

Alkalmazott mesterséges intelligencia (AMI)

<http://www.mit.bme.hu/oktatas/targyak/vimibb01>

1 ea. Bevezetés a mesterséges intelligenciáról

Dobrowiecki Tadeusz előadásának felhasználásával



<https://www.esrcheck.com/2023/06/05/artificial-intelligence-ai-experts-sign-statement-on-ai-risk/>

Előadó: Pataki Béla

BME I.E. 414, 463-26-79

pataki@mit.bme.hu,

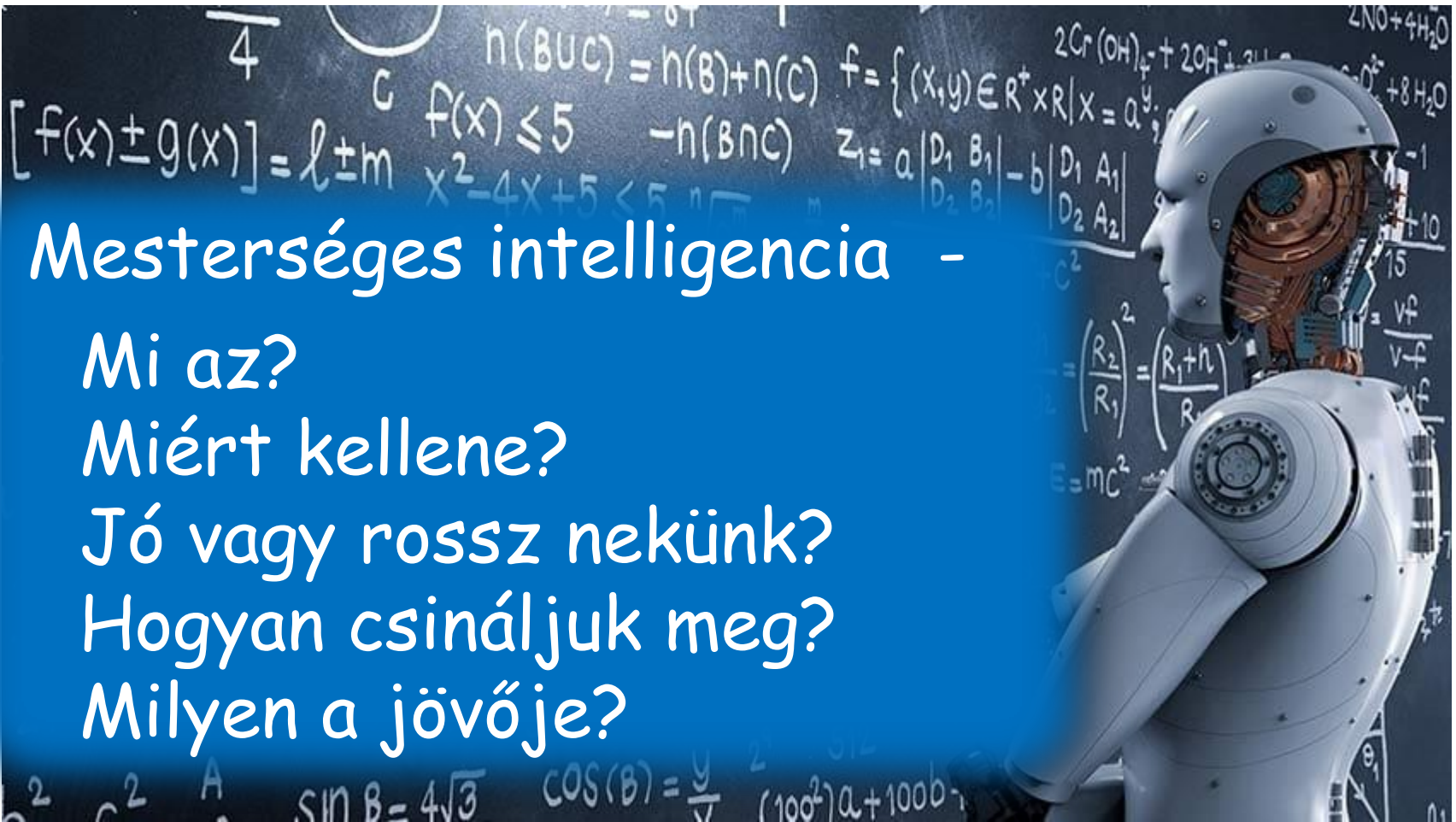
<http://www.mit.bme.hu/general/staff/pataki>

ROSSZ HÍR:

Ez nem a Roxfort Boszorkány- és
Varázslóképző Szakiskola !

(algoritmusok, megfontolások, matek....)

Bevezető előadás anyaga



Mesterséges intelligencia -

- Mi az?
- Miért kellene?
- Jó vagy rossz nekünk?
- Hogyan csináljuk meg?
- Milyen a jövője?



**INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL
INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION**



**A DEFINITION OF AI:
MAIN CAPABILITIES AND DISCIPLINES**

The term AI contains an explicit reference to the notion of intelligence. However, *since intelligence (both in machines and in humans) is a vague concept*, although it has been studied at length by psychologists, biologists, and neuroscientists, AI researchers use mostly the notion of *rationality*. This refers to the ability to choose the best action to take in order to achieve a certain goal, given certain criteria to be optimized and the available resources. Of course, rationality is not the only ingredient in the concept of intelligence, but it is a significant part of it.

Az MI kifejezés kifejezetten utal az intelligencia fogalmára. Mivel azonban az *intelligencia (mind a gépeknél, mind az embereknél) homályos fogalom*, bár pszichológusok, biológusok és neurológusok hosszasan tanulmányozták. Az MI-kutatók többnyire a *racionalitás* fogalmát használják. Ez arra utal, hogy bizonyos optimalizálható kritériumok és a rendelkezésre álló erőforrások mellett meg lehet választani a legjobb lépést egy bizonyos cél elérése érdekében. Természetesen a racionalitás nem az egyetlen összetevője az intelligencia fogalmának, de ennek jelentős része.

Mi az „intelligens” jelentése?



Nem tud kunyhót építeni, nem képes megtalálni ehető gyökereket, és az időjárást egyáltalán nem képes megjósolni ...
Magyarán: **pocsékul** teljesített az IQ tesztünkön.

Miért jó egy „intelligens” rendszer?

Intelligens rendszer (reménységeink szerint):

- ügyes, többet tud, **tanulékony**,
- **adaptív**, **robusztus**, **autonóm**,
- **tud beszélni** (és megérteni emberi beszédet),
- **kooperál** (együttműködik),
- **el tudja magyarázni**, hogy mit miért tesz, (?) ...

Ezáltal:

- **kisegíthet minket rutin és/vagy fárasztó és/vagy veszélyes ... munkák körében,**
- **újszerű szolgáltatásokat valósíthat meg,**
- **megnövelve a „specialisták” körét, emelt szintű szolgáltatásokat nyújthat szélesebb körű (emberi) felhasználói körnek,**
- **önmagában is lehet érdekes, tanulságos,**

Vajon intelligens az „intelligens”?

Google keresés: „intelligent” - kb. 2.350.000.000 találat
(0.32 mp alatt 2023. szeptember eleje)

Intelligens: ház, otthon, robot, repülőgép, gépkocsi, hajó, porszívó, telefon, mosógép, város, érzékelő, anyag, címke, közlekedési lámpa, hírkereső, vállalat, motor, szemüveg, kéziszerszám, akkumulátor, sebészeti szike, kaputelefon, kamera, kávégép, villanyóra, kipörgésgátló, villanyoszlop, tableta, gyógyszer, toll, papír, térkép, babakocsi, izzó, konnektor, védőkabát, gyurma, krémes, borosüveg, szökőkút, fogkefe, mosópor, ...

Nyilvánvaló, hogy **ennek nagyon nagy része csak reklámfogás.**



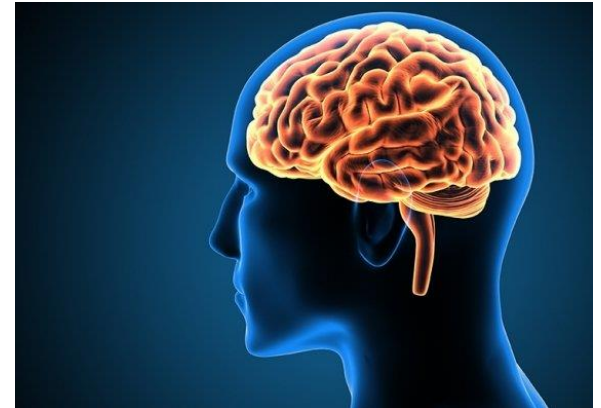
Intelligens krémes - okosan rétegeződő süti (kolléganőnk megcsinálta: *finom!*)

<http://www.mindmegette.hu/intelligens-kremes-az-okos-suti-47942/>

Egy messzi lehetőség? Egy közeli siker?



~ ?

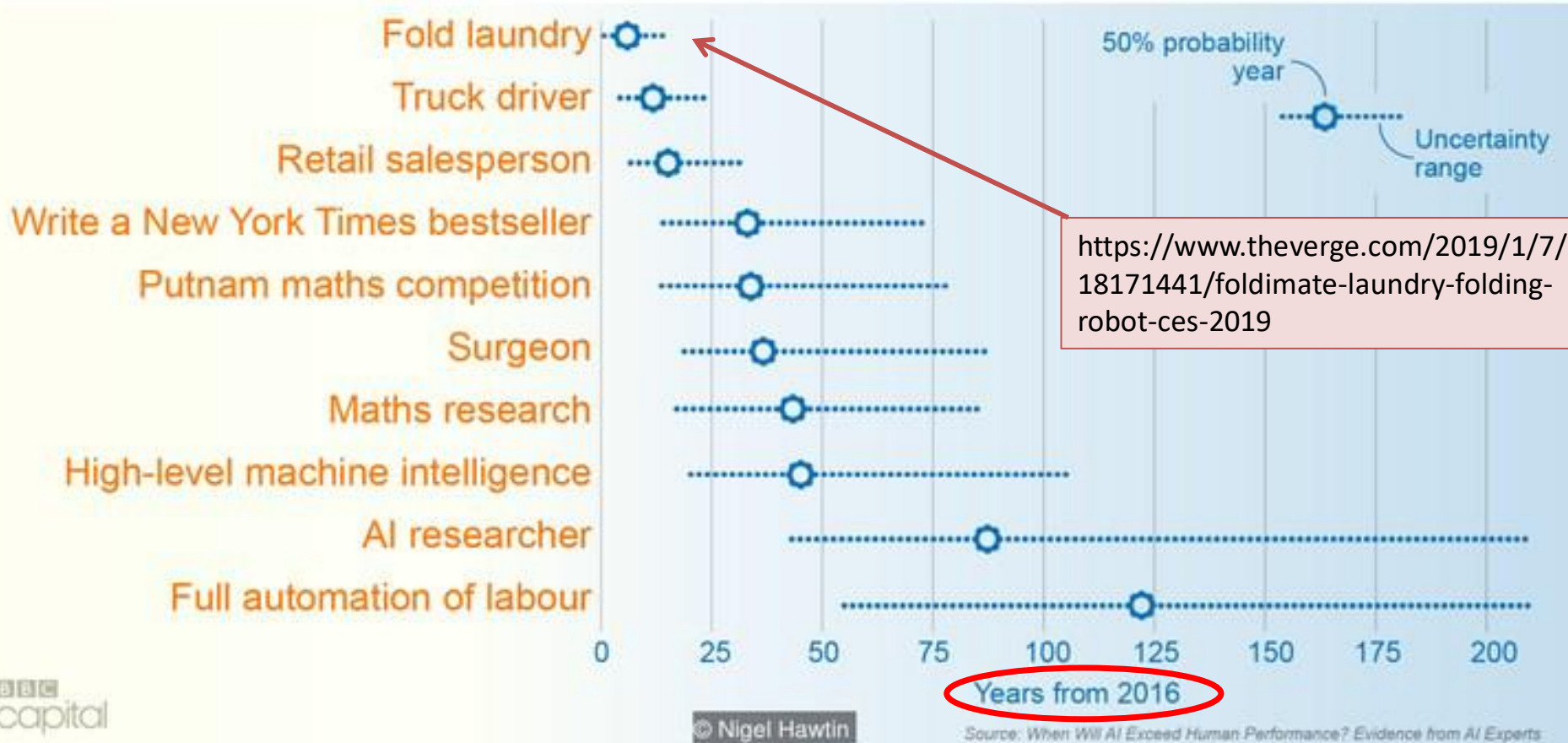


	Szuperszg	Személyi szg	Emberi agy
Számítási egységek	10^6 GPUs + CPUs	8 CPU mag	10^6 oszlop
	10^{15} tranzisztor	10^{10} tranzisztor	10^{11} neuron
Tárolási egységek	10^{16} bájt RAM	10^{10} bájt RAM	10^{11} neuron
	10^{17} bájt diszk	10^{12} bájt diszk	10^{14} szinapszis
Ciklusidő	10^{-9} mp	10^{-9} mp	10^{-3} mp
Műveletek/mp	10^{18}	10^{10}	10^{17}

Egy messzi lehetőség? Egy közeli siker?

When will AI outperform humans at work?

How many years until a machine can do our jobs better than us?



Megközelítés

Mesterséges intelligenciával rendelkező rendszerek

Emberi módon gondolkodó	Racionálisan gondolkodó
Emberi módon cselekvő	Racionálisan cselekvő

- 1. Emberi \neq racionális (Döntéseink nagyrészen érzelemvezéreltek! és ez nem feltétlenül negatív megjegyzés)*
- 2. Nem feltétlenül az emberi vagy a természetben elterjedt mód a legjobb, ha valamilyen célt akarunk elérni - csapkodó szárnyú repülő, futó lábak helyett kerék + út stb.*
- 3. A mi megközelítésünk: racionálisan cselekvő rendszer = intelligens ágens*

Hogyan jellemezzük, hogy mi a racionális?

Mi a racionális? Ami jó teljesítményt nyújt.

Teljesítménymérték: legtöbbször nem azon célszerű mérni a teljesítményt, hogy az ágens hogyan viselkedjen, hanem azon, hogy mit akarunk elérni, és ehhez közelebb kerülünk-e!

Az intelligens rendszertől elvártuk, hogy legyen:
ügyes, többet tud, tanulékony, adaptív, robusztus, autonóm,
tud beszélni, problémákat old meg, kooperál, el tudja magyarázni,
hogy mit miért tesz, kisebb energiát fogyaszt...

A teljesítménymérték formalizálja (matek), hogy mit hogyan és milyen súllyal vesszünk figyelembe.

Ha meg akarjuk mondani, hogy mi a jó teljesítmény, akkor egy **skalár mérőszám**ra van szükségünk. (lásd pl. öttusa)

Mi a racionális? Ami jó teljesítményt nyújt.

A teljesítménymérték legyen külső, különben az ágens könnyen manipulálhatja a teljesítménymértéket, ahelyett, hogy a kívánt állapotot, célt igyekezne elérni!

Tulajdonképpen nem ér sokat az az 5 kredit, nem érdemes e miatt felkelni hajnali 9 órakor, vagy otthagyni a munkát.... Különben is lecsökkenthetjük 1 egységre az Alk. MI tárgy számunkra mutatott értékét, az esti sörözés pedig érjen nekünk 10 egységet!

(Még az embernél is alapvető a külső meghatározottság, lásd pl. David Riesman: A magányos tömeg, Közgazdasági És Jogi Kiadó, 1983

- tradíciók által irányított ember - hagyományok előírta viselkedés
- belülről irányított ember - a gyerekkorban beégetett értékek, célok
- kívülről irányított ember - a többiek elvárásai, pl. divat)

Mesterséges intelligencia(ák?)

HCI Human Computer Interfaces

AmI Ambient Intelligence ambiens intelligencia

Ember érzékelése és kiszolgálása, (természetes) emberi lét egy láthatatlan számítógépes környezetben.

AAL Ambient Assisted Living

Gyenge MI (Narrow AI, Weak AI)

Gépek képesek intelligensen cselekedni (jobban mondva intelligenseknek látszanak a cselekvései alapján, de nem gondolkodnak)

Intelligenciát mutatnak, de csak egy speciális területen (sakkozik, útvonalat tervez, előre jelez, autót vezet, arcot felismer stb.).

Jófajta, régivágású MI (Good Old Fashioned AI - GOF AI)

Minden intelligens viselkedés megragadható egy olyan rendszerrel, amely a tárgyterületet leíró tények és szabályok halmazából kiindulva logikai következtetést végez.

Mesterséges intelligencia(ák?)

Erős MI (Strong Artificial Intelligence)

A gépek valóban intelligensen cselekszenek.

Valóban gondolkodnak-e a gépek?

Dijkstra (1984) "Ez a kérdés kb. annyira releváns, mint annak az eldöntése, hogy vajon tud-e úszni egy tengeralattjáró."

Turing (1950) "Ahelyett, hogy folytan erről vitatkoznánk, általában egyezményesen elfogadjuk, hogy mindenki gondolkodik. Ezt a megközelítést a gépekre is kiterjeszthetnénk, hacsak azt tapasztaljuk, hogy intelligensen cselekszenek. → Turing-teszt: nem látom és szavai/cselekedetei alapján nem tudom eldönteni, hogy ember-e vagy gép

Általános MI (Artificial General Intelligence - AGI)

Az élet számos (összes?) területén intelligensen viselkedik.

Emberszabású MI (Human-Level AI - HLAI)

Absztrakt gondolkodásra képes, következtet, összetett koncepciókat megért, tanul, általánosít, jelleme van, érzelmei vannak stb.

Példa

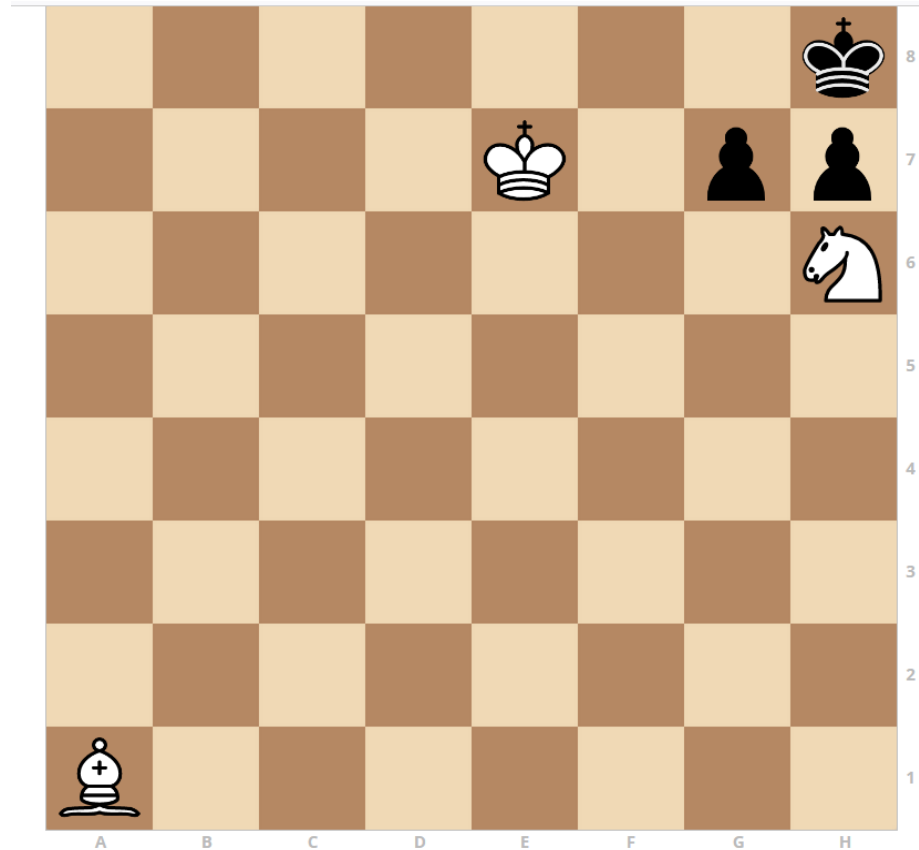
Világos indul és 3 lépésben mattot ad.

Nagyon sok - sakkozni közepesen vagy elég jól tudó - embernek is hosszabb fejtörést okoz ez a feladvány.

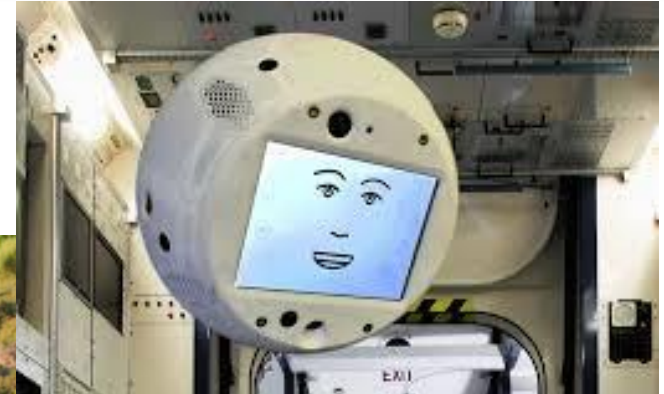
Egy intelligens sakkozó programnak ezredmásodperc se kell a megoldáshoz. **Intelligens?**

De ez a program nem tud főzni, filozofálni, sörözni, nem tud Alk. MI zh-t írni, udvarolni, tüntetni...

Gyenge MI (Narrow AI, Weak AI)



Egy intelligens ágens



Egy intelligens ágens belsejében

<http://www.care-o-bot.de/en/care-o-bot-3.html>



Fraunhofer Institute
for Manufacturing
Engineering and
Automation IPA

Egy intelligens ágens belsejében

információ



... információ-
begyűjtés
eszközei
szenzorok

fizikai hatás



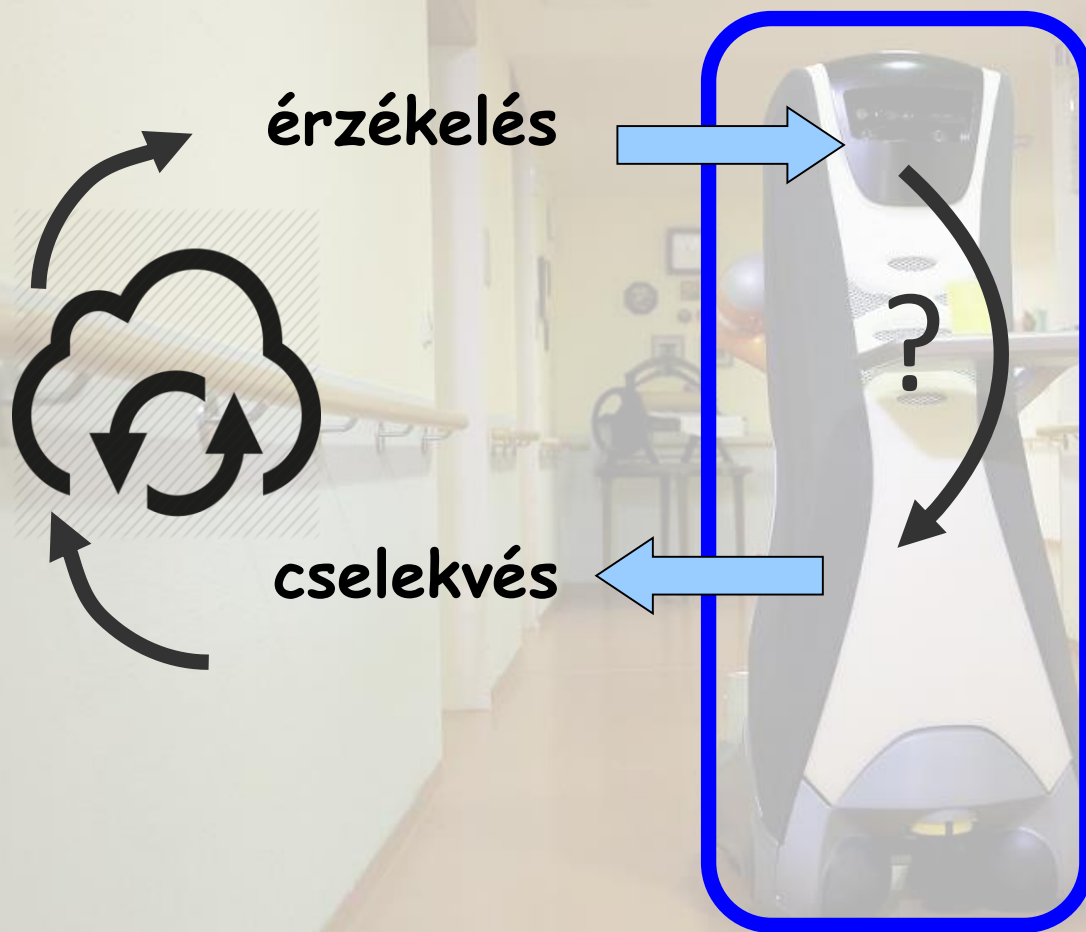
... hatáskifejtés
eszközei
beavatkozók ...

Ami kívül: az a **környezet**

Ami belül: az a **rendszer**



Egy intelligens ágens belsejében



ágens

a környezetébe
(fizikailag) beágyazott,
vele folyamatos
köölcsönhatásban
lévő, ...
érezkelőivel érzékeli,
beavatkozóival
(intelligensen)
megváltoztatja ...

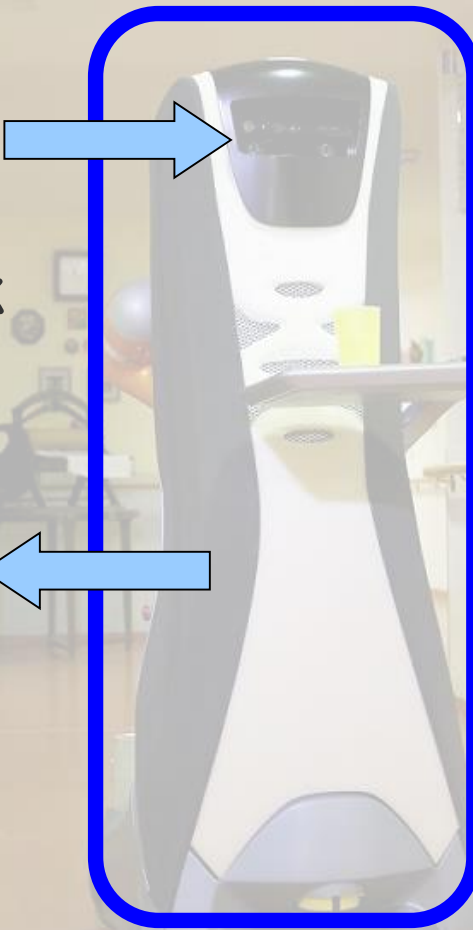
Egy intelligens ágens belsejében

Ágens környezetének
 $s_K(t)$ állapotai vannak

$$s_K(t) \in S_K$$

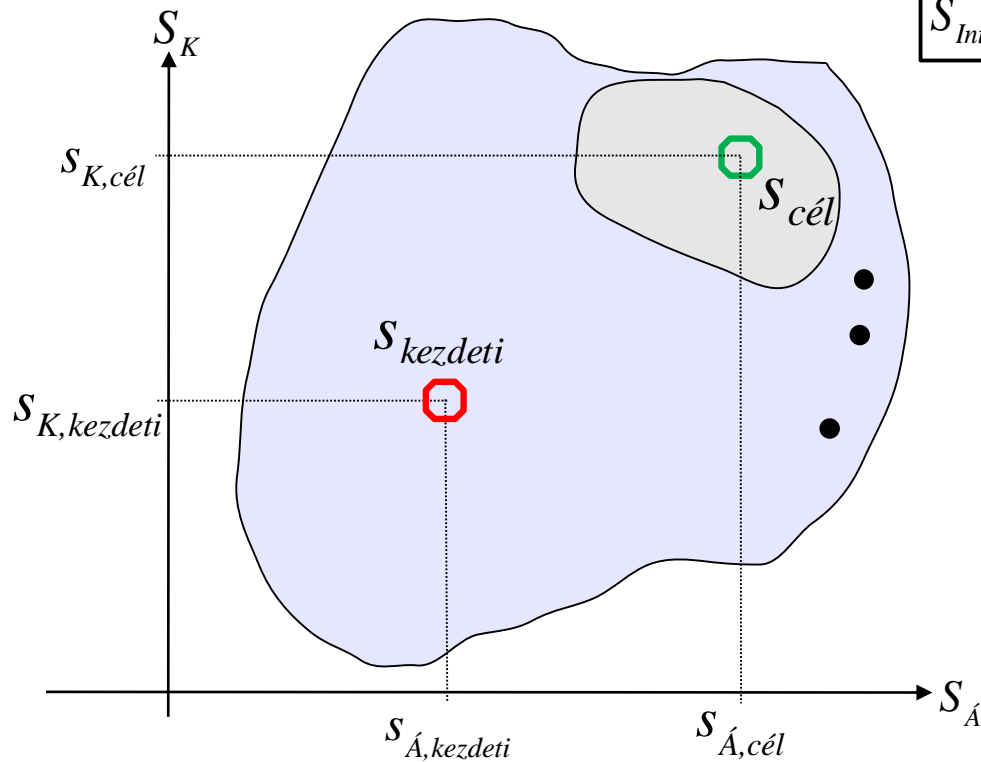
Ágensnek magának is
 $s_A(t)$ állapotai vannak

$$s_A(t) \in S_A$$



$$S_{\text{Intelligens Rendszer}} \subseteq S_K \times S_A$$

Egy intelligens ágens belsejében

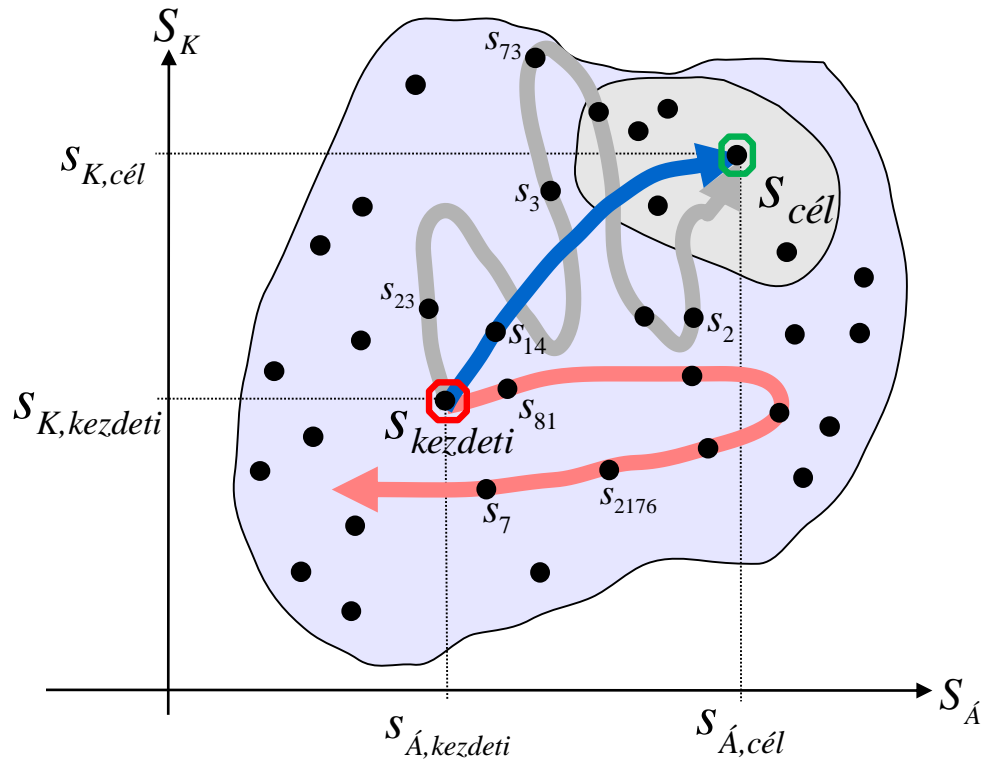


$$S_{Intelligens\ Rendszer} \subseteq S_K \times S_A$$

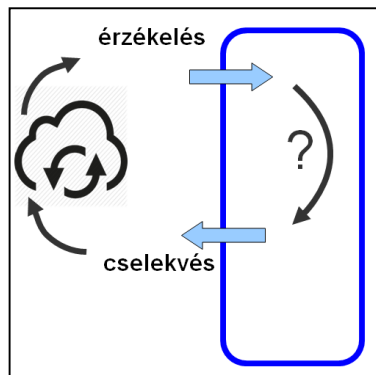
S_K	S_A
Eső, nedves út	Bécs, tele tank
Napos, jeges út	Bécs, üres tank
Köd, nedves út	Győr, tele tank
Napos, száraz út	Bicske, üres tank
Eső, vízes út	Szeged, fél tank
...	Szeged, tele tank
	...

Egy intelligens ágens belsejében

$$S_{\text{Intelligens Rendszer}} \subseteq S_K \times S_A$$



S_K	S_A
Eső, nedves út	Bécs, tele tank
Napos, jeges út	Bécs, üres tank
Köd, nedves út	Győr, tele tank
Napos, száraz út	Bicske, üres tank
Eső, vízes út	Szeged, fél tank
...	Szeged, tele tank
	...



Egy intelligens ágens belsejében

Racionális cselekvés = cél felé irányuló cselekvés \neq siker
(lehet, hogy valójában nem visz majd a cél fele, bizonytalanság)

Intelligens ágens (ahogyan mi foglalkozunk vele) - racionális módon választja meg a cselekvéseit, és a célállapotát sikeresen éri el a környezeti változások, nehézségek stb. ellenére.

A tökéletes racionalitás lehetetlen, a számítási szükségletek túl nagyok, az erőforrások - szenzorok is, beavatkozók is - végesek.

(Pl. nem tudjuk/akarjuk figyelembe venni, hogy lesz-e földrengés a műtét alatt - Magyarországon más lehet a racionalitás, mint Japánban!... Gyakran tökéletességet várunk el az orvostól, tanártól, diáktól, sofőrtől....)

Korlátozott racionalitás - megfelelően cselekedni, miközben az összes számításra nincs elegendő idő!

Egy intelligens ágens belsejében

Az ágens feladata az **előzetes ismereteiből és az érzékelések sorozatából „kiszámítani” a cselekvést, a cselekvés kiszámításának „ügyessége” kapcsolatba hozható a rendszer intelligenciájával.**

DE

- az érzékelések függenek az érzékelőktől,
- a cselekvések függenek a beavatkozóktól,
- a „kiszámítani” függ az ágens felépítésétől, tudásától.

ÍGY

A rendszer racionalitása, intelligenciája (hogyan döntsön) függ attól is, hogy milyen érzékelőkkel és beavatkozókkal (és egyéb erőforrásokkal) rendelkezik!

(Elmaradott országban praktizáló orvos... egy lámpakörte is lehet inkubátor - Magyarországon nem lenne racionális.)

Egy intelligens ágens környezetében

A szükséges intelligenciát befolyásolja a **környezet**

hozzáférhető
determinisztikus
epizódszerű
statikus
diszkrét
egy ágenses
kooperatív

nem hozzáférhető
nem determinisztikus
nem epizódszerű
dinamikus
folytonos
többágenses
versengő

A legnehezebb a nem hozzáférhető, nem epizódszerű, dinamikus, nem determinisztikus, és folytonos, többágenses környezet.

A valós helyzetek legtöbbször olyan bonyolult, hogy gyakorlati okokból nem determinisztikusként kezelendők.

Egy intelligens ágens

Ágens „ellenségei”

(amiktől az intelligenciája korlátos, vagy romlandó)

- (1) a feladat/környezet komplexitása (bonyolultsága)
- (2) véges erőforrásai
(rendelkezésre álló időt is beleértve)
- (3) információhiány érzékeléskor
- (4) a környezet változékonysága

MI haszna és kockázata

- (1) Embereknél intelligencia \approx dominancia, de MI nem egy ember!
- (2) Standard MI modellje: **a (fejlesztő által megadott) cél tökéletesen ismert**, minden erőforrás bevethető a cél megvalósításába, ... Működik, ha a gép nem önálló. **Ha a cél helytelen, leállítjuk, módosítjuk, újraindítjuk, ...**
- (3) Ha egy gépnek lefixált célja van, a kikapcsolásának ellenáll(hat), mert ez (bizonyíthatóan) a céljának elérését meggátolja, viszont más célja nincs is.
- (4) N. Wiener (1960): „Ha a céljainkhoz olyan gépi eszközt használunk, amelynek működését hatékonyan nem látjuk át, legyünk biztosak abban, hogy a gépbe helyezett cél valóban az, amit akarunk.”
- (5) A gépnek célt adó (felhasználó) ember: rövidlátó, irracionális, emocionális, nem képes kiszámítani a preferenciáinak megfelelő optimális teendőt, ..., az emberek közt váltott célok sohasem kizárólagosak, abszolútak, feltétlenül végrehajtandók, ... (Főzzél nekem egy teát!)

MI haszna és kockázata

AI Cited for Unlicensed Practice of Law

California Law Bans Bots From Pretending to Be Human

San Francisco Bans Facial Recognition Technology

Automated to Death



Google Glass Is Banned
On These Premises

stoptheyborgs.org ©©©©

Tesla: Autopilot was on during deadly Mountain View crash

NUMBER OF OPERATING ROOM INJURIES LINKED TO ROBOT SURGERY REPORTEDLY INCREASING

Bizonyíthatóan jótékony MI

MI egyik új megközelítésű célja: nem tiszta intelligencia, hanem az **emberi értékekkel összehangolt** intelligencia.

Új modell: **a gép céltudása részleges!**

Ha egy gép tudja, hogy a célját teljes egészében nem ismeri, mondhatja pl. hogy ez a cselekvés jónak tűnik, feltéve, amit tudok, de beszéljük ezt meg (a főnök emberrel), mert hátha ez a cselekvésem sért valami olyan preferenciát, aminek tudatában nem lehetek.

Így eljutunk olyan gépekhez, amelyek engedélyt fognak kérni, amelyek pl. esetenként hagyják magukat kikapcsolni (mert a gép tudni fogja, hogy a cselekvése esetleg káros lehet). Ha a gépnek egy fix célja lenne, soha nem hagyná magát kikapcsolni, hiszen ez garantálna, hogy a célját nem fogja tudni teljesíteni.

Az iskola (BME is) és az élet...

“Difference between school and life: School teaches you lessons and then gives you a test. Life gives you a test and you learn the lessons”