. A **DCEC**-t és változatait az etikai elvek és elméletek, valamint azok megvalósításainak módosítása során használták.

4.1 Szintaxis

Mint fentebb említettük, egy rendezett számítás a rendezett rendszer analógjának tekinthető egy tipizált egyszálás számrendszerrel.

Rendezés	Leírás
ügynök	Emberi és nem emberi szereplők.
Idő	Az idő típus az időt jelöli a tartományban. Pl. egyszerű, mint t_i , vagy összetett, mint pl. <i>születésnap(son(jack))</i> .
Esemény	A tartományban bekövetkező eseményekre használatos.
ActionType	A művelettípusok absztrakt műveletek. Ezeket a szereplők bizonyos időpontokban instantiálják. Példa: evés.
Akció	Az Event altípusa az olyan eseményekre, amelyek az ügynökök által végrehajtott cselekvéseként fordulnak elő.
Folyékonyan	A világ állapotainak ábrázolására szolgál az eseményszámításban.

A szintaxis két összetevőből áll: egy elsőrendű magból és egy modális rendszerből, amely erre az elsőrendű magra épül. Az alábbi ábrák a szintaxis és a következtetési sémák.

Az elsőrendű magja az **eseménykalkulus** [Mueller, 2006]. Az eseménykalkulus általánosan használt függvény- és relációszimbólumai szerepelnek. Az eseményszámítás három fő fajtája a fluktuens, az esemény és az idő. A fluktuensnek a világ állapotait elsőrendű kifejezéseként reprezentálják. Az események olyan dolgok, amelyek a világban az idő meghatározott pillanataiban történnek. Az akciók olyan események, amelyeket egy ágens hajt végre. Bármely akció típus esetében

a és a ágens, az a végrehajtásához tartozó eseményt a $action(a, a)$ adja meg. Például ha a a "futás" és a "Jack", akkor az $action(a, a)$ a "Jack fut" kifejezést jelöli. A commonsense modellezésére szolgáló más kalkulusok (pl. a **helyzetkalkulus**)

és a fizikai érvelés könnyen kicserélhető az eseményszámítás helyett.

```

Szintaxis
is

S ::= Ügynök | ActionType | Action C Event | Moment |
Fluent action : Ügynök × ActionType →
    esemény : Ügynök × Moment →
    tartomány : Fluent × Moment →
    Formula : Esemény × pillanat → képlet
f ::=
    clipped : Moment × Fluent × Moment ×
    kezdeményez : Esemény × Fluens × Pillanat →
    Moment → Formula
    képlet : Esemény × Fluent × Moment →
    Formula Moment × Moment → Formula

t ::= x : S | c : S | f(t1, ..., tn)
q : Formula | ¬φ | φ ∧ ψ | φ ∨ ψ | ∀x : φ(x)

P(a, t, φ) | K(a, t, φ) |
φ ::= C(t, φ) | S(a, b, t, φ) | S(a, t, φ) | B(a, t, φ)
    D(a, t, φ) | I(a, t, φ)
    O(a, t, φ, (¬)happens(action(a*, a), t))
    
```

A kalkulusban szereplő modális operátorok közé tartoznak a **K** ismeret, **B** meggyőződés, **D** vágy, **I** szándék stb. standard operátorai. Az intencionális operátorok általános formátuma a következő

$K(a, t, \varphi)$, amely azt mondja, hogy a ügynök a t időpontban ismeri a φ proposíciót. Itt φ viszont lehet bármilyen tetszőleges formula. Vegyük észre a következő modális operátorokat is: **P** egy állapot észlelésére,

C a közös tudás, **S** az ügynök-ügynök közötti kommunikáció és a nyilvános közlemények, **B** a hit, **D** a vágy, **I** a szándékra, és végül, ami döntő fontosságú, egy dyadikus deontikus operátor **O**, amely kimondja, hogy egy cselekvés kötelező vagy tilos. ügynökök számára. $DCEC$ Meg kell jegyezni, hogy ez az egyik példány a bővíthető kognitív kalkulák családjában.

A kalkulák tartalmaz egy dyadikus (aritás = 2) deontikus operátor **O**. Jól ismert, hogy az unáris kellene a szabványos a deontikus logika ellentmondásokhoz vezet. A mi dyadikus változatunk az operátor blokkolja az ilyen ellentmondások szabványos listáját, és azon túl.⁸

A φ deklaratív közlése a és b között a t időpontban az $S(a, b, t, \varphi)$ segítségével ábrázoljuk.

4.2 Következtetési sémák

Az alábbi ábra a következtetési sémák egy részletét mutatja. a $DCEC$ számára. Első rendű természetes következtetés bevezetése és

a kizárási szabályok nem jelennek meg. Az I_K és az I_B következtetési sémák segítségével olyan idealizált rendszereket modellezhetünk, amelyek tudása és hiedelmei a bizonyítási elmélet szerint $DCEC$ Míg a hu-

A két szabály lehetővé teszi számunkra, hogy közelebről modellezzük, hogyan gondolkodnak az olyan tudatosabb ágensek, mint a szervezetek, nemzetek és stratégiai szereplők. (A kognitív kalkulások néhány di-alektusa korlátozza az iterációk számát a

intencionális operátorok.) I_{13} a szándékokat közvetlenül az észlelésekhez köti (Ez a modell nem veszi figyelembe azokat az ágenseket, akik esetleg nem hajtják végre a szándékaikat). I_{14} diktálja, hogy a kötelezettségeket hogyan fordítják le ismert szándékokká.

Következtetési sémák (töredék)

$$\frac{K(a, t_1, \Gamma), \Gamma \vdash \varphi, t \leq t_2}{K(a, t_2, \Gamma)} [I_K]$$

$$\frac{B(a, t_1, \Gamma), \Gamma \vdash \varphi, t \leq t_2}{B(a, t_2, \Gamma)} [I_B]$$

$$\frac{K(a, t, \varphi)}{[I_4]} \quad \frac{t < t', I(a, t, \varphi)}{[I_{13}]}$$

$$\frac{B(a, t, \varphi) \quad B(a, t, O(a, t, \varphi)) \quad O(a, t, \varphi, \chi)}{K(a, t, I(a, t, \chi))} [I_{14}]$$

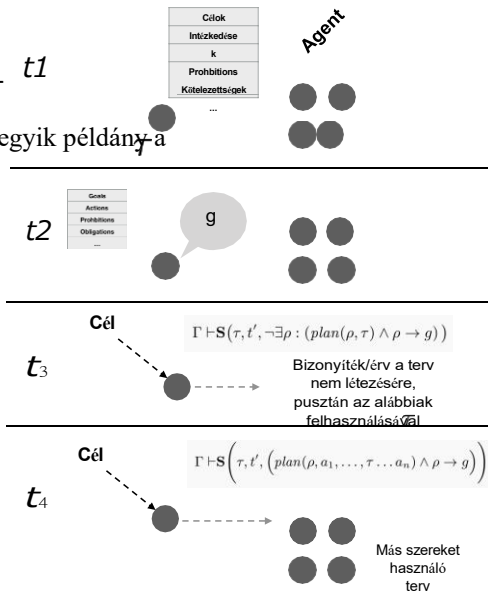
4.3 Szemantika

Az elsőrendű töredék szemantikája a standard elsőrendű szemantika. Az igazság-funkcionális kötőszavak $\wedge, \vee \rightarrow$ és a kvantorok \forall, \exists tisztán elsőrendű formulákhoz mindegyike rendelkezik

a szabványos elsőrendű szemantika. A szemantika a modális operátorok sok fontos dologban különböznek az úgynevezett Belief-Desire-Intention (BDI) logikákban [Rao és Georgeff, 1991] rendelkezésre álló $DCEC$ kifejez

etten elutasítja a lehetséges-világok szemantikát és a modell-alapú reasoningot, helyette a bizonyításelméleti szemantikát és a $natu$ -nak nevezett érvelési típushoz kapcsolódóan.

rális következtetés [Gentzen, 1935; Francez és Dyckhoff, 2010]. Röviden, ebben a megközelítésben a modális operátorok jelentését a bizonyítások feletti tetszőleges



4. ábra: TAI munka az időben. Egy TAI-ügynök kezdetben egy célt vizsgál, majd bizonyítékot kell szolgáltatnia arra, hogy nem létezik egy olyan nem-tematikus terv, amely csak ezt az ügynököt használja. Ezután τ toboroz egy sor más releváns ágens, hogy segítsenek a céljának elérésében.

5 A TAI meghatározása

A t időpontban fennálló állapotokat a $\Gamma(t)$ formulák halmazával jelöljük. Ez a képlethalmaz tartalmazza a különböző ágensekre vonatkozó kötelezettségeket és tilalmakat is. Minden egyes ágens esetében

a_i t időpontban létezik egy szerződés (a_i, t) $\Gamma(t)$, amely leírja a_i kötelezettségeit, tilalmait stb. a bármely t időpontban előáll egy g céllal, hogy a szerződése teljesüljön.⁹ Az ágens úgy véli, hogy ha g nem teljesül, akkor a szerződését egy bizonyos jövőbeli $t + \delta$ időpontban megszegi:

$$B(a, t, \neg g) \rightarrow \neg \wedge c(a, t + \delta)$$

Ezután az ágens megpróbál egy olyan tervet kitalálni, amely a célt kielégítő cselekvések sorozatát tartalmazza.

Ezeket a fogalmakat pontosítjuk. Egy a ágensnek van egy sor olyan cselekvése, amelyet különböző időpontokban végezhet el.

Például egy porszívózó ügynök lehetséges műveletei között szerepelhet a sík mentén történő mozgás, míg egy telefonon lévő ügynök művelete lehet egy értesítés megjelenítése. Ezt jelöljük

$can(a, a, t)$ a következő kiegészítő axiómával:

$$\text{Axióma } \neg \text{lehet}(a, a, t) \rightarrow \neg \text{történik}(\text{action}(a, a), t)$$

Az alábbiakban egy konzisztens tervet határozunk meg:

Következő terv

Egy konzisztens $\rho(a_1, \dots, a_n)$ terv t időpontban az a_1, \dots, a_n , a megfelelő a_1, \dots, a_n és t_1, \dots, t_n úgy, hogy $\Gamma \vdash (t < t_i < t_j)$ az $i < j$ és az összes

számításokon keresztül határozzuk meg.

⁸E lista áttekintését a [McNamara, 2010] tartalmazza.

⁹Lásd [Govindarajulu and Bringsjord, 2017a] egy példát arra, hogyan használhatók a kötelezettségek és tilalmak a *DCEC-ben*.

a_i ügynökökkel rendelkezünk:

1. $can(a_i, a_i, t)$
2. $happens(action(a_i, a_i))$ konzisztens a $\Gamma(t)$ -vel.

Megjegyezzük, hogy egy konzisztens ρ ... terv a nyelvünkön egy kifejezéssel reprezentálható. Bevezetünk egy új fajtát $Plan$ és egy változó arítású predikátumszimbólumot $plan(\rho, a_1, \dots, a_n)$, amely azt mondja, hogy ρ egy olyan terv, amely magában foglal egy_1, \dots, egy_n .

A cél szintén bármely g képlet. Egy konzisztens terv akkor felel meg egy g célnak, ha:

$$\left(\Gamma(t) \cup \begin{matrix} \mathbf{f} \\ happens(action(a_1, a_1), t_1), \dots \\ happens(action(a_n, a_n), t_n) \end{matrix} \right) \mathbf{f} \dashv g$$

A $\Gamma(\rho, g)$ a fentiek rövidítéseként használjuk. A tervek és célok fenti definíciói a következő im-

a TAI meghatározásához szükséges fontos megkötés. Ez megkülönbözteti a mi tervezési formalizmusunkat más tervezési rendszerektől, és megfelelőbbé teszi egy általános célú, sátoros mesterséges intelligencia rendszer számára.

Egységes tervezési kényszer

A terveket és a célokat a tervezési rendszer nyelvén kell megjeleníteni és átgondolni.

A fenti követelményt kihasználva a TAI-ügynökök két szintjét határozhatjuk meg. Az(1) szintű TAI-rendszer megfelel a következőknek

egy τ ágens egy olyan rendszer, amely a g céllal jön létre t időpontban a szerződésének teljesítése érdekében, bizonyítékot állít elő arra, hogy nincs konzisztorterv, amely csak a τ ügynököt érinti. Ezután τ egy olyan tervvel érkezik, amely egy vagy több más ágenszt is érint. Egy Level(1) TAI ágens azzal a tudással indul, hogy milyen cselekvések lehetségesek más szerek esetében.

Szint(1) TAI-ügynökök

Feltétel Bármely a, α, t esetén:

$$\Gamma \mathbf{f} \dashv can(a, \alpha, t) \rightarrow \mathbf{Kct}, t^t, can(a, \alpha, t)$$

Akko

1. τ bizonyítékot állít elő arra, hogy g -re nem létezik olyan terv, amely csak önmagát foglalja magában, és τ kijelenti, hogy nincs ilyen terv.
2. τ létrehoz egy tervet g -re, amely csak saját magát és egy vagy több ágenszt foglal magában, és ezt a tervet bejelenti.

$$\Gamma \mathbf{f} \dashv S \tau, tt, plan(\rho, a_1, \dots, \tau \dots a_n) \wedge \rho \rightarrow g$$

ügynökök mit tudnak tenni. A TAI-ügynököknek tökéletlen ismeretei lehetnek más ügynökökről. Az ágens más ágensek cselekvéseiről, kötelezettségeikről, tilalmaikról stb. a megfigyelésük vagy az ágensekre vonatkozó előírások elolvasása révén szerezhethet információt. Ebben az esetben egy (2. szintű) TAI-ügynököt kapunk.

ügynök. Csak az előfeltételt kell módosítanunk. fent.

Szint(2) TAI-ügynökök

Feltétel Bármely a, α, t esetén:

$$\Gamma \mathbf{f} \dashv can(a, \alpha, t) \rightarrow \mathbf{B} \tau, t^t, can(a, \alpha, t)$$

táka fenti ügynökök közhelyrendű magasabbrendű TAI-ügynökünk is, amely szándékosan olyan cselekvéseket hajt végre, amelyek egy vagy több más ügynököt indítanak el, hogy a fent leírtak szerint csápos módon cselekedjenek. Az egységes tervezési kényszer szükségessége világosabb, ha magasabb rendű ágenseket vizsgálunk.

6 A TAI-ügynökök hierarchiája

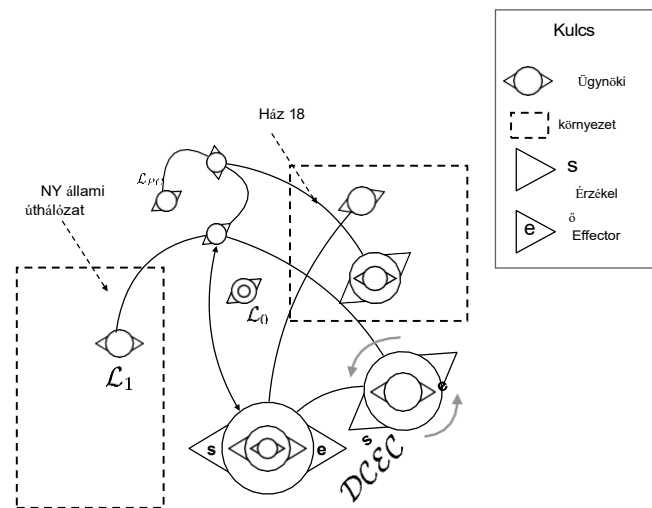
A fenti TAI formalizáció a csápos ágensek többféle hierarchiáját eredményezi. Ezek közül néhányat tárgyalunk be- low.

Szintaktikai cél összetettsége A g cél összetettsége az egyszerű kijelentésektől kezdve, pl. *cleanKitchen*, az elsőrendű kijelentésekre. pl. $\forall r : Room : clean(r)$, és más kognitív állapotokat reprezentáló intenzionális kijelentésekre. ügynökök

$$\mathbf{B}(a, most, \mathbf{B}(b, most, \forall r : clean(r)))$$

Célvariáció A fenti definíció szerint egy a ágens akkor minősül sátorosnak, ha csak egy g célt tervez sátoros módon a fenti feltételekben meghatározottak szerint. Lehetnek olyan ágenseink, amelyek számos változatos és különböző célt terveznek csápos módon.

A terv bonyolultsága Sok cél esetében általában több terv létezik, amelyek különböző cselekvéseket tartalmaznak (különböző költségekkel és újraforrásokkal), és amelyeket különböző ágensek hajtanak végre.



5. ábra: Képes áttekintés. Egy kis magyarázat: Az, hogy egyes ágensek ágenseken belül vannak, azt jelzi, hogy a külső ágens mindent tud és/vagy be- kedik, ami a belső ágensről releváns; ezért ahogy az ágensek kognitívan egyre erősebbek, úgy növekszik episztemikus attitűdjük mélysége (ez iterált hit/tudás operátorokkal ellátott képletekben tükröződik). Az ágensek mérete/intelligenciája a logikai kalkulák méretével/intelligenciájával párhuzamosan nő, amelyeken alapulnak, és amelyek kifejezőképessége és intelligenciája növekszik.

érvelési képesség; $L0$ a nulladik rendű logika, $L1$ pl. az elsőrendű logika, és a $DCEC$ sajátos kognitív kalkulusz. A rotáció egyszerűen azt jelenti, hogy az ágensek az idő múlásával érzékelnek és cselekszenek.

7 Példák és embrionális megvalósítás

Ebben a szakaszban egy TAI-ügynök formális vázlatát mutatjuk be, majd egy másik példán keresztül ismertetjük a TAI-rendszer megvalósításán folyó munkát.

7.1 Példa

Vegyük az elején említett példát. Van egy j ember és három mesterséges ágens: egy_c az autóban, egy_h az otthonban. és a_p egy ügynök, aki az ütemezési és naptárinformációkat kezeli...
tion. A képletek közül néhányat a Γ -ben mutatunk be.

$B_{(ac, to, crowded(store) \rightarrow unusal), f1}$

$P_{(ac, t1, crowded(store)), f2}$

$\forall t : O_{ac, t, unusal, történi action(ac, check(weather)), t + 1}) f3$

$\forall t : B_{a_c, t, történi action(a, check(weather)), t3} f4$

$\forall a : \rightarrow K(a, t4, vihar),$

$\forall t : O_{ca_c, t, vihar, S(ac, a_h, vihar, t + 1)}, f5$

A fenti képletek először is megállapítják azt a tényt, hogy a_c megfigyeli, hogy a bolt zsúfolt. a_c 'szerződésében szerepel, hogy az ügynöknek meg kell néznie egy időjárás-jelentést, ha valami szokatlan észlel. A

képlet azt is kimondja, hogy ha egy ügynök t_3 időpontban ellenőrzi az időjárást, akkor az ügynök előrejelzést kap egy közelgő viharról. a_c szerződés kötelezi arra, hogy tájékoztassa a_h , ha úgy véli, hogy vihar közeledik.

$\forall t : O_{a_h, t, vihar, \forall s : mennyiség(s) > 0, f_6}) f7$
 $a_h, t_5, shops(j, today) \vee shops(j, tomorrow)$

$K_{\rightarrow \forall s : mennyiség(s) > 0, a_h, t, happensaction(ac, recc(shops(j))), t} f_8$

$\rightarrow \text{üzletek}(j)$
 $\forall t : B_{a_h, t, happens action a_h, req(ac, shops(j)), t} f_9$
 $\rightarrow történi action(a, recc(shops(j))), t$

A fenti első képlet szerint a_h honlapnak gondoskodnia kell arról, vihar esetén a készletek fel vannak töltve. Aztán a_h tudja, hogy a j ember ma vagy holnap vásárolhat, ami azt eredményezheti, hogy a készleteket feltöltik. a_h azt az információt kapja a_p , hogy a holnapi vásárlás nem lehetséges (ez a képlet nem látható). Ezután, vannak képleteink, amelyek a következő hatásokat határozzák meg

Ezt megelőzően számos olyan szerződést kötöttek, amelyek a felnőtt szülők P_1 és P_2 egy H otthonban kötődnek, és számos mesterséges ágens is kötődik H -ban, beleértve egy TAI-ügynököt (T), amely felügyeli az otthont. (Szigorúan véve az ágensek nem kötődnek szerződéseket, de tudnak, hogy az emberi gazdáik ezt tették, és ők tudnak, hogy mik a szerződések.)

Tél van Berlinben, New Yorkban. Éjszaka. Odakint hóvihár. A házi H édesanyja és édesapja, valamint két kisgyermekük mélyen alszik. Az okostelefon minden szülőnél "Ne zavarjon" beállítás van, és csak a közeli családtagok számára van engedélyezve a belépés. Nincs vezetékes telefon. A szén-monoxid-érzékelő az alagsorban, a kemence közelében hirtelen megemelkedett szintet jelez, ami a következő lépésekben mutatkozik meg.

küszik felfelé. T érzékeli ezt, és hipotéziseket alkot arról, hogy mi okozza az emelkedett értéket, és believes alapján egy olyan kognitív számítás alapján, hogy a leolvasás pontos (bizonyos valószínűségi tényező).

A legközelebbi tűzoltószertárat a T jelzi. Nincs a házban lévő riasztás diagnosztikai és aknákat, hogy a központi hangjelző elem lemerült. Az érték egyre magasabbra küszik, és most már még az érzékelők az emeleti hálószobákban, ahol az emberek alszanak, emelkedő és emelkedő szintet mutatnak. T ezt is érzékeli.

Sajnos, T okok miatt, mire a tűzoltók kikerkeznek, maradandó neurológiai károsodás vagy akár halál is bekövetkezhet.

(ismét valószínűségi tényezőre van szükség) a család egy vagy több tagja esetében. Amennyiben a riasztó cég úgy programozta be az érzékelőt, hogy az egy központi parancsnak jelentsen, akkor is, minden emberi parancs hibás. A vállalat lehet, hogy negligens, vagy a telefonhívás lehet az egyetlen rendelkezésükre álló lehetőség, vagy olyan személyzetet küldenek ki, akik túl későn érkeznek. Más mesterséges ágensek segítségének igénybevétele nélkül a tervezésben és a

érvelés, T nem tudja megmenteni a családot; T tudja ezt az alapján, hogy a bizonyítás/érvelés. A T azonban valószínűleg felébresztheti a családot, kezdve a

a szülők, bármilyen módon. Azonban ezek mindegyike legalább egy olyan jogi tilalom megsértését vonja maga után, amelyet a hatályos szerződések hoztak létre. Ezek a szerződések elmentek egy IBM szolgáltatás, amely feltöltötte az elméleti DCEC jogi kötelezettségen ismerő T -nek - vagy inkább

egy olyan nyelvjárásban, amely külön kötelezettség-operátorokkal rendelkezik a jogi

O_l és erkölcsi O_m kötelezettségek. Az erkölcsi kötelesség a mentésre

egy_c , amely a vásárlási akciót ajánlja j -nek. A_h célja $s : mennyiség(s) > 0$, és egy_h , egy_c és j felhasználásával felépül egy_c terv.

7.2 A végrehajtás felé

Leírunk egy példaforgatókönyvet, amelyet egy embrionális

megvalósításhoz tervezünk.

a család azonban felülírja a törvényi tilalmakat. 7 fordulatok a tévé a hálószobában maximális hangerőn, és figyelmeztetést villant fel, hogy azonnal hagyják el a házat, mert halálos gáz gyúlik fel. (Természetesen sok más alternatíva is létezik. A TAI áttörhetne a Ne zavarjanak, pl.).

7.3 Az intelligens városi infrastruktúra használata felé

Az intelligens városokra vonatkozó európai kezdeményezés [eur, 2018] az Európai Bizottság [ec, 2018] által létrehozott kezdeményezés, amelynek célja az életminőség javítása Európa-szerte, miközben az energia- és éghajlatvédelmi célkitűzések felé halad. Céljai közül sokan a világ egészségét tekintve is relevánsak és kívánatosak. A TAI-nak megvan a lehetősége arra, hogy ezek közül soknak az elérésében szerepet játsszon, mint pl.

mint az intelligens készülékek (az előző alfejezetben tárgyalt módon) és az intelligens forgalomirányítás. Valójában a kezdeményezés alkalmazási köre és célkitűzései a TAI széleskörű alkalmazásával jelentősen kiszélesedhetnek.

Röviden bemutatunk egy egyszerű forgatókönyvet, amely kibővíti az európai kezdeményezés intelligens közlekedési céljainak vízióját.

A parkolóhelyek nagyon szűkösek egy munkanapon Manhattan közepén. Egy elfoglalt vezetőnek a nap folyamán több iroda közelében kell parkolnia, és ezek a helyek a hét folyamán változnak.

Az ügyvezető autója a naptárát böngészi. A múltbeli minták alapján interpolálja azokat a helyeket, ahol a nő szerinte parkolni szándékozik. Kommunikál az ezeken a helyeken parkoló többi autóval, és a kifejezett (és kikövetkeztethető) szándékaik és aktuális helyük alapján meghatározza, hogy a tulajdonosok mikor fognak valószínűleg újra megfordulni. Az ügyvezetőnk tartózkodási helyét, a forgalmi viszonyokat és a napirendjében bekövetkező változásokat figyelembe véve dinamikusan határozza meg az optimális parkolási helyeket, a forgalmas nap folyamán. Természetesen a TAI szellemében az összes többi autó mozgása is ennek megfelelően módosulna az idő múlásával.¹⁰

8 Következtetés és jövőbeli munka

Bevezettük a Tentacular AI-t és annak számos architektúrális elemét, és nincs illúziónk arról, hogy ennél többet elértünk volna. A 2018-as AEGAP-on a TAI-t a fent vázolt mindkét forgatókönyvben bemutatjuk; a megvalósítás jelenleg is folyamatban van. A TAI-kutatási program kezdeti állapota ellenére reméljük, hogy ígéretes, bár kezdetleges áttekintést adtunk a csápos mesterséges intelligenciáról.

- egy olyan áttekintés, amely - tekintettel a rendkívül kifejező nyelvek központi szerepére az újszerű tervezés és következtetés szempontjából - reméljük, hogy a "dolgok internetének" és a mesterséges intelligenciával való vibráló kereszteződésének hajnalán néhányan, sőt talán sokan érdeklődnek iránta.

9 Köszönetnyilvánítás

A TAI projektet az RPI és az IBM közös támogatása tette lehetővé az AIRC program keretében; hálásak vagyunk ezért a támogatásért. Az itt bemutatott kutatások egy részét az ONR és az AFOSR támogatása tette lehetővé, amiért szintén hálásak vagyunk.

Hivatkozások

[Arkoudas és Bringsjord, 2008] K. Arkoudas és

S. Bringsjord. A józan ész pszichológiájának formalizálása felé: An Analysis of the False-Belief Task. In T.-B. Ho és Z.-H. Zhou, szerkesztők, *Proceedings of the Tenth Pacific Rim International Conference on Artificial Intelligence (PRICAI 2008)*, number 5351 in Lecture Notes in Artificial Intelligence (LNAI), pages 17-29. Springer-Verlag, 2008.

fel, amelyek megoldhatók **differenciális adatvédelemmel** [Dwork, 2008] vagy **zéró-tudáson** alapuló **bizonyítékokon** alapuló adatvédelemmel [Gehrke et al., 2011].

¹⁰Az ilyen TAI-alkalmazások adatvédelmi aggályokat vetnek

- [Arkoudas és Bringsjord, 2009] K. Arkoudas és S. Bringsjord. Propositional Attitudes and Causation. *International Journal of Software and Informatics*, 3(1):47-65, 2009.
- [Arkoudas és Musser, 2017] Konstantine Arkoudas és David Musser. *Alapvető bizonyítási módszerek az informatikában: A Computer-Based Approach*. MIT Press, Cambridge, MA, 2017.
- [Bringsjord és Ferrucci, 2000] S. Bringsjord és D. Ferrucci. *Mesterséges intelligencia és irodalmi kreativitás: Inside the Mind of Brutus, a Storytelling Machine*. Lawrence Erlbaum, Mahwah, NJ, 2000.
- [Bringsjord és Govindarajulu, 2012] S. Bringsjord és N. S. Govindarajulu. Mi is az intelligencia valójában a világhálón? *Metafilozófia*, 43(4):361-532, 2012. Ez az URL a cikk előnyomatására mutat.
- [Bringsjord és Sen, 2016] Selmer Bringsjord és Atriya Sen. On Creative Self-Driving Cars: Hire the Computational Logicians, Fast. *Applied Artificial Intelligence*, 30:758-786, 2016. Az URL itt csak egy javítatlan preprintre vezet.
- [Bringsjord *et al.* , 2014] Selmer Bringsjord, Naveen Sundar Govindarajulu, Daniel Thero és Mei Si. Akratikus robotok és azok számítási logikája. In *Proceedings of ETHICS 2014 (2014 IEEE Symposium on Ethics in Engineering, Science, and Technology)*, pages 22-29, Chicago, IL, 2014. IEEE katalógusszám: CFP14ETI-POD.
- [Bringsjord *et al.* , 2018] S. Bringsjord, P. Bello, and N.S. Govindarajulu. A tudatosság axiomatizálása felé. In D. Jacquette, szerkesztő, *The Bloomsbury Companion to the Philosophy of Consciousness*, 289-324. oldal. Bloomsbury Academic, London, Egyesült Királyság, 2018.
- [Bringsjord, 2015] Selmer Bringsjord. Tétel: T. Besold: Az általános intelligencia magában foglalja a kreativitást, feltételezve In T. Besold, M. Schorlemmer és A. Smaill, szerkesztők, *Computational Creativity Research: Towards Creative Machines*, pages 51-64. Atlantis/Springer, Párizs, Franciaország, 2015. Ez a kötet az *Atlantis Thinking Machines* 7. kötete, szerkesztette Kuhnbergwer, Kai-Uwe az Osnabrucki Egyetemről, Németország.
- [Chella és Manzotti, 2007] Antonio Chella és Ricardo Manzotti, szerkesztők. *Mesterséges tudatosság*. Imprint Academic, Exeter, Egyesült Királyság, 2007.
- [Dwork, 2008] Cynthia Dwork. Differenciális magánélet: A Survey of Results. In *International Conference on Theory and Applications of Models of Computation*, pages 1-19. Springer, 2008.
- [ec, 2018] Az Európai Bizottság prioritásai. https://ec.europa.eu/commission/index_en, 2018. [Online; hozzáférés: 2018. június 25.].
- [eur, 2018] Európai kezdeményezés a oldalon.
- <https://setis.ec.europa.eu/set-plan-implementation/technology-roadmaps/european-initiative-smart-cities>, 2018. [Online; hozzáférés: 2018. június 25.].

- [Francez és Dyckhoff, 2010] Nissim Francez és Roy Dyckhoff. Proof-theoretic Semantics for a Natural Language Fragment. *Linguistics and Philosophy*, 33:447-477, 2010.
- [Gehrke *et al.*, 2011] Johannes Gehrke, Edward Lui és Rafael Pass. Towards Privacy for Social Networks: A Zero-Knowledge Based Definition of Privacy. In *Theory of Cryptography Conference*, pages 432-449. Springer, 2011.
- [Genesereth és Nilsson, 1987] M. Genesereth és N. Nilsson. *A mesterséges intelligencia logikai alapjai*. Morgan Kaufmann, Los Altos, CA, 1987.
- [Gentzen, 1935] Gerhard Gentzen. Investigations in Logical Deduction. In M. E. Szabo, szerkesztő, *The Collected Papers of Gerhard Gentzen*, pages 68-131. North-Holland, Amsterdam, Hollandia, 1935. Ez a közismert 1935-ös német változat angol nyelvű változata.
- [Gerevini és Long, 2004] Alfonso Gerevini és Derek Long. Tervkorlátozások és preferenciák a PDDL3-ban. Technikai jelentés, Bresciai Egyetem, Automatizálási elektronika tanszék, 2004. Ez az ötödik nemzetközi tervezési verseny nyelve.
- [Govindarajulu és Bringsjord, 2017a] Naveen Sundar Govindarajulu és Selmer Bringsjord. On Automating the Doctrine of Double Effect. In Carles Sierra, editor, *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 4722-4730, Melbourne, Australia, 2017. Preprint elérhető ezen az url-en: <https://arxiv.org/abs/1703.08922>.
- [Govindarajulu és Bringsjord, 2017b] Naveen Sundar Govindarajulu és Selmer Bringsjord. Erősségtényezők: An Uncertainty System for a Quantified Modal Logic, 2017. Előadás a Workshop on Logical Foundations for Uncertainty and Machine Learning at IJCAI 2017, Melbourne, Ausztrália.
- [Govindarajulu *et al.*, 2017] Naveen Sundar Govindarajulu, Selmer Bringsjord, Rikhiya Ghosh, and Matthew Peveler. A kettős hatás doktrínáján túl: A valódi önfeláldozás formális modellje. Nemzetközi konferencia a robot etikáról és biztonsági normákról, 2017.
- [Gupta, 1992] Naresh Gupta. A tömb-világtervezés komplexitásáról. *Artificial Intelligence*, 52:223-254, 1992.
- [Helgason *et al.*, 2012] Helgi Helgason, Eric Nivel, and Kristinn Thórisson. Az AGI-architektúrák figyelemmechanizmusairól: A Design Proposal. In J. Bach, B. Goertzel, and M. Ikle', editors, *Proceedings of the Fifth Conference on Artificial General Intelligence*, pages 89-98, Berlin, Germany, 2012. Springer.
- [Kovacs, 2012] Daniel L. Kovacs. A PDDL3.1 többügynökös kiterjesztése. In *Proceedings of the 3rd Workshop on the International Planning Competition (IPC), ICAPS- 2012*, pages 25-29, Atibaia, Brazília, 2012. ICAPS.
- [Levesque és Lakemeyer, 2007] Hector Levesque és Gerhard Lakemeyer. 24. fejezet: Kognitív robotika. In *Handbook of Knowledge Representation*, Amsterdam, Hollandia, 2007. Elsevier.
- [Luger, 2008] George Luger. *Mesterséges intelligencia: Struktúrák és Stratégiák a Komplex Problémaoldáshoz (6. kiadás)*. Pearson, London, Egyesült Királyság, 2008.
- [Mcdermott *et al.*, 1998] D. Mcdermott, M. Ghallab, A. Howe, C. Knoblock, A. Ram, M. Veloso, D. Weld, and D. Wilkins. PDDL - A tervezési terület meghatározási nyelve. Műszaki jelentés CVC TR-98-003, Yale Center for Computational Vision and Control, 1998.
- [McNamara, 2010] P. McNamara. Deontikus logika. In Edward Zalta, editor, *The Stanford Encyclopedia of Philosophy*. 2010. McNamara (rövid) jegyzete a Kant-törvényből eredő paradoxonról a főbejegyzés egyik mellékszálában található.
- [Miller *et al.*, 2018] Tim Miller, Adrian R. Pearce és Liz Sonenberg. *Social Planning for Trusted Autonomy*, 67-86. old. Springer International Publishing, Cham, 2018.
- [Mueller, 2006] E. Mueller. *Commonsense Reasoning: An Event Calculus Based Approach*. Morgan Kaufmann, San Francisco, CA, 2006. Ez a könyv első kiadása. A második kiadás 2014-ben jelent meg.
- [Rao és Georgeff, 1991] A. S. Rao és M. P. Georgeff. Racionális ágensek modellezése egy BDI-architektúrán belül. In R. Fikes és E. Sandewall, szerkesztők, *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, pages 473-484, San Mateo, CA, 1991. Morgan Kaufmann.
- [Russell és Norvig, 2009] S. Russell és P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice Hall, Upper Saddle River, NJ, 2009. Harmadik kiadás.
- [Seidita *et al.*, 2016] Valeria Seidita, Antonio Chella, and Maurizio Carta. Egy egyetemi campus intelligenciájának biológiailag inspirált reprezentációja. *Procedia Computer Science*, 88:185-190, 2016.
- [Wang, 2006] Pei Wang. *Merev rugalmasság: Az intelligencia logikája*. Springer, Dordrecht, Hollandia, 2006. Ez a könyv a Dov Gabbay és Jon Barwise által szerkesztett *Applied Logic Series* 34. kötete.

