

**Források:** Stanford University – Artificial Intelligence course: [www.ai-class.com](http://www.ai-class.com)

Alison Cawsey: Mesterséges Intelligencia, Panem könyvkiadó 2002, ISBN 9635452853

Stuart Russel és Peter Norvig: Mesterséges Intelligencia, Panem könyvkiadó 2005, ISBN 9635454112

**Az előadás 1-20 slidejait a források alapján szerkesztette: Dr. Esztergár-Kiss Domokos**

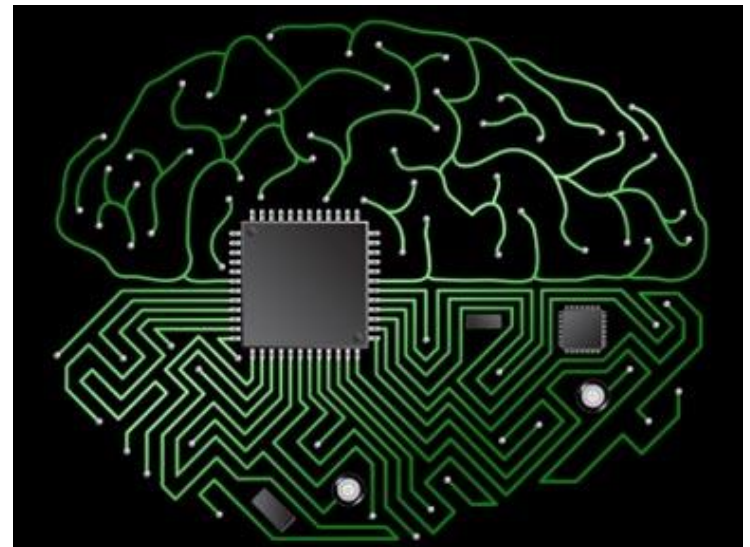
## A mesterséges intelligencia alapjai, alapelvek



„Intelligens, aki többet tud, mint amit tanult!”

*(Gyenes Károly)*

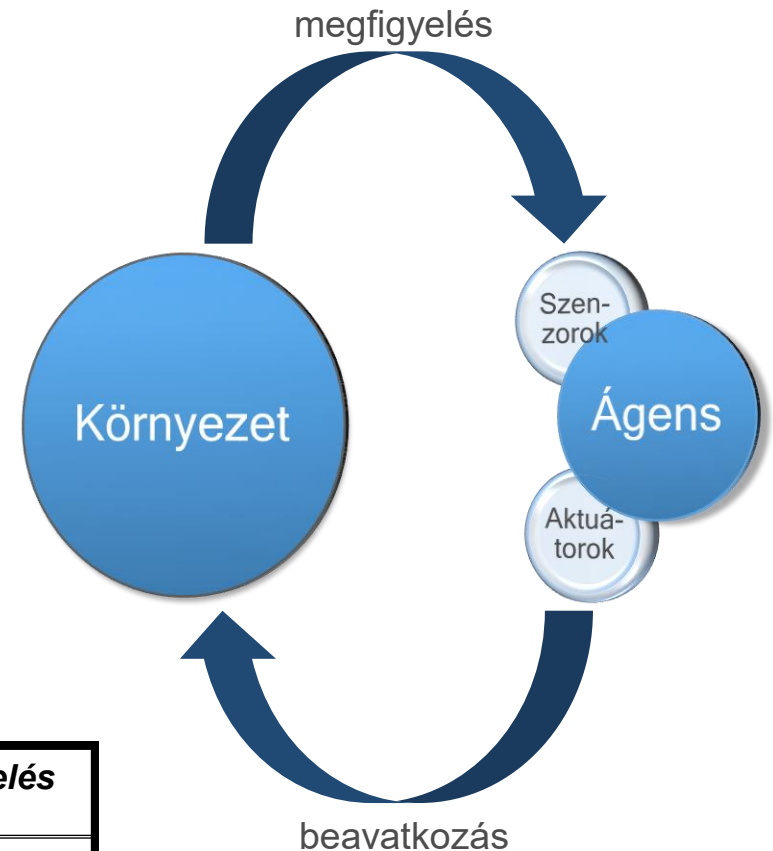
- egy rendszer eléri a kitűzött célokat a környezeti változások ellenére (adaptivitás)
- az emberi intelligencia egyes komponenseinek számítási modellekkel történő megvalósítása
- 1990-es évek: neurális hálók, valószínűségi hálók, business intelligence, machine learning



# Intelligens rendszerek modellje

## Ágensek

- Szenzorok: érzékelők
  - Aktuátorok: beavatkozók
- racionális: adott célok elérésére irányuló cselekvés
  - cél: környezet egy kívánatos állapota
  - állapot elérésének módja: legmagasabb hasznosság alapján



<i>ágens</i>	<i>környezet</i>	<i>beavatkozás</i>	<i>megfigyelés</i>
robot	utca, helyiség	gyorsítás, lassítás, elfordulás	kamera, mikrofon, gyorsulásmérő
diszpécser-támogató eszköz	forgalom-irányító központ	ajánlott döntések	forgalmi helyzet

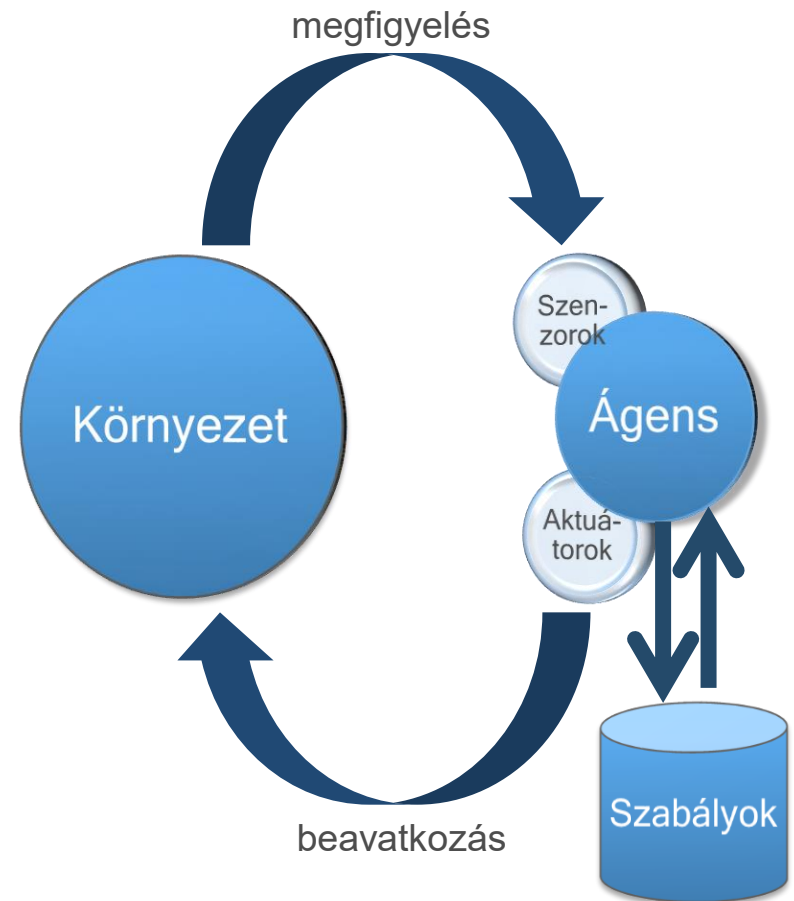
## Ágensek típusai:

- Reflexszerű ágens:

- HA-AKKOR szabály:  
HA állapot, AKKOR cselekvés,

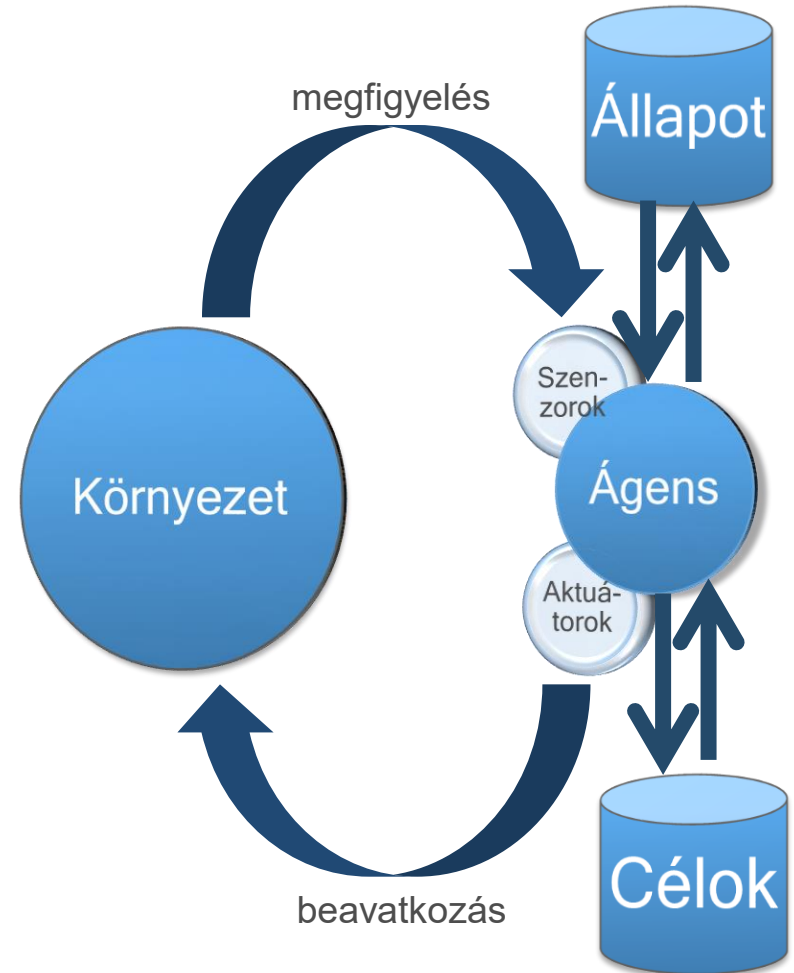
pl: HA az előző autó fékez, AKKOR lassíts!

- implementálás: keresőtábla  
(feltétel-cselekvés szabályok)
- jól működik, ha a helyes döntés  
meghozható az aktuális észlelés  
alapján
- Probléma:
  - fejlesztés komplexitása
  - merevség



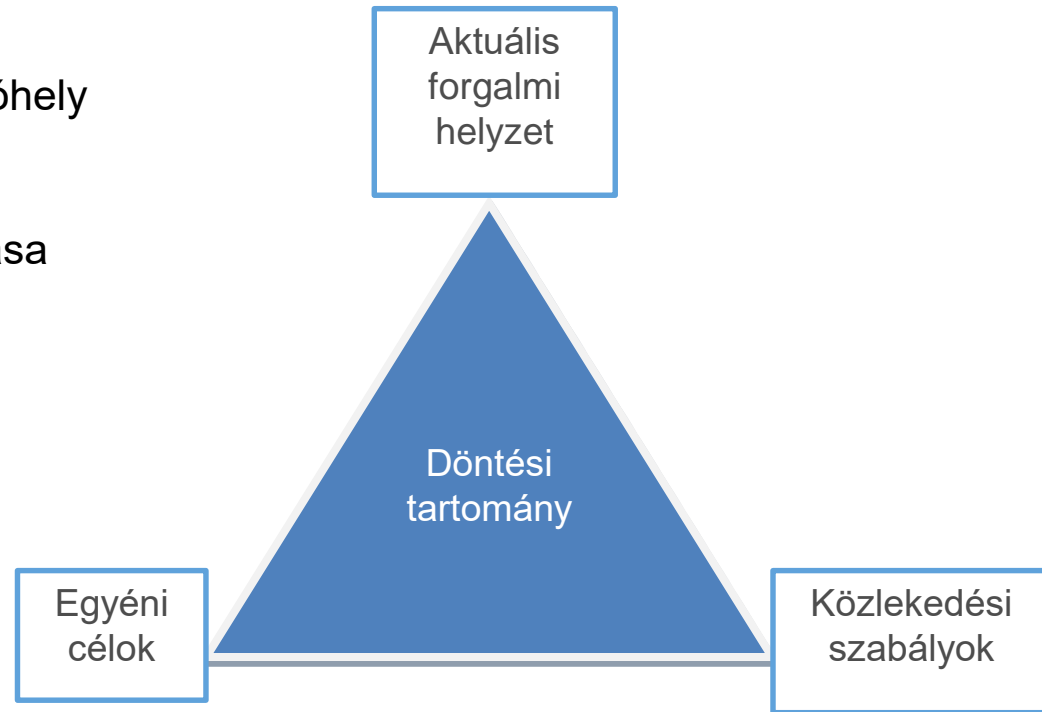
- Cél-orientált ágens:

- a környezet jelenlegi állapotának ismerete nem elég a döntéshez – több cselekvés feltétele egyszerre teljesül
- cél: a környezet egy kívánatos állapota
- modell:
  - összeveti a lehetséges cselekvések eredményeit
  - meghatározza, hogy melyik az a legjobb cselekvés, ami a céljához vezet



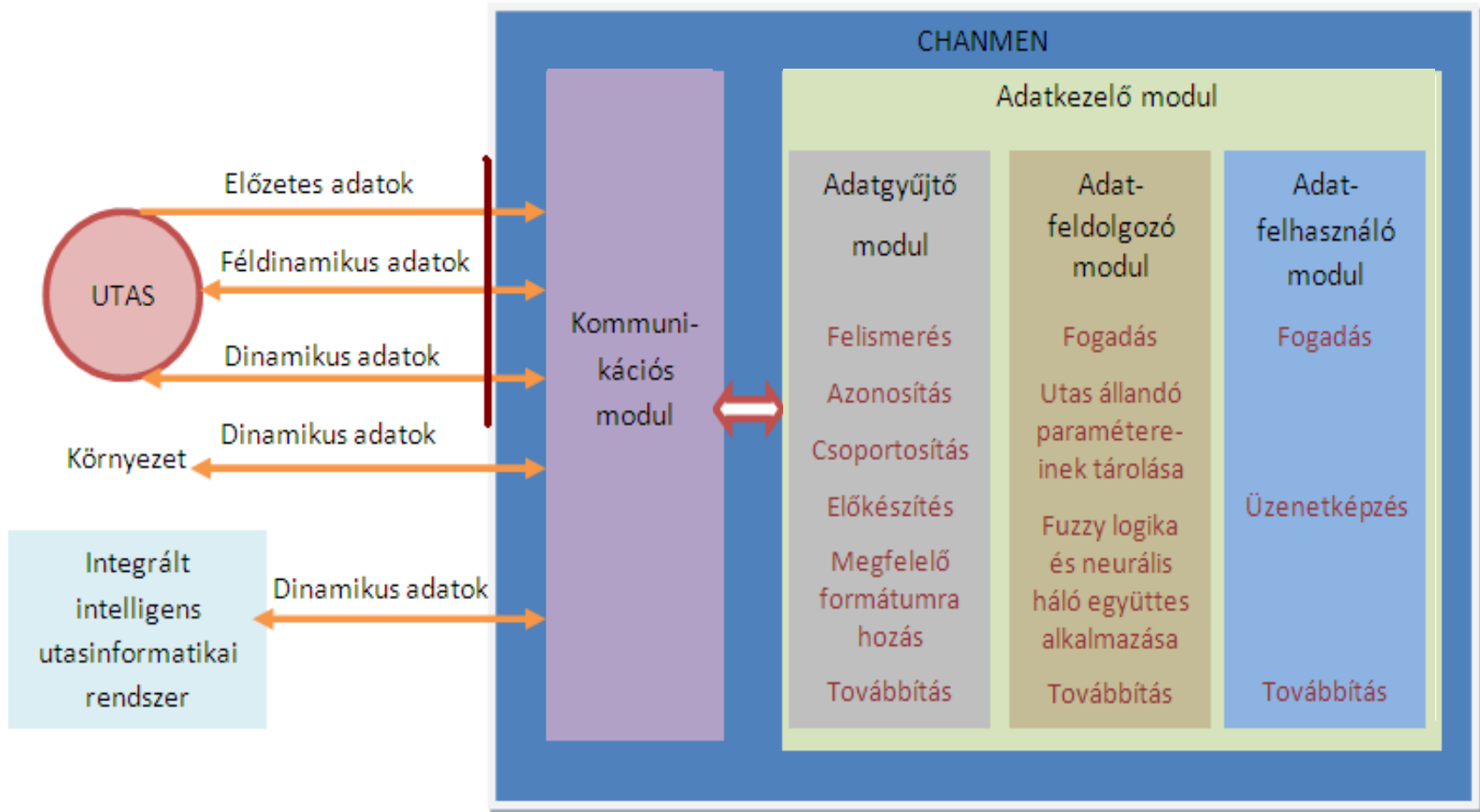
# Átszállás management

- Racionális ágens működésének feltételei:
  - teljesítmény-mérőszám → idő, költség
  - (eddig megfigyelt) észlelési sorozat → utasok korábbi döntései
  - információ környezetről → átszállóhely mikrokörnyezete
  - végrehajtható cselekvések → mozgások, létesítmények látogatása
- Cél:
  - dinamikus (real-time),
  - előre jelzett és
  - személyre szabott
  - információk szolgáltatása



# Átszállás management

Forrás: Lukács Gábor: Átszállás Menedzser - a mesterséges intelligencia alkalmazása a személyek mozgási folyamatainál, TDK dolgozat 2011

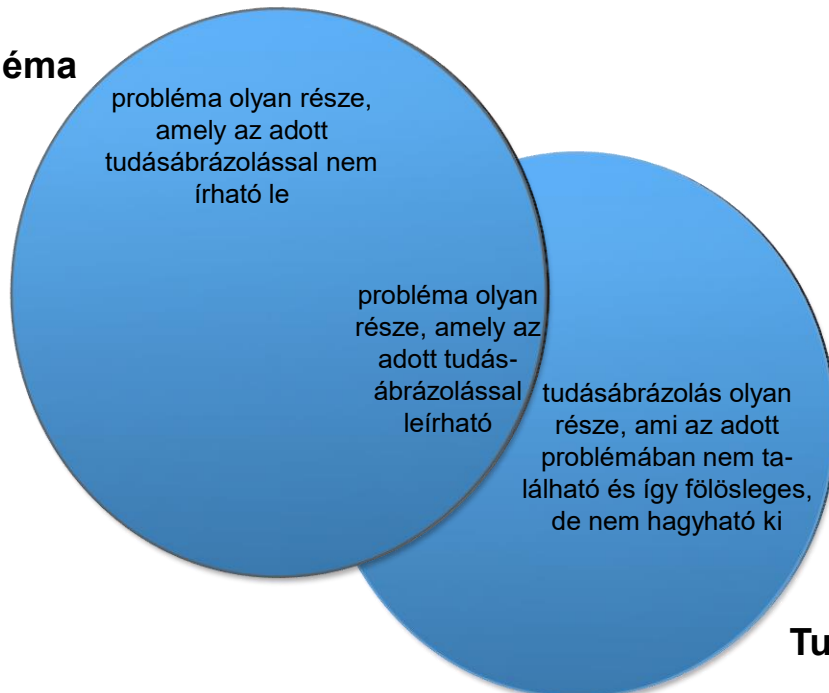


<b>Környezet</b>	<i>teljesen / részben megfigyelhető</i>	<i>determinisztikus / sztochasztikus</i>	<i>díszkrét / folytonos</i>	<i>statikus / dinamikus</i>	<i>epizódszerű / folyamatos</i>	<i>kooperatív / versengő</i>	<i>egyágenses / többágenses</i>
<b>átszállás management</b>	teljesen	sztochasztikus	folytonos	dinamikus	folyamatos	kooperatív	többágenses

# Tudásreprezentálás

- mennyiségek reprezentálása → adatszerkezetek
- manipulálási procedúrák → algoritmusok
- *Kérdések:*
  - Mi a tudás?
  - Hogyan lehet a tudást reprezentálni?
  - Hogyan lehet a tudást manipulálni?
- *Tudásszervezés → tudásbázis építése*
  - Miről beszélünk?  
objektumok, tények, relációk
  - Hogyan kódoljuk a tudást?  
logikai állítások (axiómák)

## Probléma



reprezentáció redukálása egy manipulálható formára, amely lehetővé teszi a célok elérését egy alkalmas következtetési mechanizmus révén



Motiváció

Intelligens  
rendszer  
modellje

- Ágensek
- Környezet
- Tudásreprezentálás

**Valószínűség**

- Döntéshozatal bizonytalan ismeretek mellett
- Valószínűségi hálók
- Fuzzy logika

Gépi tanulás

- Felügyelt tanulás
- Neurális hálók
- Megerősítéses tanulás

## Döntéshozatal bizonytalan ismeretek mellett

- sok esetben a logikai szabályok szükségszerűen hiányosak, mert
  - túl sok feltételt kellene figyelembe venni
  - bizonyos feltételeket egyszerűen lehetetlen ismertnek feltételezni

### Mennyi idő alatt lehet kiérni a Liszt Ferenc repülőtérre?

- 90p-terv
  - a check-in előtt 90 perccel elindulni
  - a kötelező sebességkorlátot betartani
- 90p-terv helyes, kijutunk a reptérre, HACSAK:
  - a kocsival nem történik valami
  - nem fogy ki a benzin
  - nem lesz baleset útközben
  - a gép nem indul korábban
  - nem lesz földrengés

### Helyes a 90p-terv megválasztása?

- helyes, ha  $A1 \wedge A2 \wedge A3 \wedge A4 \wedge \dots \wedge A998 \wedge A999 \wedge A1000 \rightarrow 90p$  logikai állítás bebizonyítható
- kizáró információ hiányában, nagy az „esély”, hogy egy ilyen döntés helyesnek bizonyul

## Baleset példa:

### Elsőrendű logika:

- minden balesetnek ittas vezetés az oka
- DE lehet:
  - figyelmetlenség
  - járműhiba
- ÉS lehet, hogy az ittas vezetésből nem lesz baleset!

	<i>ittas vezetés</i>	<i>nem ittas vezetés</i>
<i>baleset</i>	0,04	0,06
<i>nem baleset</i>	0,01	0,89

baleset és ittas vezetés közötti kapcsolat egyik irányban sem feltétlen logikai következmény!

$$P(B | A) = \frac{P(A | B) * P(B)}{P(A)}$$

**Bayes-tétel**

a világból kiolvasható tudás

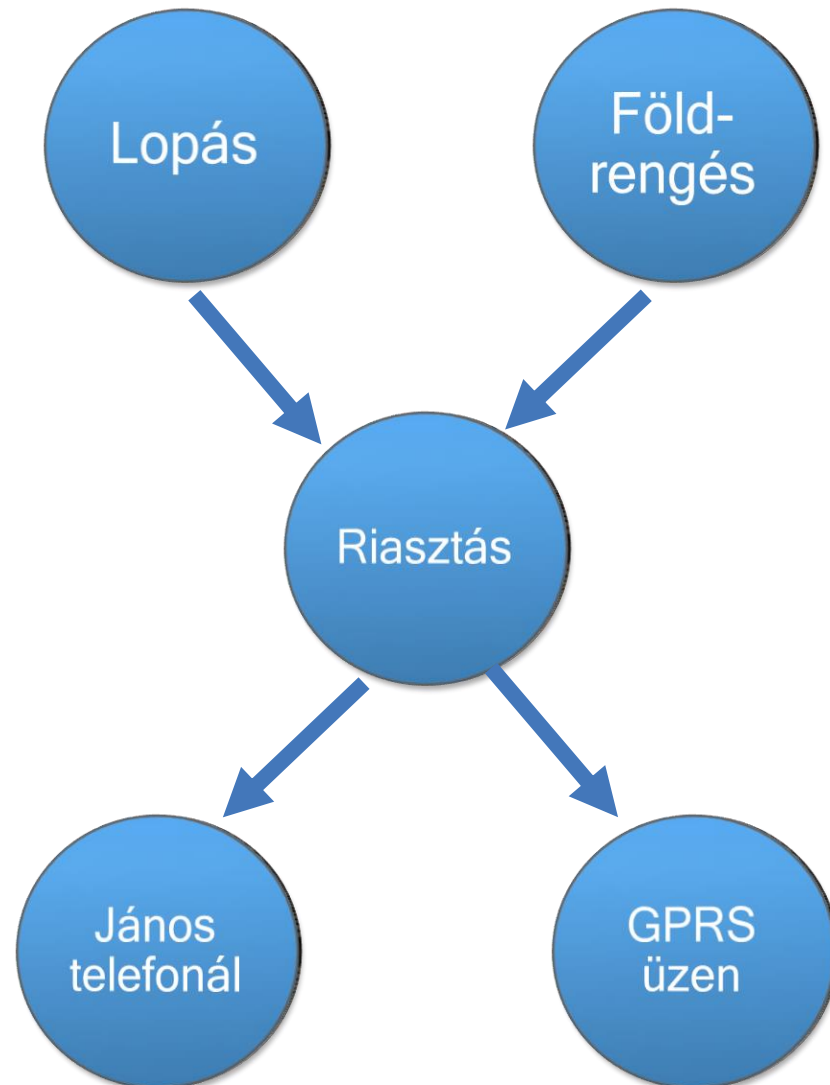
$$P(\text{baleset} | \text{ittas vezetes}) = \frac{P(\text{baleset} \wedge \text{ittas vezetes})}{P(\text{ittas vezetes})} = \frac{0,04}{0,04 + 0,01} = 0,8$$

# Valószínűségi hálók

## Autóriasztó példa – valószínűségi változók

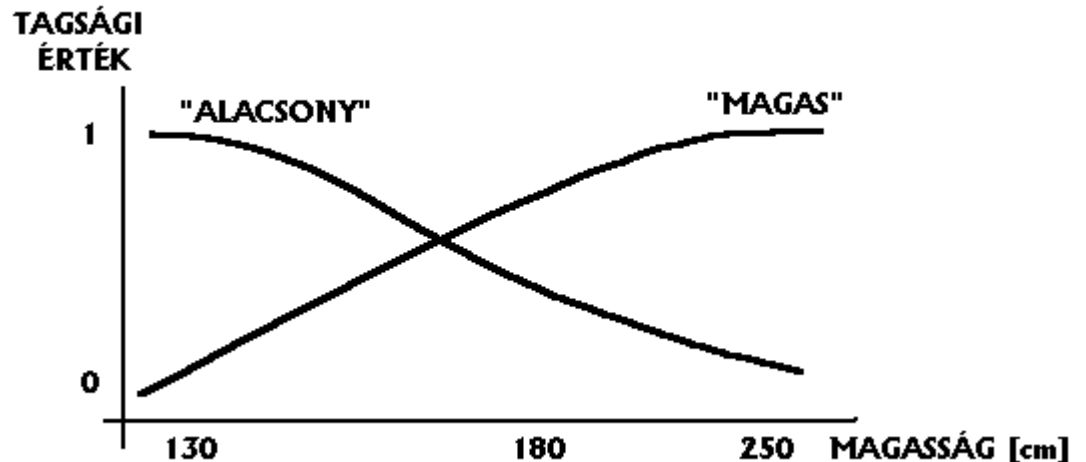
- Autónkba egy új riasztót szereltünk fel.
- Ez megbízhatóan észleli a lopási kísérleteket, de kisebb földrengések esetén is jelez.
- Egy szomszédunk, János megígérte, hogy felhív, ha meghallja a riasztónkat.
- János mindig felhív minket, ha szól a riasztó, de néha összekeveri a telefoncsörgést a riasztó csengésével és ekkor is telefonál.
- Van egy GPRS modul is az autóban, ami üzenetet küld a telefonunkra a riasztó megszólalása esetén.
- Azonban a nem jó a lefedettség arra felé, így néha nem tudja elküldeni az üzenetet.

***A hívások bekövetkezte vagy hiánya alapján szeretnénk megbecsülni a lopás valószínűségét.***



# Fuzzy logika

- egy elem halmazba tartozása nem állapítható meg egyértelműen
- mennyire tartozik bele a halmazba



- Példa:
  - emberek magassága:  $E = \{130, 131, 132, \dots, 183, \dots, 250\}$
  - magas emberek:  $M = \{170, 171, \dots, 250\}$
  - alacsony emberek:  $A = \{130, 131, \dots, 169\}$
- „körülbelül 155 cm magas illető, **nagyjából** alacsonynak mondható”  
 $A = \{130^{(1)}, 140^{(1)}, 150^{(1)}, 160^{(0.8)}, 170^{(0.5)}, 180^{(0.1)}, 190^{(0)}\}$   
 $M = \{130^{(0)}, 140^{(0)}, 150^{(0)}, 160^{(0.2)}, 170^{(0.5)}, 180^{(0.9)}, 190^{(1)}\}$

*nem minden fekete és fehér*

Motiváció

Intelligens  
rendszer  
modellje

- Ágensek
- Környezet
- Tudásreprezentálás

Valószínűség

- Döntéshozatal bizonytalan ismeretek mellett
- Valószínűségi hálók
- Fuzzy logika

**Gépi tanulás**

- Felügyelt tanulás
- Neurális hálók
- Megerősítéses tanulás

## Gépi tanulás definíciója

- környezetéből nyert ismeretek alapján javítja a teljesítőkéességét
- *alkalmazhatóság:*
  - minták állnak rendelkezésre
  - direkt megoldás nem ismert vagy túl bonyolult
  - a környezet változásaira adaptív módon kell reagálni
- *típusai:*
  - **felügyelt tanulás:** az ágens bemenetét és kimenetét is észlelni tudjuk
  - **megerősítéses tanulás:** az ágens az általa végrehajtott tevékenység bizonyos értékelését kapja meg
  - felügyelet nélküli tanulás: semmilyen utalás sem áll rendelkezésünkre a helyes kimenetről

# Felügyelt tanulás

cselekvés =  $f$ (észlelések, régebbi cselekvések, tudásbázis)

- cél:  $f(x)$  felállítása a bemenetek és a kimenetek ismerete alapján

## Helyes döntés tanulása:

True Positive: az autót meg kell javítani és ezt felismerjük

True Negative: az autónak nincs baja és ezt felismerjük

False Negative: az autót meg kell javítani és ezt nem ismerjük fel

False Positive: az autónak nincs baja és ezt nem ismerjük fel

		<i>aktuális érték <math>x</math></i>	
		<u>P</u> ositive	<u>N</u> egative
<i>predikció <math>f(x)</math></i>	P'	True Positive	False Positive
	N'	False Negative	True Negative

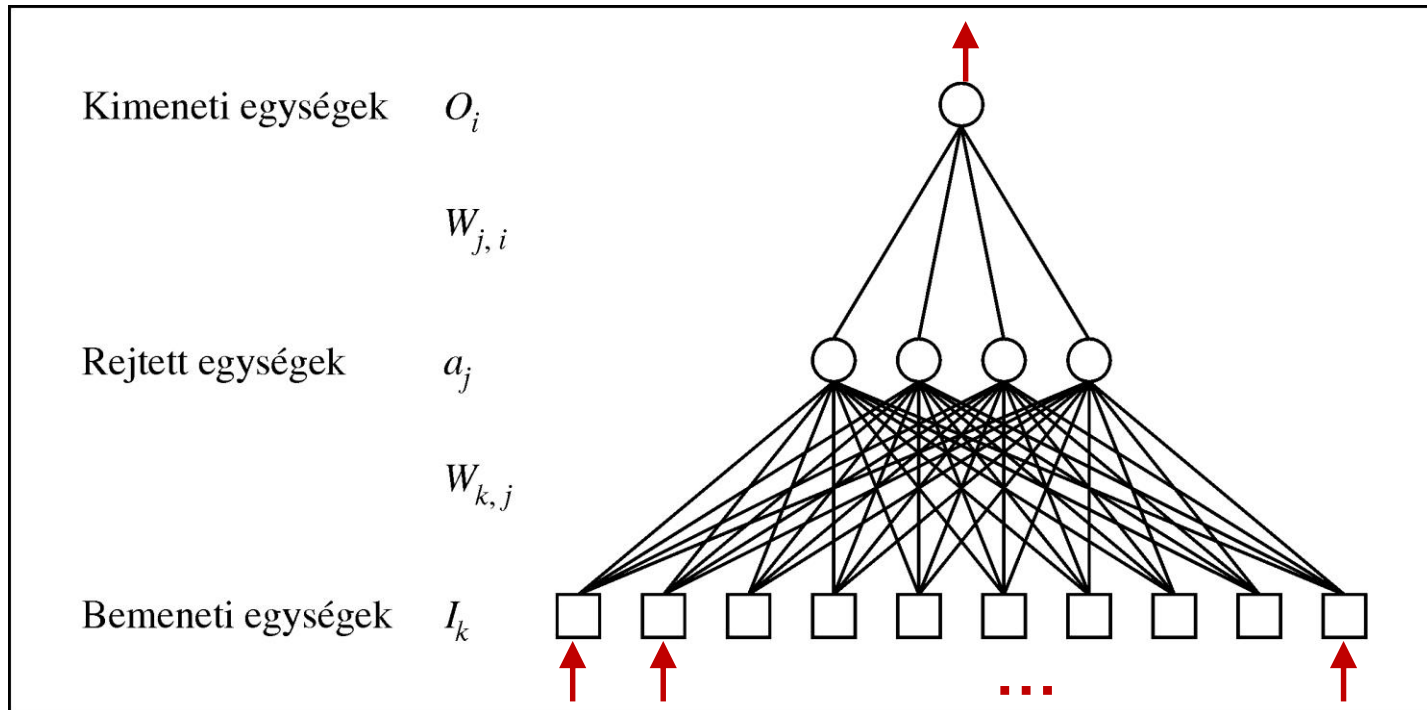
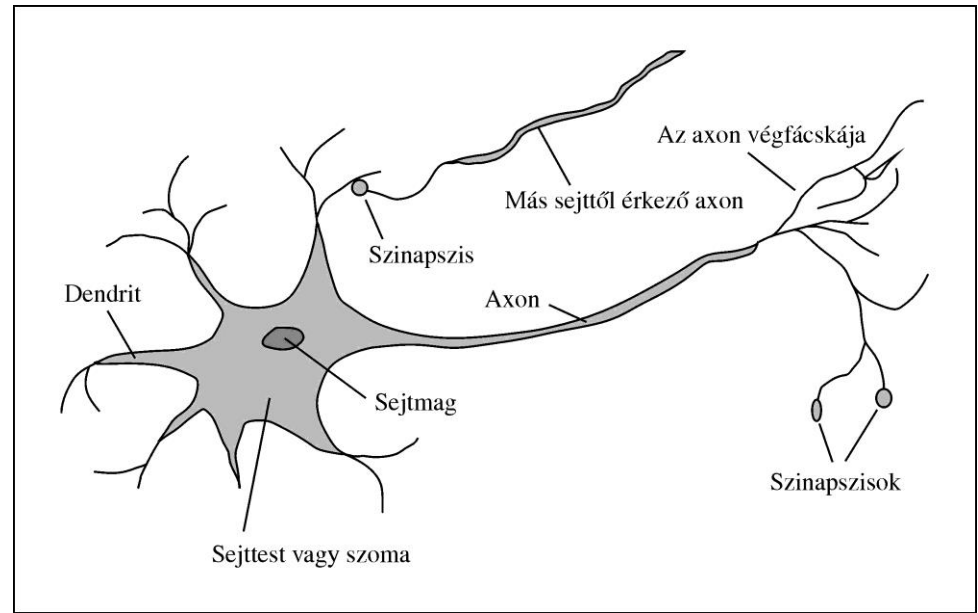
tanító halmaz és teszt halmaz

$$f(x)=x$$



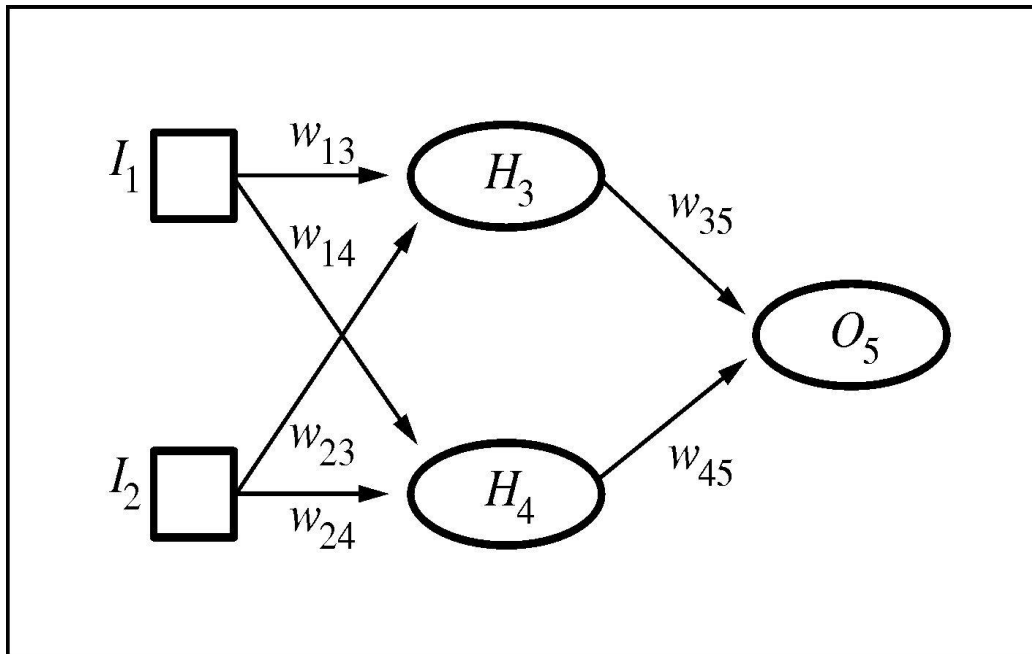
# Neurális hálók

- induktív tanulási algoritmussal tanítható matematikai struktúra
- ahol:  $W$ ..súlyok



**Példa:**

$$\begin{aligned} a_5 &= g(W_{3,5}a_3 + W_{4,5}a_4) \\ &= g(W_{3,5}g(W_{1,3}a_1 + W_{2,3}a_2) + W_{4,5}g(W_{1,4}a_1 + W_{2,4}a_2)) \end{aligned}$$



**Hálózati struktúrák:**

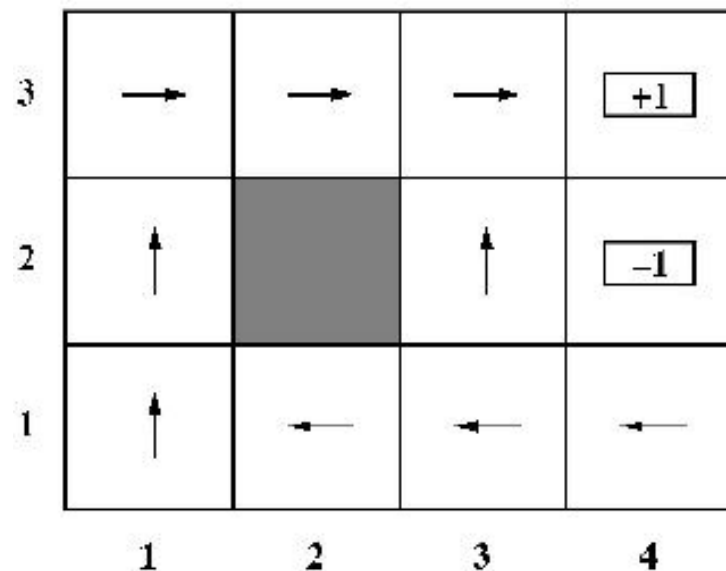
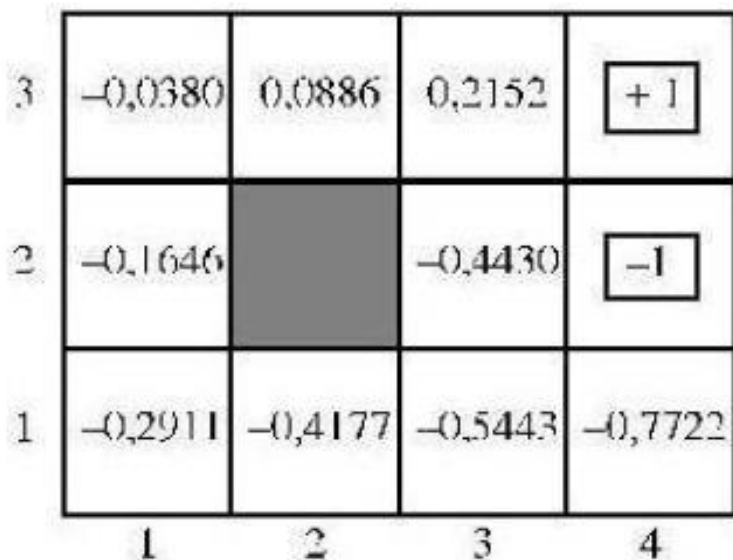
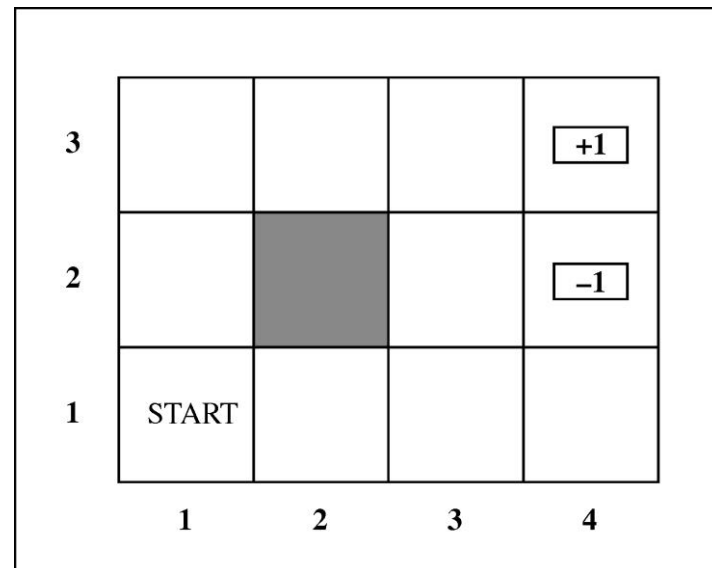
- visszacsatolt
- egy rétegű
- több rétegű
  - rejtett rétegek

# Megerősítéssel tanulás

- ágens sohasem tudja, hogy mik a jó lépések, és azt sem, melyik jutalom melyik cselekvésből ered
- ágens tudása:
  - induláskor tudja a környezetet és cselekvéseinek hatását
  - induláskor nem tudja (*ezt a modellt is meg kell hogy tanulja*)
- megerősítés:
  - csak a végállapotban
  - bármelyik állapotban
- ágens aktivitása:
  - *passzív tanuló*: egyszerűen figyeli a világ alakulását, és a különböző állapotok hasznosságát tanulja
  - *aktív tanuló*: a megtanult információ birtokában cselekszik (*információt cselekvéssel gyűjt*)
- környezet/ágens modell megléte esetén:
  - az ágens az állapotok  $U(i)$  hasznosság függvényét tanulja
  - ennek alapján választja ki cselekvéseit
  - hogy az elérhető hasznosság várható értékét maximalizálja

## Passzív tanulás ismert környezetben:

- az ágens észleli az állapotváltozásokat (tudja, melyik állapotban van)
- tanító szekvencia:
  - az ágens az (1,1) állapotból indul
  - az állapotátmenetek valamilyen sorozatát észleli
  - amíg el nem érkezik a (4,2), illetve a (4,3) végállapotba
  - ahol egy megerősítést kap
- cél: a jutalomról kapott információ alapján  $U(i)$  várható hasznosság (az egyes állapotokban) megtanulása



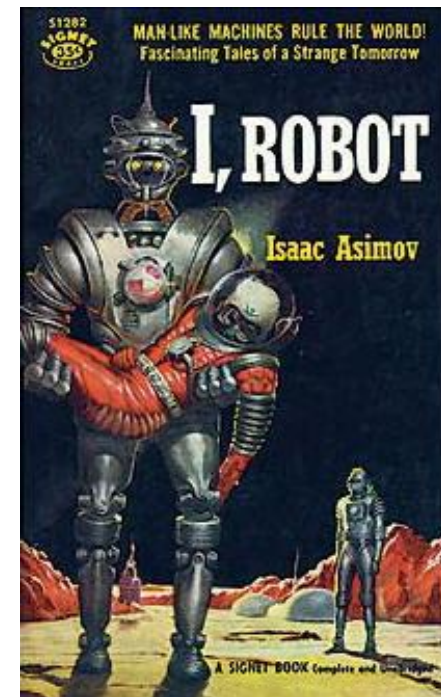
# Principles

## Asimov's Three Laws of Robotics (1942)

1. A robot may not injure a human being or allow a human being to come to harm
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws

*Isaac Asimov (1920-1992):*

*American writer and professor of biochemistry*



## Objectives for Artificial Intelligence (AI) applications of Google (2018):

1. Be socially beneficial
2. Avoid creating or reinforcing unfair bias
3. Be built and tested for safety
4. Be accountable to people
5. Incorporate privacy design principles
6. Uphold high standards of scientific excellence
7. Be made available for uses that accord with these principles

### Avoiding

- technologies that cause or likely to cause harm
- weapons or technologies which cause or facilitate injury to people
- technologies that gather or use sensitive information
- technologies whose purpose is break the law or human rights