

Intelligens ágensek – Mesterséges Intelligencia rendszer technikai gyökerei

Dobrowiecki Tadeusz

Mérés és Információs Rendszerek Tanszék

Habilitációs előadás

BME-VIK, október 2013

1. Lehet-e intelligens rendszerekről beszélni egyszerűen, mégsem triviálisan, nem elveszítve az analízis és a szintézis gondolati apparátusát?
2. Láttatni, hogy egy intelligens rendszer valóban egy szolgáltató rendszer és az intelligenciája nem más, mint egy adagolható szolgáltatás, mely tervezhető és tesztelhető.
3. Ez az erősen leszűkített felfogás mérnöki szempontból helytálló, mert:
 - elkerülhetők meddő „tudományos” gumicsontok
 - algoritmusokban, módszertanokban lehet előrehaladni
 - egyre hasznosabb rendszereket lehet készíteni.



Jogos igények **Józan ész**
Rendszerelmélet **Kibernetika**
Szabályozáselmélet ~



Intelligens ágens



Care-O-bot 3

Fraunhofer Institute for Manufacturing Engineering and Automation

Sensor Head

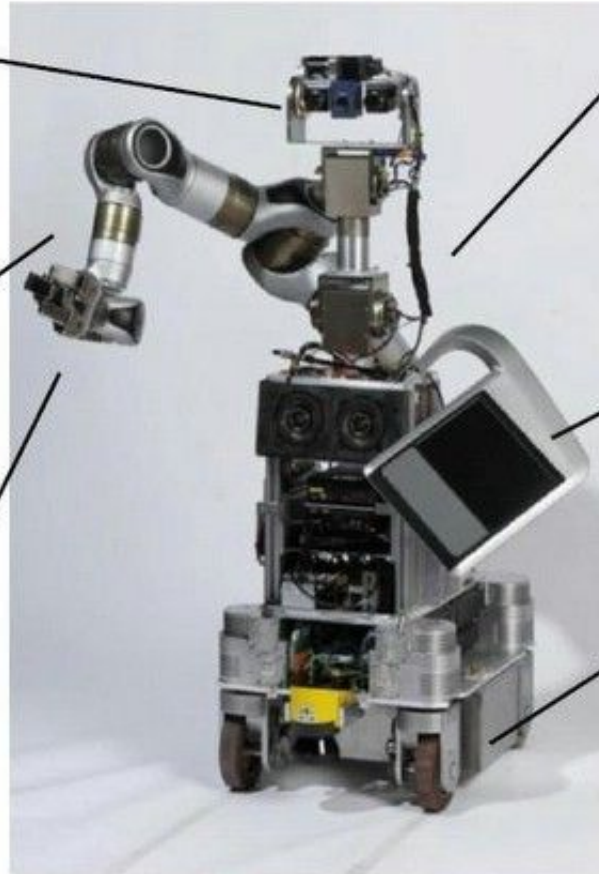
- Stereo camera
- 3D sensor
- 1 DOF to flip sensors back and forth

Manipulator

- 7 DOF, redundant
- TCP and joint control
- Hollow shaft for cables

Gripper

- 7 DOF, 3 fingers
- In-finger tactile sensors



Basis

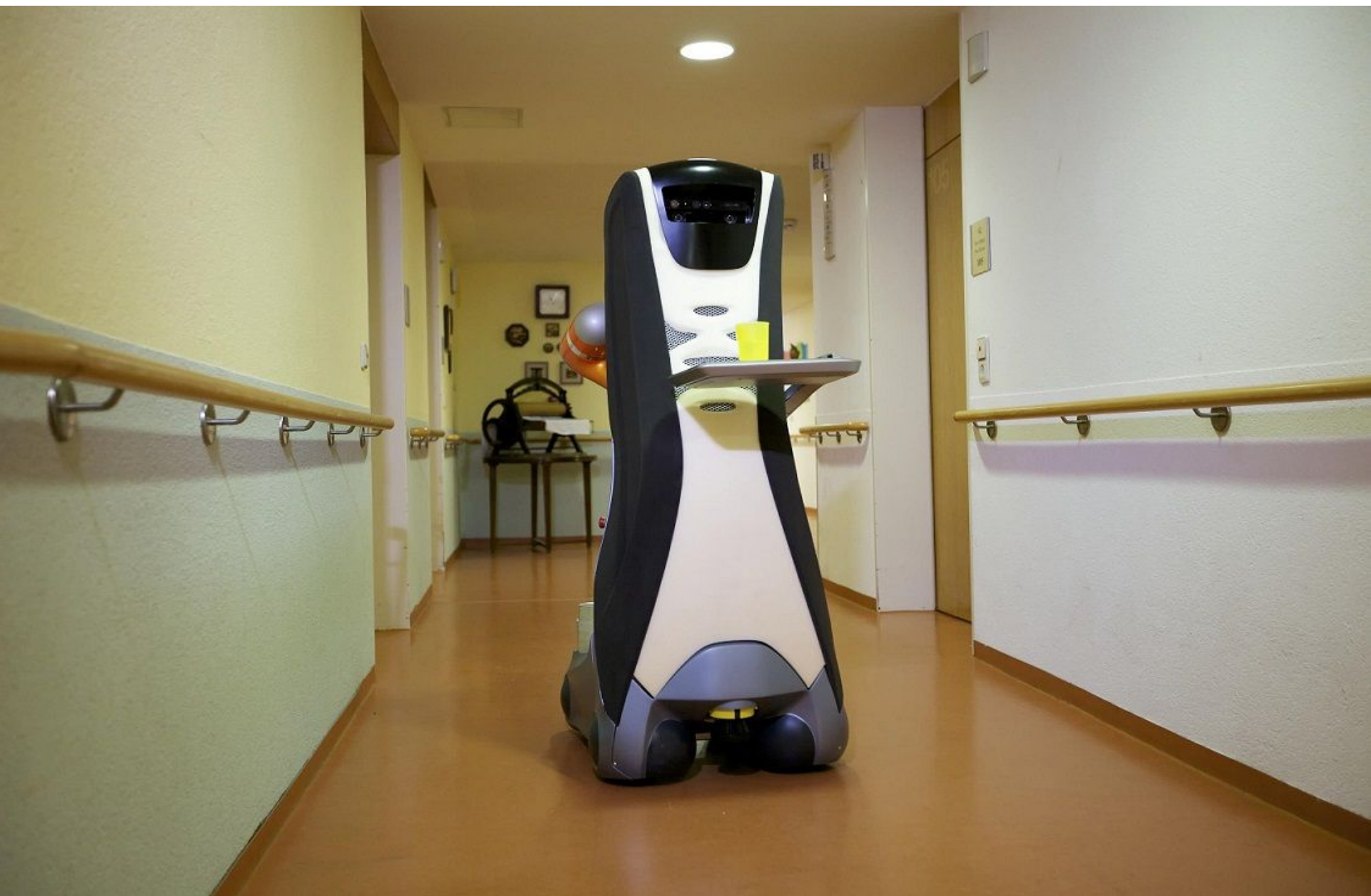
- 4 DOF kinematic for body expressivity
- Stereo speakers
- 3 control PCs

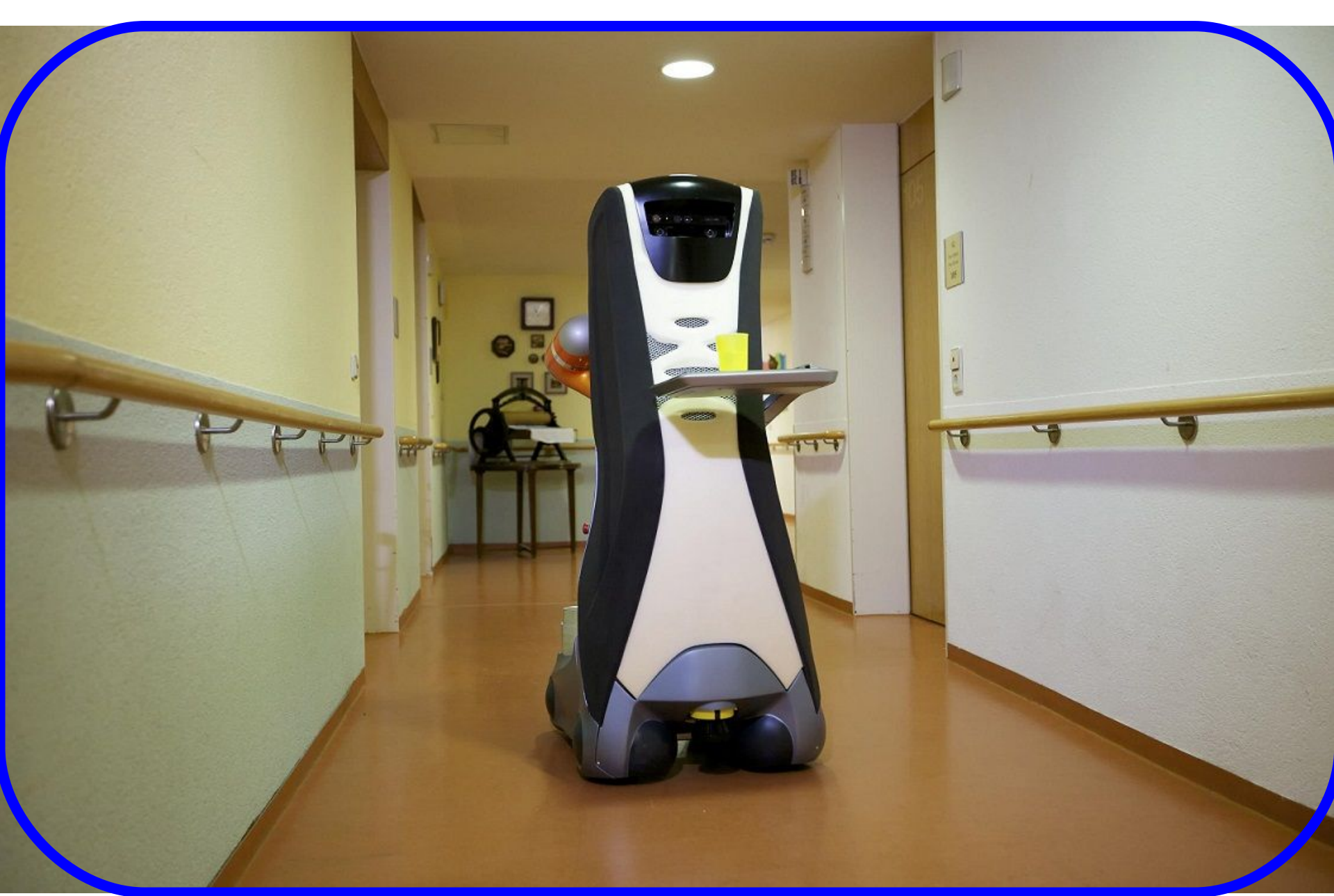
Tray

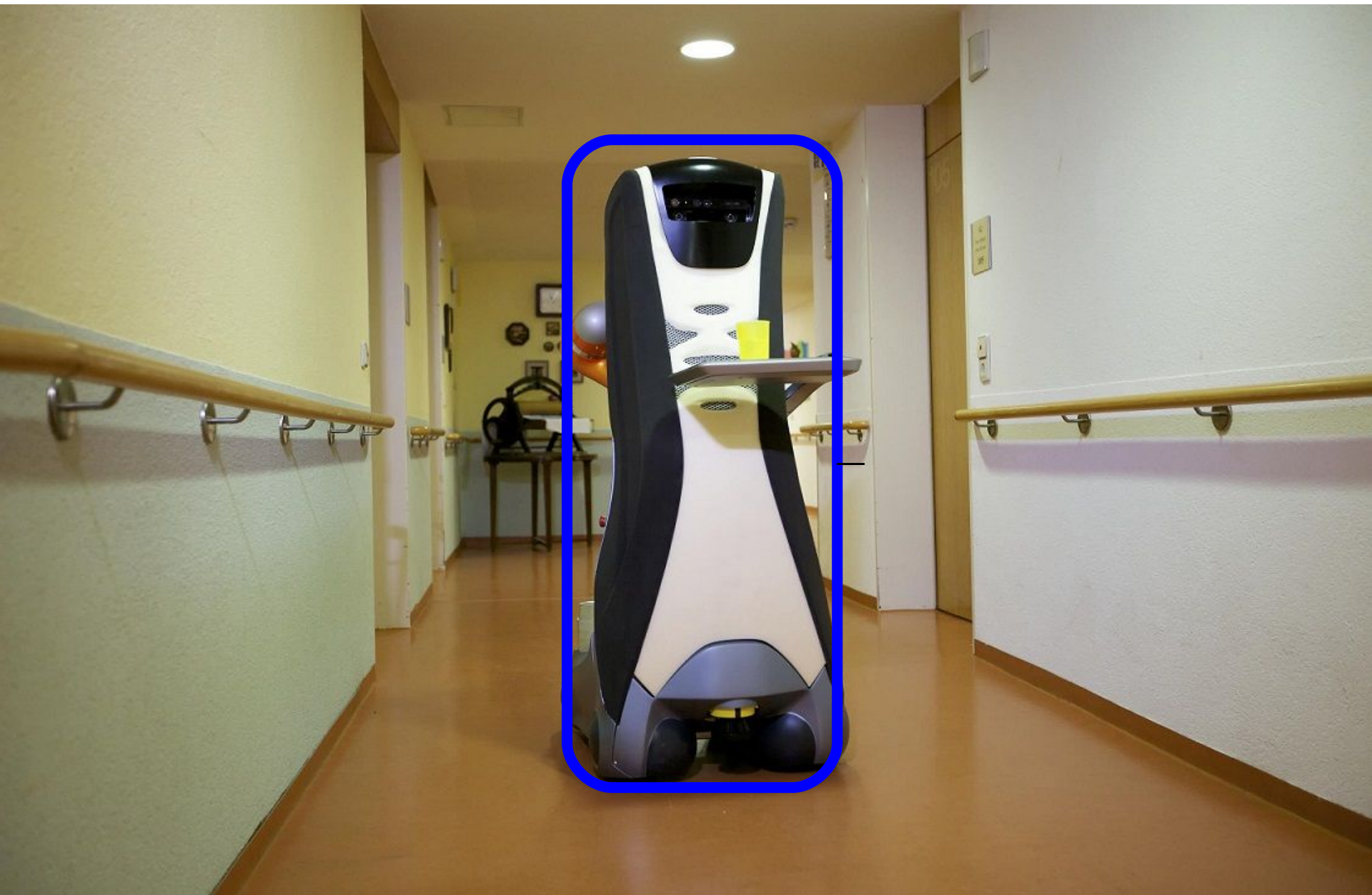
- 1 DOF
- Touch screen

Omnidirectional Platform

- 4 wheeled omni-directional drive system
- 3 laser scanners
- Li-ion or lead-acid battery







(rendszer)
környezet(e)



rendszerhatár



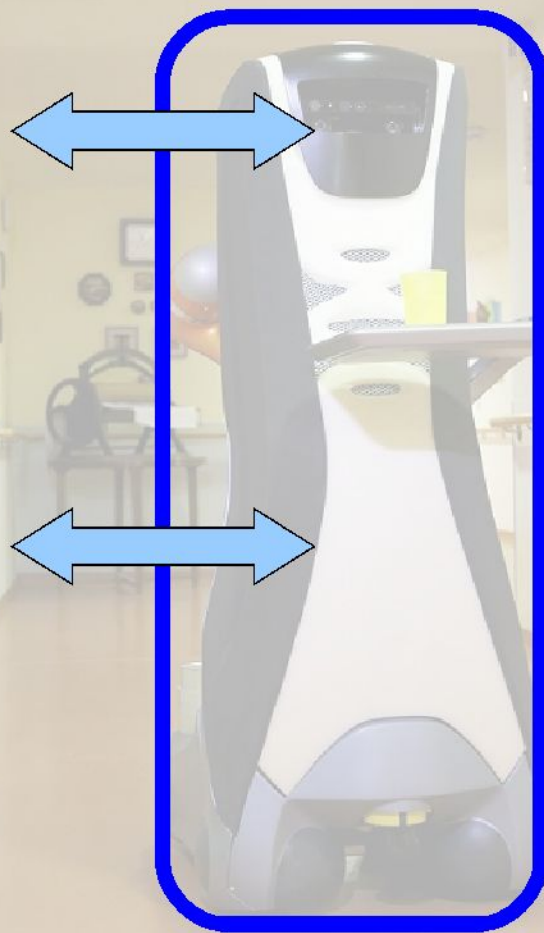
rendszer



rendszer a
rendszerben = struktúra



információ
hatás
anyag



rendszerhatár

= vmi bent
vmi kinnt
és a
kölcshatás



információ



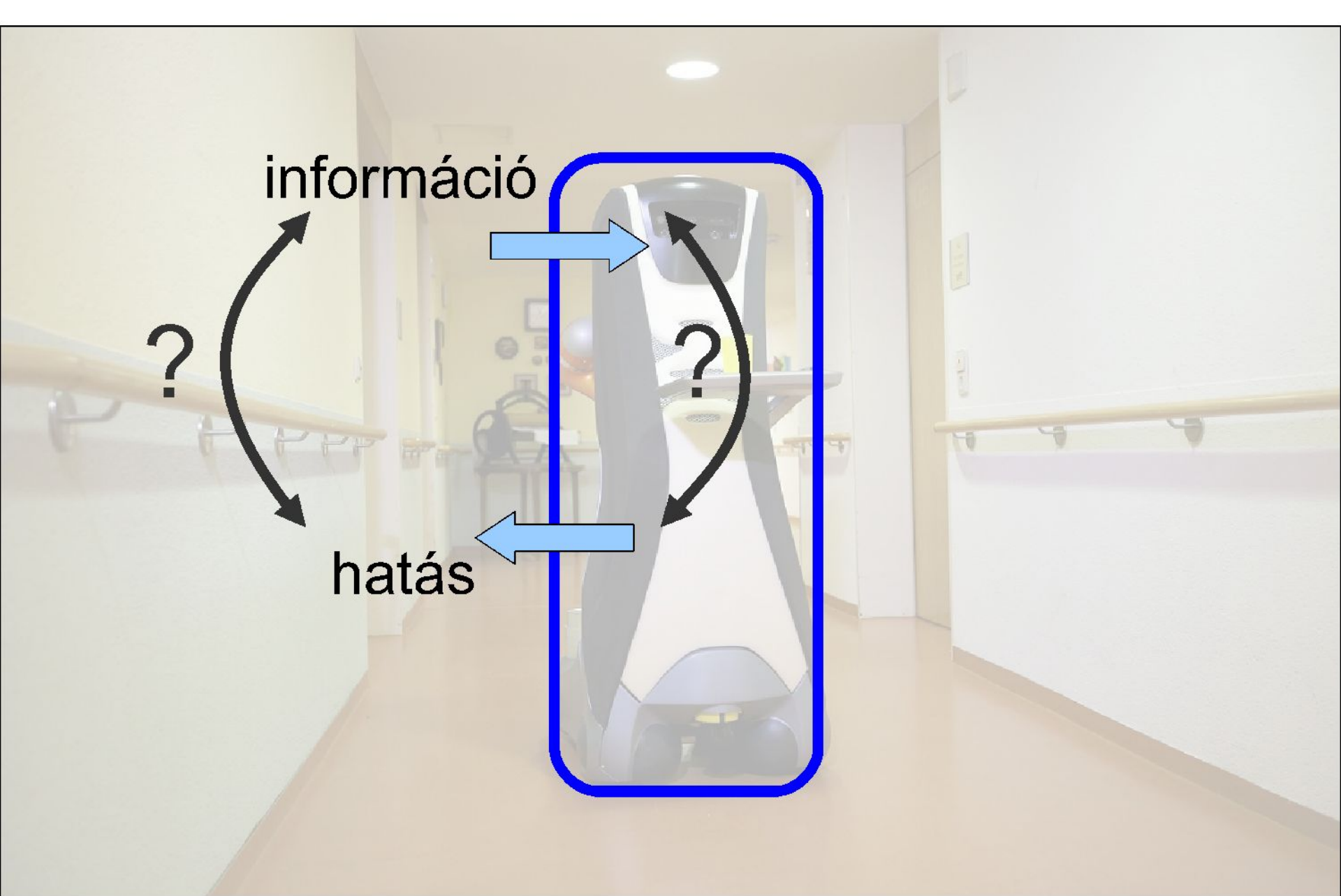
hatás

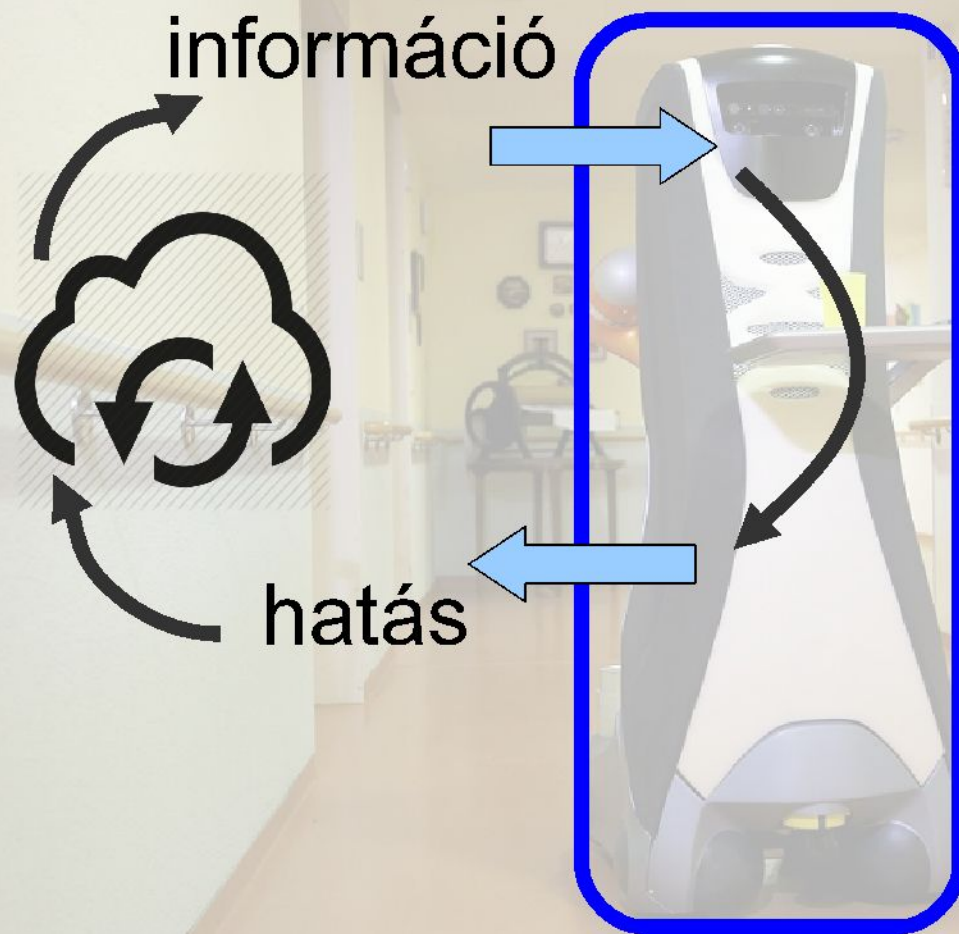


... információ-
begyűjtés
eszközei
- szenzorok

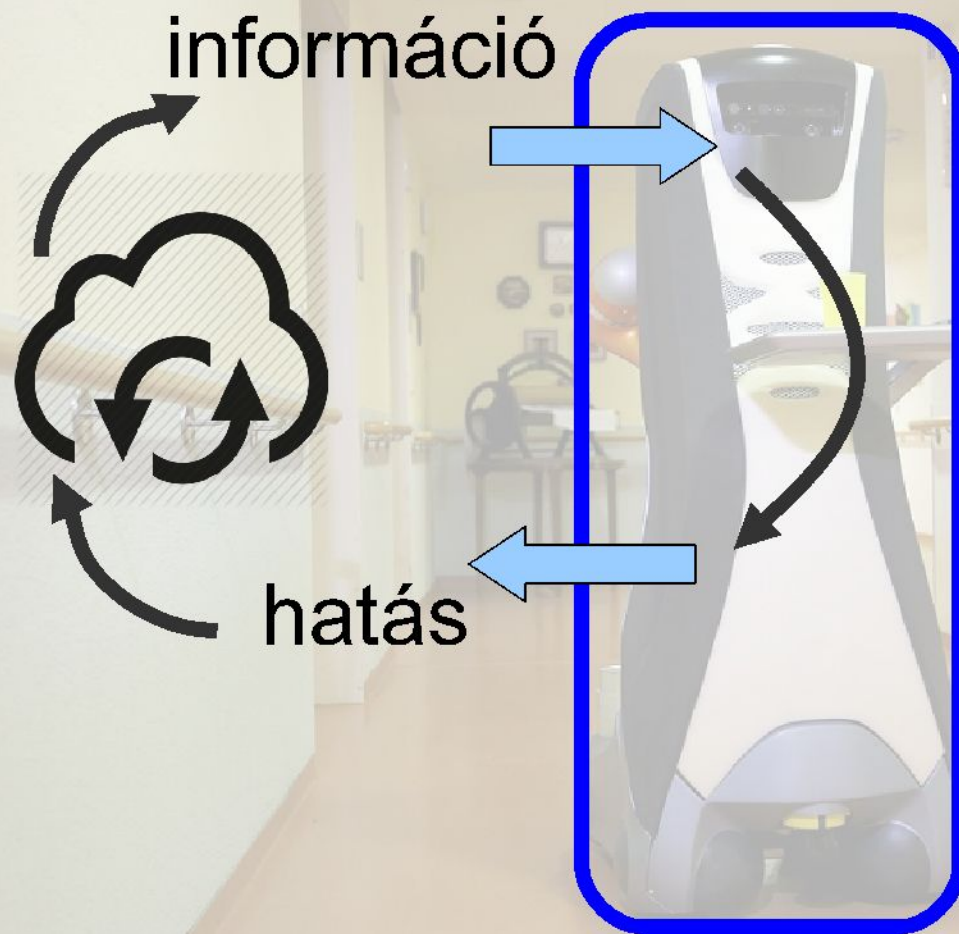
... hatáskifejtés
eszközei
- beavatkozók ...







... körforgás
... (vissza)csatolás?
... szabályozás?



ágens =

a környezetébe
ágyazott, vele
folyamatos
kölsönhatásban
lévő, ...
érzékelőivel
érzékel,
beavatkozóival
megváltoztatja



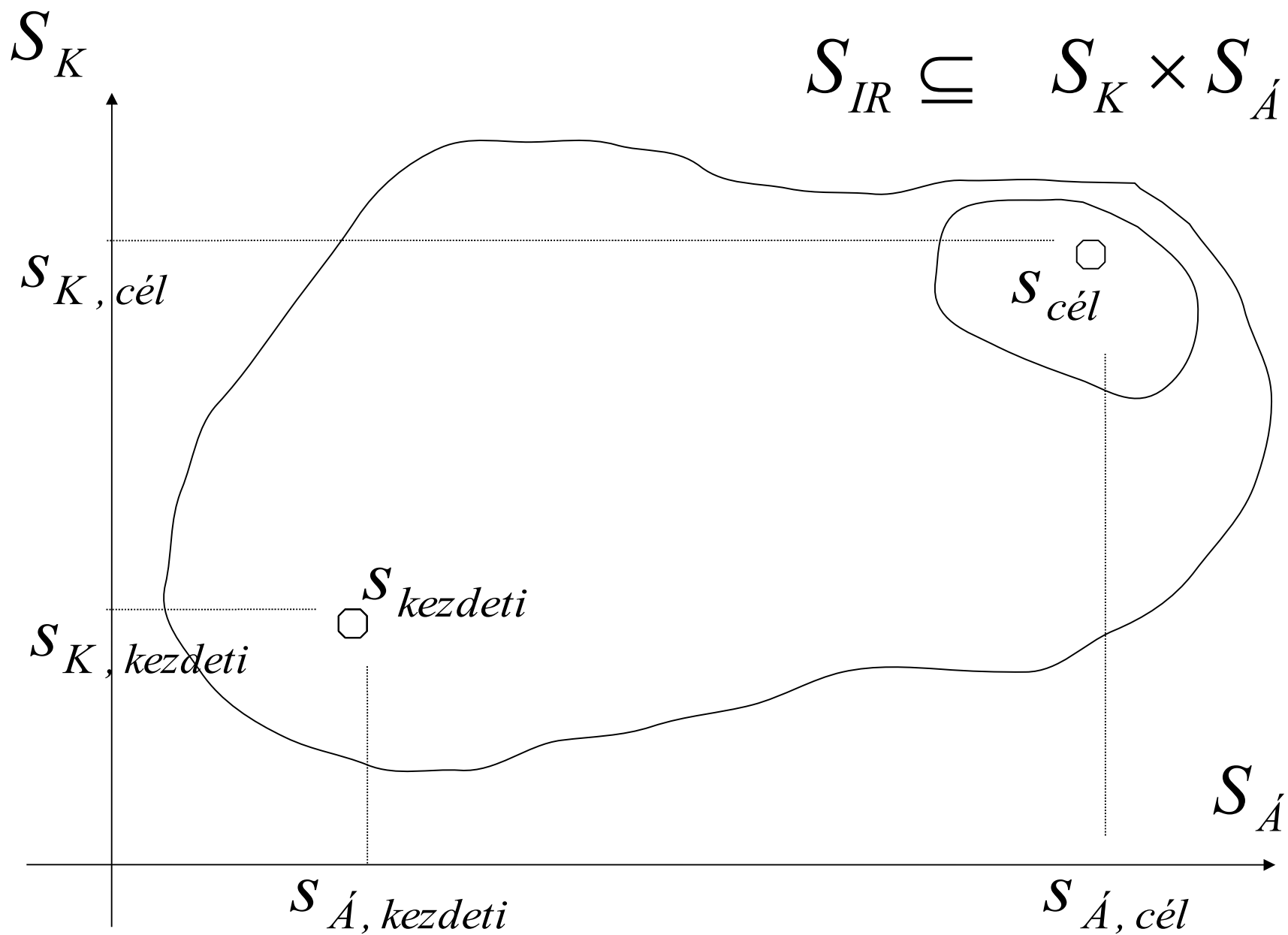
környezetnek
 $s_K(t)$ állapotai
vannak

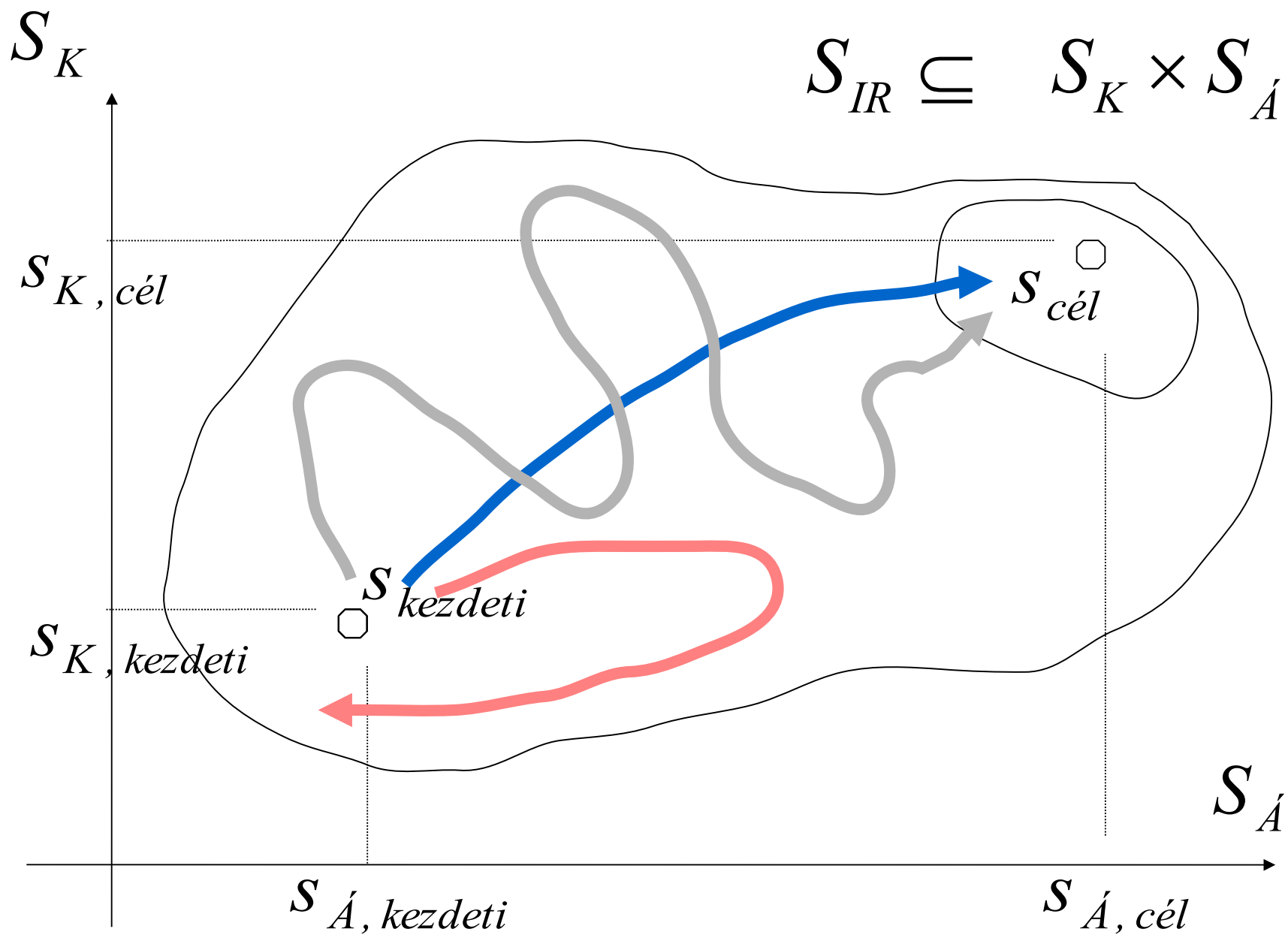
$$s_K(t) \in S_K$$

ágensnek
 $s_A(t)$ állapotai
vannak

$$s_A(t) \in S_A$$

$$S_{IR} \subseteq S_K \times S_A$$





Néhány meglehetősen triviális észrevétel

Ágens beavatkozása a környezetére irányul.

Egyáltalán miért cselekszik? Mert van tennivalója.

A környezet nem olyan, ami neki kellene.

Milyen környezet kellene? $S_{\text{cél}}$

Ágens „célja” a környezetének egy meghatározott állapota (állapothalmaz), ami számára kívánatos.

Érzékelés/ beavatkozás ciklusában ágens az S_{IR} térben egy **trajektória** mentén halad.

Trajektóriák nem egyformák.

Van, ami a „cél” felé visz, flottul. Van, ami épp csak.

Van, ami egyenesen eltereli az ágenst a céljától.

Olyan ágens, ami különböző állapotokból kiindulva a célállapotaihoz mindig a jó (hatékony, rövid, ...) trajektóriát választja, nyilván más ágensnél ügyesebb.

Néhány meglehetősen triviális észrevétel folyománya

Racionális cselekvés = adott célokra irányuló cselekvés

Intelligens ágens = célorientált módon választja a cselekvéseit és a célállapotait sikeresen (hatékonyan) éri el a környezete által mutatott változások, nehézségek, és komplexitások ellenére.

Összetett környezetben a tökéletes racionalitás lehetetlen, mert a számítási szükségletek egyszerűen túl nagyok.

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

Mesterséges Intelligencia művelése „mérnöki” felfogásban korlátosan racionális ágensek R&D-je

nyilván nem az összes szempont, de kellően gazdag, akár emberi, akár gépi viselkedés esetén lehetőséget ad gyakorlatilag fontos kérdések kivizsgálásához.

Pragmatikus szempontból fontos, hogy ilyen intelligens rendszerek

- fokozottan kiségitenek minket rutin, veszélyes ... munkák körében,
- megnövelve a „specialisták” körét, emelt szintű szolgáltatásokat nyújthatnak szélesebb (emberi) felhasználói populációnak,
- újszerű szolgáltatásokat valósíthatnak meg,
- önmagában is lehetnek érdekesek, tanulságosok,

Akkor tervezzünk intelligens ágenst!

Kell egy mechanizmus, ami megfelelő trajektórián tartja a rendszert, amíg különbséget érez a pillanatnyi és a célállapot között.

Rendszerelmélet, szabályozáselmélet:

állapot-, megfigyelési egyenlet, egyensúlyi állapot, attraktor, stabil, vezérelhető, megfigyelhető, lineáris, dinamika, nemlineáris, előrecsatolás, visszacsatolás, szabályozó, ...

Kibernetika:

sokaság elve, jó szabályozó elve, ...

Józan ész(?):

legyen jó és egyszerű és olcsó és ...

Előzetes konklúzió:

intelligens ágens valamiféle „ügyes szabályozó”
(nagyon bonyolult matematikájú feladatban)

De vannak problémák ...

De vannak problémák ...

állapotegyenlet

megfigyelési egyenlet

egyensúlyi állapot, attraktor, stabil

vezérelhetőség

megfigyelhetőség

lineáris

előrecsatolás

visszacsatolás

szabályozó tervezése

sokaság elve

jó szabályozó elve

dinamika

...



állapotegyenlet
megfigyelési egyenlet



$$? = f(?)$$

Min múlik mi adott pillanatban az intelligencia?
a siker mérőszáma
az eddigi megfigyelések
amit az ágens a környezetéről tud
mennyire cselekvőképes

Ágensfüggvény - az ágens „diff”-egyenlete

$$cs_t = f(m_t, m_{múlt}, cs_{múlt}, TB_t(S_k, S_{cél}, d(...), ...))$$
$$X(t+1) = f(X(t), ...)$$

$$cs_t = f(m_t)$$

$$cs_t = f(cs_{t-1})$$

$$cs_t = f(TB_t), \dots$$

több intelligencia, több önállóság,
több furmányososság, több ...

egyensúlyi állapot,
attraktor, stabil állapot



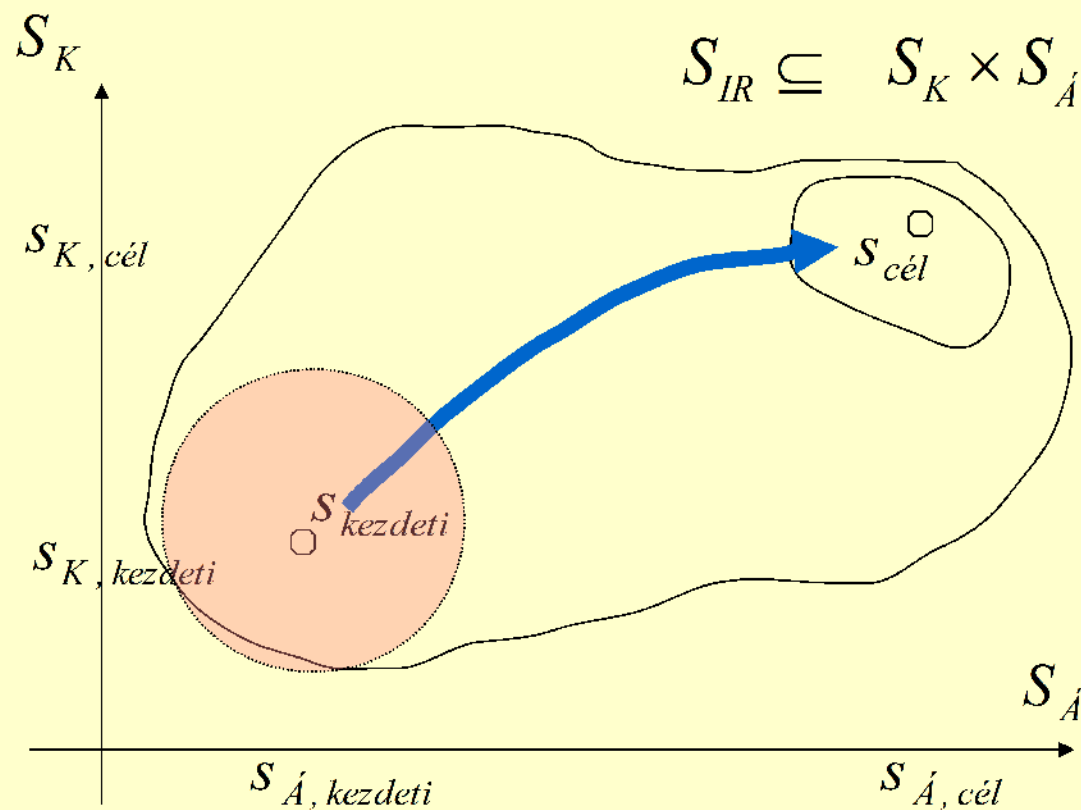
3-féle cselekvés és folyamánya:

- kifelé a fizikai környezetre ható
jó trajektória = jó **feladatvégzés**
- befelé saját kognitív struktúrákra ható
jó trajektória = sikeres **adaptáció, tanulás**
- kifelé más ágensekre ható (kommunikáció)
jó trajektória = sikeres protokolláris
együttműködés, szervezkedés

vezérelhetőség

X

egyedi cselekvés hatósugara **lokális**
cél távolsága **globális**, l. később



megfigyelhetőség



- (1) Szenzorok? - környezet részben,
v. egyáltalán **nem hozzáférhető**
- (2) Megfigyelés =
evidencia +
affordancia (felmerülő lehetőségek)
- (3) A megfigyelés **befolyásolja a megfigyelőt**:
 - (a) érzékelt evidencia megváltoztatja
a TB állapotát → ágensfüggvény
 - (b) ágensek közötti kommunikáció =
tényszerű tudás csere +
mentális állapotátmenetek
(ACL szabványok)

lineáris

X

Szuperpozíció? Arányosság?



előreacsatolás visszacacsatolás



Előreacsatolás:

potenciálisan hibamentes
a baj megtörténte előtt
hosszú távon divergálhat
tanulás más hibájából
ZH-re előre tanuló hallgató

Visszacacsatolás:

nem lehet hibátlan
a bajnak meg kell előbb történnie
hosszú távon konvergál
tanulás saját hibájából
ZH-t felkészülés nélkül megíró hallgató
sorsa



Ágens feladata **érzékelésből ügyesen „kiszámítani” a cselekvést**

- érzékelések függenek az érzékelőktől.
- cselekvések függenek a beavatkozótól.
- „kiszámítani” függ az ágens felépítésétől.

Ágensfüggvény

$$cs_t = f(m_t, \dots, Tb_t)$$

argumentumok matematikája

$f(\cdot)$ algoritmizálhatósága

komplexitási kérdések (NP leselekedik)

egyetemes **keresési** algoritmus

agensfüggvény kiszámítása

tervezői asztalon

agens által, működése közben

agens által, működésre előkészítve

→ **architektúrák**

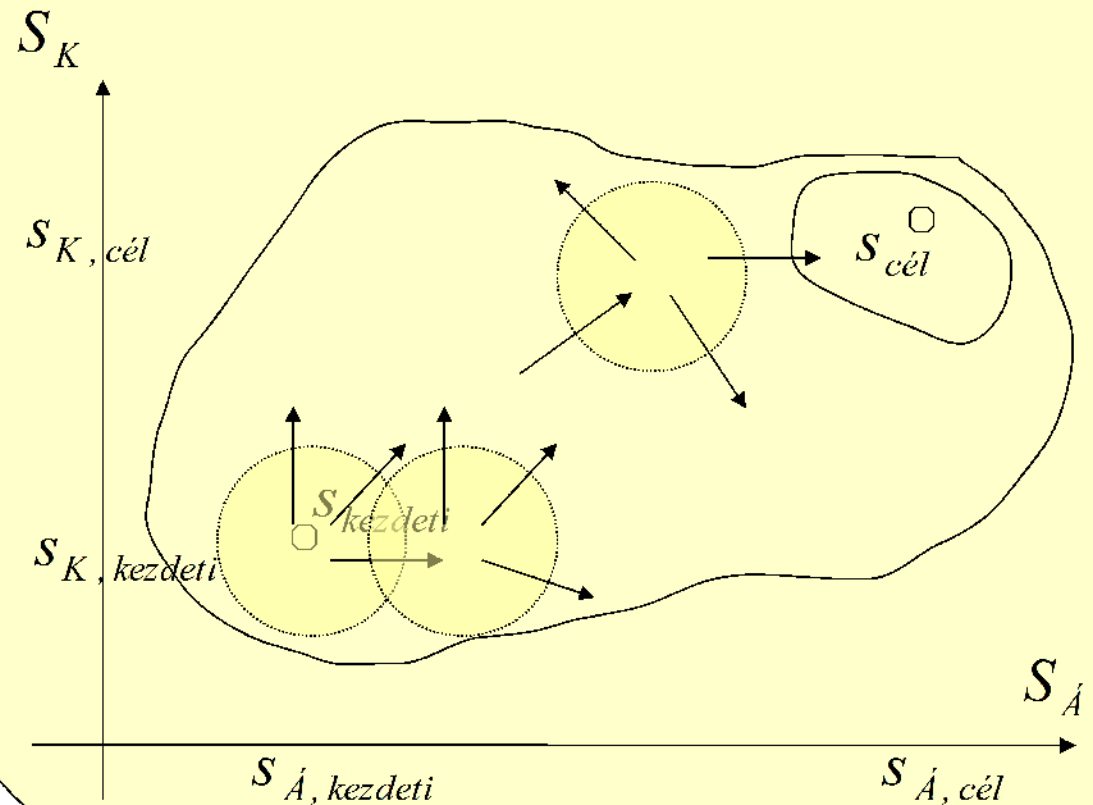
szabályozó tervezése



Egyetemes **keresési** algoritmus

mi lenne, ha ... lépünk,

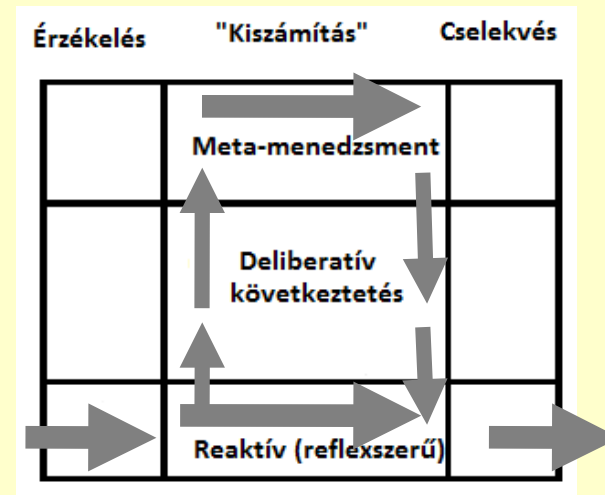
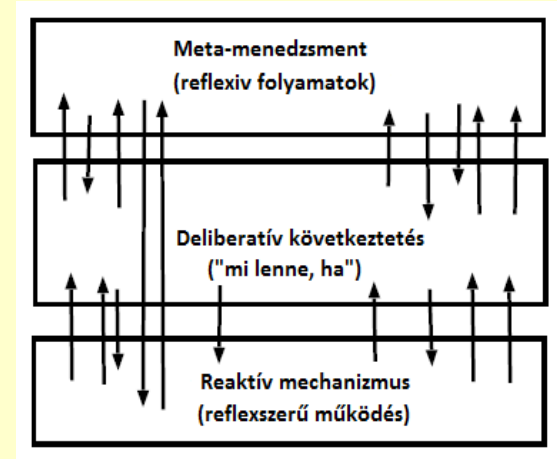
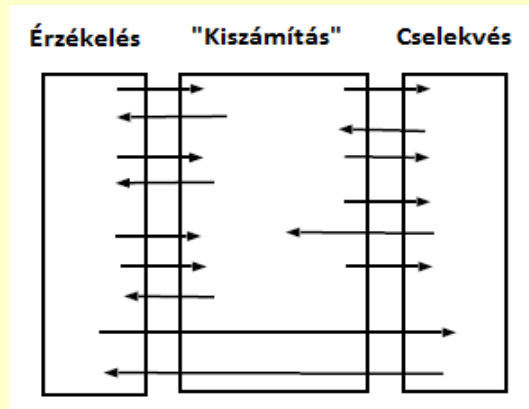
és akkor mi lenne, ha tovább ... lépünk
és akkor ...



szabályozó tervezése



Ágens architektúrák



sokaság elve



Kibernetika

Sokaság elve

„minél több vezérlő cselekvéssel rendelkezik egy rendszer, annál több zavaró tényezőt tud kompenzálni” (Ashby, 1956)

Jó szabályozó elve

"egy jó szabályozó szükségszerűen jó modellje a szabályozott rendszernek" (Conant, 1970)

Triviálisan igaz, mert:

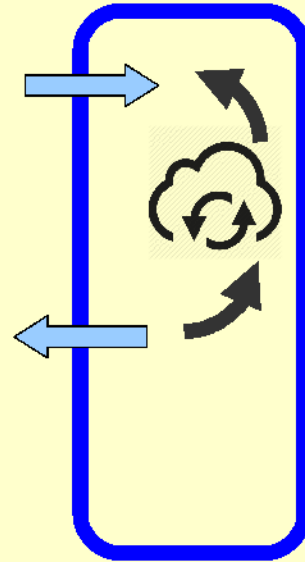
TB elve

„egy ágens tudásbázisa legyen kellően gazdag, és ezzel a gazdagsággal fedje le a feladatainak modelljét (minden lényeges tudást)”
(Dobrowiecki, 2013) :-)

dinamika



$d/dt(S_K)$



$d/dt(S_A)$

Ki a gyorsabb?

A környezet? Ágens adaptációja, tanulása soha nem lesz eredményes.

Az ágens? Sikeres lesz változó környezet ellenére.

Egyformák? Igazi probléma több ágenses környezetben. Kölcsönös tanulásuk nem fog konvergálni!

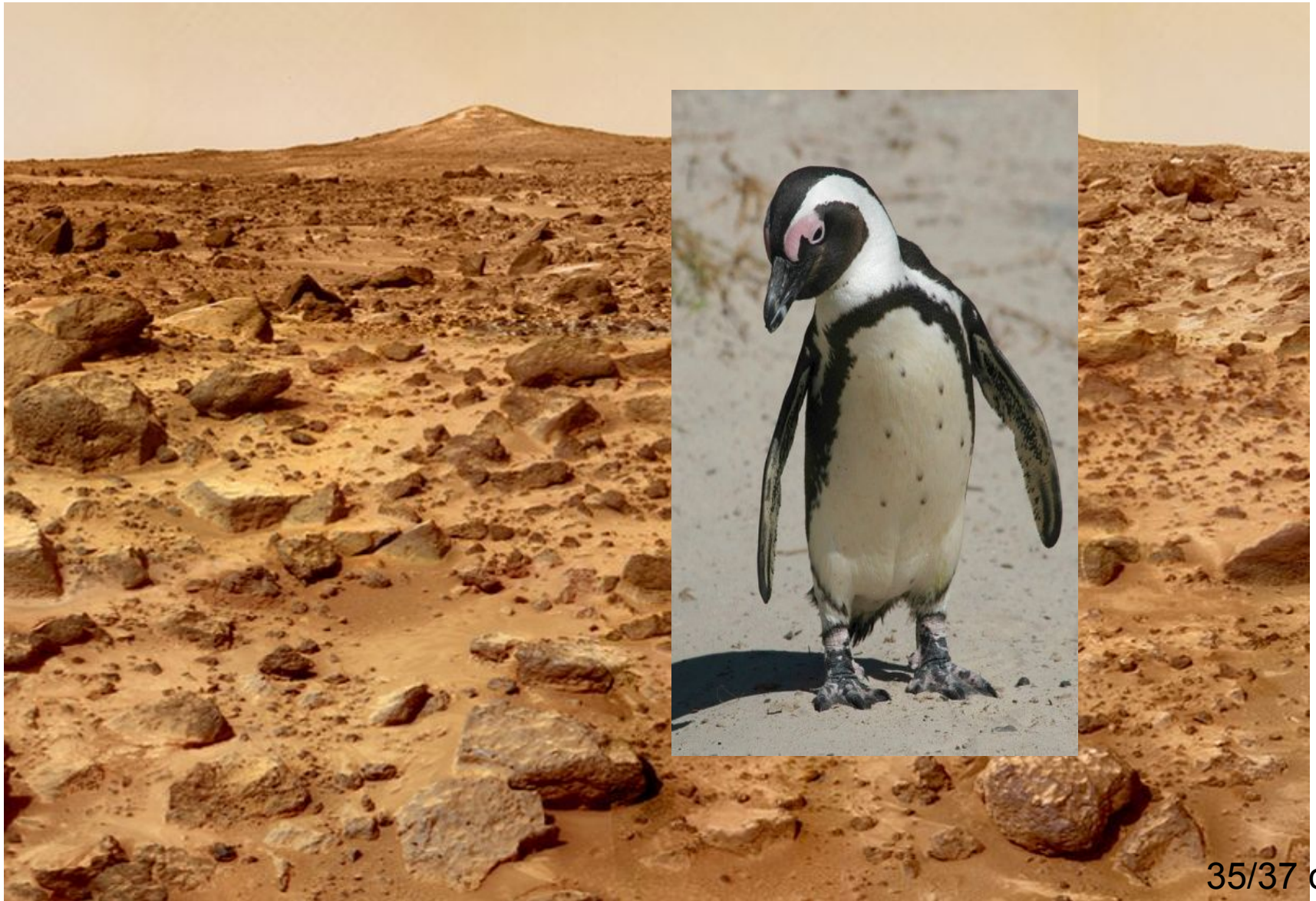
Eddig szó sem esett a számítógépről!

Az igazság az, hogy a komplex feladatok megoldásához tartozó „ágensfüggvények” kiszámítása és azáltal a klasszikus értelemben vett nem vezérelhető és nem megfigyelhető feladatok megoldása olyan algoritmusok megvalósítását igényli, amelyeket ma kizárólag programozottan implementált környezetben tudunk csak elképzelni.

Szóval intelligens rendszer manapság mégis egy számítógépes rendszer, bár nem szükségképpen, valóban opportunista módon.

És a végére, az eddigiekből adódóan, két egyszerű elv.

Pingvin-elv – egy rendszer csak a natív környezetében lehet igazán intelligens



Ferrari-elv –

egy rendszer igazán intelligens,
ha a feladat szükségleteinek
megfelel. Annál kevesebb,
de annál több intelligencia is
– handicap.



Köszönöm szépen a figyelmet!