

# Mesterséges intelligencia - 1. előadás

---

## 1. fejezet - Bevezetés

---

### Mi az MI?

- 4 fajta megközelítés létezik:
  - Emberi módon cselekedni: Turing-teszt megközelítés
    - Turing-teszt
      - **természetes nyelvfeldolgozás** (*natural language processing*): a párbeszédhez
      - **tudásreprezentáció** (*knowledge representation*): ismert és hallott információ tárolásához
      - **automatizált következtetés** (*automated reasoning*): tárolt infok alapján válaszadás vagy új következtetés
      - **gépi tanulás** (*machine learning*): új körülményekhez való adaptálódás, mintázatok detektálása és általánosítása
    - teljes Turing-teszt
      - **gépi látás** (*computer vision*): az objektumok érzékeléséhez
      - **robotika** (*robotics*): objektumok mozgatásához
    - negatívum: a repülő sem úgy lett, hogy galambot akartunk reprodukálni, hanem aerodinamikát kezdünk tanulmányozni
  - Emberi módon gondolkodni: a kognitív modellezés
    - ehhez az emberi gondolkodást kell elemezni: önelemzés és pszichológiai kísérletek
    - **kognitív tudomány** (*cognitive science*): ezek kutatásával foglalkozó interdiszciplináris terület
      - precíz és verifikálható eredmények megfogalmazása az emberi elme működéséről
  - Racionálisan gondolkodni: a gondolkodás törvénye
    - Arisztotelész -> megcáfolhatatlan következtetési folyamatok törvényekbe foglalása: **szillogisztika** (*syllogisms*)
    - helyes premisszákból mindig helyes következményt adnak -> ebből lett a **logika** (*logic*)
    - **logicista** (*logician*): ezen elvre alapozva akarnak intelligens rendszereket létrehozni
    - elképzelés problémái:
      - formális elemekkel informális tudást kifejezni nehézkes, főleg ha az a tudás nem biztos tudás
      - egy probléma elvi és gyakorlati megoldása közt nagy különbség lehet
      - elképesztően hatalmas lenne az erőforrásigénye
  - Racionálisan cselekedni: a racionális ágens
    - **ágens** (*agent*): valami, ami cselekszik.
      - számítógépes ágens: több, mint egy "mezei" program -> pl autonóm vezérlés felügyelte cselekvés, környezet észlelése, stb
    - **racionális ágens** (*rational agent*): a legjobb (várható) kimenetel érdekében cselekszik

- racionális következtetésnek része a korrekt következtetés, de a racionálisban következtetés nélküli cselekvések is vannak (mint pl forró kályhától a kezünk elkapása)
- a Turing-teszt teljesítéséhez szükséges képességek mind lényegesek a racionális cselekvés megvalósításakor
- két előnye -> miért ez lesz a legnyerőbb?
  - a gondolkodás törvényénél általánosabb, mert a korrekt következtetés csak egy része a racionális ágensnek
  - tudományosan is könnyebb kezelni, mint az emberi megközelítést
    - a racionalitás mértéke jól definiált és teljesen általános (az ember meg egy bonyolult evolúciós folyamat következménye)
- fontos: összetett környezetben tökéletesen racionálisan cselekedni lehetetlen!
  - **korlátozott racionalitás** (*limited rationality*): úgy megfelelően cselekedni, hogy nem áll elég idő az összes kívánt számítás elvégzésére

## A mesterséges intelligencia alapjai

- Filozófia
  - szillogizmusok: korrekt következtetések informális rendszere (Arisztotelész)
  - Pascal: "az aritmetikai gép olyan hatást prűdokol, ami közelebbinek tűnik a gondolathoz, mint az állatok összes cselekvése"
  - Descartes: elme és anyag közti kapcsolat -> elme nem lehet tisztán fizikai, mert nem maradna hely a szabad akaratnak
    - **dualizmus**: elmének egy része (lélek) nem része a fizikai természetnek (nem vonatkoznak rá a fizika törvényei)
    - az állatok nélkülözik ezt a dualista természetet -> kb gépként viselkednek
  - **materializmus**: dualizmus alternatívája, miszerint magának az agynak a fizikai törvényei valósítják meg az elmét
  - tudás forrásának meghatározása
    - Lock: empiricista mozgalom = "nincs semmi a megértésben, ami előbb ne létezne az érzékszervekben"
    - David Hume: **indukció** = az általános elveket a komponenseik ismétlődő kapcsolataiból emeljük ki
    - **logikai pozitivizmus**: minden tudást végső soron az érzékszervi bemeneteknek megfelelő *megfigyeléses állítások*on alapuló logikai elméletekkel meg lehet magyarázni
    - **igazoló elmélet** (*confirmation theory*): miképp gyarapítja a tudást a tapasztalat
  - tudás és cselekvés kapcsolata
    - Arisztotelész: jogosak a cselekvések, mert a célok és a cselekvések kimenetelei között logikai kapcsolat lehetséges
- Matematika
  - George Boole: Boole-logika kidolgozása (görög filozófikusokra alapozva)
  - Gottlob Frege: Boole-logika kiterjesztése objektumokkal és relációkkal
  - Alfred Tarski: referenciaelmélet, melyel a logikai és valós világ objektumait rendelehetjük össze
  - első algoritmus: euklidészi algoritmus (legnagyobb közös osztó kiszámolására)

- algoritmusok tanulmányozása: al-Hvárizmi (perzsa matematikus)
  - David Hilbert (1900): 23 matematikai probléma -> 23.: létezik e természetes számokra vonatkozó tetszőleges logikai állítás igazságát eldöntő algoritmus? (Eldönthetőségi probléma)
  - Gödel: **Nemteljességi tétel** = minden természetes számok tulajdonságait kifejezni képes nyelvben léteznek igaz, de abban az értelemben eldönthetetlen állítások, hogy az igazságuk algoritmikusan nem mutatható ki!
  - Church-Turing-tézis: a Turing-automata képes minden kiszámítható függvényt kiszámítani! (nincs nála jobb)
    - azaz NINCS olyan automata, amely általánosságában el tudja dönteni egy adott program egy adott bemenetre választ fog-e adni, vagy örökké futni fog!
  - **kezelhetetlenség** (*intractability*): ha az adott osztálybeli problémapéldányok megoldásához szükséges idő legalább exponenciálisan nő a problémapéldányok nagyságától függően
  - kezelhetetlen problémák felismerése -> **NP-teljesség**: bármilyen problémaosztály, amire visszavezethető egy NP-teljes problémaosztály, az kezelhetetlen lesz!
  - **valószínűségszámítás** (*probability*): Bayes-tétel és Bayes-analízis -> bizonytalan következtetések megközelítésének alapja
- Gazdaságtan
    - Adam Smith: gazdaságtan mint tudomány, gazdaság = saját gazdasági jólétét maximalizáló különálló ágens
    - Léon Walras: **hasznosság** matematikai kezelése, azaz a kíváncsú eredmények
    - **döntéelmélet** (*decision theory*): valószínűség-elmélet és hasznosságelmélet összekapcsolója -> ez a nagy gazdaságokra igaz (ahol mások hatásait nem kell figyelembe venni)
    - **játékelmélet** (*game theory*) (Neumann és Morgenstern): az egyik játékos cselekvései befolyással lehetnek a másik játékos hasznosságára -> ez a kis gazdaságokra igaz
    - **operációkutatás** (*operational research*): hogyan kell racionálisan dönteni, ha a haszon nem azonnali, hanem több egymást követő cselekvés sorozatának eredménye
      - **Markov döntési folyamat** (*Markov decision process*)
  - Neurális tudományok
    - hogyan keletkeznek a gondolatok az agyban?
    - Paul Broca: beszédzavar kutatása -> beszéd előállítás a bal félteke Broca-féle területén történik -> azaz léteznek konkrét kognitív funkciókért felelős részei az agynak
    - Camillo Golgi: neuronok festési módszere (az agyban)
    - fontosabb következtetések:
      - az agytól eljutunk az elméhez (azaz egyszerű sejtek gyűjteménye elvezethet a gondolathoz, a cselekvéshez és a tudathoz)
      - bár a számítógép a nyers kapcsolási sebességben milliószor gyorsabb, az agy végeredményben százszázszor gyorsabban oldja meg a feladatait
  - Pszichológia
    - probléma: túl szubjektív, nehezen mérhető, DE az állatok mérése viszont könnyű és objektív módszertan lett (H. S. Jenkins)
    - John Watson: állatos nézőpont átvétele az emberekre -> **behaviorista** mozgátlom (*behaviorism*)

- mentális folyamatok minden elméletét elutasította
  - inger-válasz objektív mérce szerinti tanulmányozása
- William James: az agy egy információfeldolgozó eszköz -> **kognitív pszichológia** (*cognitive psychology*)
- Craik: tudás alapú ágens működésének 3 kulcsfontosságú lépése
  - az ingert egy belső reprezentációra le kell fordítani
  - e reprezentációt kognitív folyamatok manipulálják és új belső reprezentációkat származtatnak belőle
  - ezen reprezentációk ismét cselekvésre fordítódnak vissza
- Magyarázat: "Abban az esetben, ha egy szervezet magában hordja a külső valóság és a saját lehetséges cselekvéseinek "kisméretű modelljét", képes különféle alternatívákat kipróbálni, a számára legjobb mellett dönteni, a jövőbeli helyzetekre azok bekövetkezése előtt reagálni, a múltbeli események ismeretét a jelen és a jövő kezelséében felhasználni, és a felmerülő szükségéhselyzetekre minden vonatkozásban kimerítőbb, biztonságosabb és kompetesebb módon reagálni"
- Számítógépes tudományok
  - Turing munkacsoportja (1940): Heath Robinson (német üzenetek dekódolására)
  - Colossus (1943): első általános célú gép Turing-ék által
  - Z-3: első programozható számítógép (1941, Konrad Zuse)
  - teljítőképesség 18 havonta megduplázódik
  - MI kutatás hatására ihletett találmányok
    - időosztásos (time-sharing) operációs rendszer
    - interaktív értelmezők, személyi számítógép albakokkal és egérrel
    - láncolt listás adatszerkezet, automatikus tárolókezelés
    - objektumorientált, szimbolikus, funkcionális és dinamikus programozás több fontos fogalma
- Irányításelmélet és kibernetika
  - Ktesibios (i.e 250): önszabályzó vízi óra -> egy nem élő dolog is képes a viselkedésével a környezeti változások hatására reagálni
  - Watt (18.század): önszabályzó visszacsatolt szabályzó rendszerek
  - sztochasztikus optimális szabályozás -> célfüggvény időben való maximalizálása a cél
- Nyelvészet
  - Chomsky: a behaviorista elmélet nem kezeli a nyelvben meglevő kreativitást
  - **természetes nyelvfeldolgozás** (*natural language processing*): a modern nyelvészet is az MI köztes területe

## A mesterséges intelligencia története

- A mesterséges intelligencia érlelődése
  - Warren McCulloch és Walter Pitts -> mesterséges neuron modell ötlete
    - neuronok vagy bekapcsoltak vagy kikapcsolak
    - bekapcsolás történik, ha kellő másik neuron stimulálja
    - neuron állapota = logikai állítással ekvivalens, ami az ingert kiváltotta
  - **Hebb-tanulás** (*Hebbian learning*): megfelelően kialakított háló képes tanulni
  - 1951: első neurális számítógép = Snarc (40 neuronból álló hálózatot stimulált)
- A mesterséges intelligencia megszületése
  - 1956 - találkozó

- Allen Newell és Herbert Simon: Logic Theorist következtetőprogram
  - itt született meg a **mesterséges intelligencia** név (*artificial intelligence*)
- Korai lelkesedés, nagy elvárások
  - GPS - General Problem Solver: az emberi problémamegoldás protokollját imitálta
    - a részcélok és lehetséges cselekvések számbavételében már hasonlított az emberi gondolkodáshoz
  - **fizikai szimbólumrendszer** hipotézis (*physical symbol system*): "a fizikai szimbólumrendszerek az általános intelligens cselekvés szükséges és elégséges eszközeivel rendelkeznek"
  - Herbert Gelernter: Geometry Theorem Prover: geometriai tételbizonyító program
  - Arthur Samuel: dámajátékot játszó program, ami tanulni is tudott!
  - John McCarthy: kulcsfontosságú eredmények
    - **Lisp**: elsődleges MI-programozási nyelv
    - időosztás kitalálása
    - Advice Taker
      - világra vonatkozó általános tudással rendelkezett
      - új axiómákat tudott elfogadni működése közben
      - ezeknek köszönhetően átprogramozás nélkül is tudott új területeken kompetenciát mutatni
  - **mikrovilágok** (*microworlds*): korlátos problématerületek
  - SAINT: zárt alakra hozható integrálszámítási feladatok elvégzése
  - ANALOGY: IQ-teszt szerű geometriai analógia jellegű problémák megoldása
  - STUDENT: algebrai feladványokat oldott meg
  - kockavilág: egyik legismertebb mikrovilág, ami asztalra helyezett tömör geometriai testekből áll
  - fontos nehézség: Az a tény, hogy egy program egy megoldás megtalálására elvben alkalmas, nem jelenti azt, hogy a program bármi olyan mechanizmust is tartalmaz, amely a megoldás gyakorlati megvalósításához szükséges!
    - a mikrovilág feladatait könnyű megoldani, mert kevés a tény, de már picit nehezebb feladatok megoldása is könnyen kezelhetetlen lesz
  - **gépi evolúció** (*machine evolution*) = **genetikus algoritmus** (*genetic algorithm*): ha egy gépi kódú programot megfelelően kicsi mutációk révén változtatunk, tetszőleges, egyszerű feladatot jól megoldó programhoz juthatunk el.
    - probléma: nem tudjuk leküzdeni a kombinatorikus robbanást
- Tudásalapú rendszerek
  - **gyenge módszer** (*weak method*): pl olyan általános célú keresők, amelyek a teljes megoldás megtalálásának érdekében szekvenciákba fűzték az elemi következtetési lépéseket
  - DENDRAL (Buchanan, 1969)
    - bemenet: molekula alapképlete és tömegspektruma
    - kezdeti verzió: minden lehetséges struktúrát előállított -> nagy molekuláknál ez kezelhetetlen lett
    - megoldás: vegyészek is a jól felismerhető csúcsmintákat keresik
    - ez lett az első tudásintenzív rendszer
  - heurisztikus programozási projektek (HPP) -> **szakértőrendszer**-ek (*expert system*) módszertanát alkalmazva

- MYCIN: vérrel kapcsolatos fertőzések diagnosztizálása
  - nehézség: nem voltak fix szabályok, ehhez orvosokat kellett kikérdezni és az alapján megalkotni ezeket
  - **bizonyossági tényező** (*certainty factor*) volt szükséges
- Az MI iparrá válik
  - 1980-as évek végére berobbant, nagyon népszerű lett a cégek körében
- A neurális hálók visszatérése
  - visszaterjesztéses tanuló algoritmus
  - **konnekcionista** (*connectionist*) modell: a szimbolistákat támadták
- Az MI tudománnyá válik
  - intuíciók helyett szigorúan vett tételek és komoly kísérletek kerültek a célkeresztbe
  - **rejtett Markov-modell** (*hidden Markov model*, HMM)
    - szigorú matematikai elméleten alapulnak
    - valós, nagyméretű beszédgyűjteményt felhasználó tanulási folyamat során hozták létre -> robosztus működés
  - **Bayes-hálók** (*Bayesian networks*): bizonytalan tények hatékony ábrázolása, és következtetés ilyen tények alapján
- Az intelligens ágensek kialakulása
  - olyan ágens, ami az adott szituációban a legjobb cselekvéshez folyamodik

## A mesterséges intelligencia jelenlegi helyzete

- **Autonóm tevékenítés és ütemezés:** NASA Remote Agent programja, az űrhajó műveleteit ütemezi egy Földről küldött magas szintű célból, a műveleteket felügyeli, hibákat detektál, stb
- **Kétszemélyes játékok:** IBM Deep Blue, első számítógépe sakkprogram, ami legyőzte a világbajnok Kaszparov-ot
- **Autonóm szabályozás:** ALVINN, közlekedési sávban tartja az autót, keresztülment az USA-n 98%-ban magától
- **Diagnózis:** diagnosztizálnak, és levezetik, hogy hogy jutottak az adott következtetésre
- **Logisztikai tervekészítés:** DART, automatikus logisztikai tervekészítés és szállítás ütemezése hetek helyett pár nap alatt elkészült
- **Robotika:** mikrosebészet, HipNav, ami leképzi a páciens belső anatómiáját, majd robotszabályozással elvégez egy csípőprotézis behelyezést
- **Nyelvmegértés és problémamegoldás:** PROVERB keresztretjétnyfejtő

## Összefoglalás

- MI 4 megközelítése
- **intelligens ágens:** adott szituációban a legjobb cselekvéshez folyamodik

## Intelligens ágens

### Bevezetés

- **ágens** (*agent*): egy olyan valami, ami az **érzékelői** (*sensors*) segítségével érzékeli a **környezetét** (*environment*), és **beavatkozái** (*actuators*) segítségével megváltoztatja azt!
  - pl emberi ágens: érzékelői-> (szem, fül, stb), beavatkozái-> (kéz, láb, száj)
  - pl robot ágens: érzékelői-> (kamera, infra távolságérzékelő), beavatkozái-> (robotkar)
  - pl szoftveres ágens: érzékelői-> (billentyűleütés, fájl tartalom, hálózati adatcsomag), beavatkozái-> (képernyőn kijelzés, fájl írása, hálózati csomag küldése)

- **érzékelés** (*percept*): ágens érzékelő bemeneteinek leírása egy tetszőleges pillanatban
- **érzékelési sorozat** (*percept sequence*): az ágens érzékeléseinek teljes története
  - azaz HA az összes lehetséges érzékelési sorozathoz meg tudjuk határozni az ágens lehetséges cselekvéseit, akkor mindent elmondtunk az ágensről
- **ágensfüggvény** (*agent function*): adott érzékelési sorozatot egy cselekvésre képez le → *absztrakt matematikai leírás*
- **ágensprogram** (*agent program*): ami az ágensfüggvényt, a mesterséges ágens belsejében megvalósítja → *konkrét implementáció*
- példa: porszívóvilág
  - A és B helyszín létezik
  - ágensfüggvény: ha az aktuális négyzet koszos, szív fel a port, különben menj át a másik négyzetbe
- **S<sub>Intelligens Rendszer</sub>  $\subseteq S_K \times S_A$** , ahol az ágens **környezetének**  **$S_K(t)$** , az **ágensnek** magának pedig  **$S_A(t)$**  állapotai vannak
- **trajektória**: Érzékelés és cselekvés közben egy ágens egy **trajektória** mentén halad
  - létezik hatékonyabb, és kevésbé hatékonyabb trajektória egy ágens céljának elérésére

## Jó viselkedés: a racionalitás koncepciója

- **racióális ágens** (*rational agent*): olyan ágens, amely helyesen cselekszik (pl az ágensfüggvény táblázatának minden bejegyzése helyesen van kitöltve)
  - mit jelent helyesen cselekedni? -> helyes az, ami az ágens a legsikeresebbé teszi → mérjük a sikerességet
  - **teljesítménymérték** (*performance measure*): ágens sikerességének kritériuma
    - megkérdezhetjük az ágensről, hogy elégedett-e a teljesítményével -> probléma: egyesek nem tudják megállapítani, mások alulbecsülik
    - => objektív teljesítménymérték kell, amit a tervező határoz meg
      - pl 8órás műszakban feltakarított por mennyisége -> ha mindig kiborítja, akkor lesz maximum :(
      - pl a tiszta padló díjazása -> ez jobban hangzik, bár a végrehajtás módjára nem tér ki :)
      - pl a tiszta padló díjazása minél rövidebb ideig tartó zaj mellett, kevés energia felhasználásával -> na ez már böki :D
    - a teljesítménymértéket jobb aszerint megállapítani, hogy mit akarunk elérni a környezetben, mint aszerint, hogy miképp kellene az ágensnek viselkednie
  - **racionalitás**
    - egy adott pillanatban való racionalitás befolyásolói:
      - a siker fokát mérő teljesítménymérték
      - az ágens eddigi tudása a környezetről
      - a cselekvések, amiket az ágens képes végrehajtani
      - az ágens érzékelési sorozata az adott pillanatig
    - **racióális ágens** (*rational agent*): az ideális racióális ágens minden egyes észlelési sorozathoz a benne található tények és a beépített tudása alapján minden elvárható dolgot megtesz a teljesítménymérték maximalizálásáért
    - porszívós példa:

- **teljesítménymérték:** egy pont plusz minden tiszta négyzetért minden időpillanatban (1000 időpillanat egy élettartam)
- **eddiggi tudás a környezetről:** a priori ismert, de a piszok eloszlása és az ágens kezdeti pozíciója nem. A tiszta négyzet tiszta is marad, a felszívás megtisztítja az adott négyzetet. A Balra és Jobbra cselekvés balra és jobbra mozgatja az ágens, kivéve ha az kikerülne a környezetből, ekkor az ágens helyben marad.
- **ágens lehetséges cselekvései:** Balra, Jobbra, Szív, Semmittevés
- **ágens érzékelési sorozata az adott pillanatig:** az ágens helyesen érzékeli jelenlegi helyzetét és azt, hogy van e ott kosz.
- ha ezekből bármi is hiányozna, akkor irracionális lenne az ágens (pl oszcillálna jobbra balra, kimenne a pályáról, stbstb)
- **Mindentudás, tanulás és autonómia**
  - **mindentudás** (*omniscience*): tudja cselekedetei valódi kimenetelét, DE gyakorlatban a mindentudás lehetetlen!
  - **racionalitás** -> elvárt teljesítményt maximalizálja, DE a **tökéletesség** -> tényleges teljesítményt maximalizálja
  - racionalitásnál fontos az **információgyűjtés** (*information gathering*): hasznos információk beszerzése a teljesítmény maximalizálásának céljából
    - a **felfedezés** (*exploration*) is egyfajta információgyűjtés, ami a teljes környezet megismerését foglalja magába
  - **tanulás** (*learn*): ha a környezet (priori = kezdeti környezet) nem teljesen ismert, a megfigyelések alapján tanulni kell
  - sikeres ágens ágensfüggvény kiszámításának 3 periódusa:
    - ágens tervezésekor (a tervezők végzik)
    - következő cselekvés megfontolásakor (az ágens végzi)
    - tapasztalatokból tanuláskor (az ágens végzi)
  - egy ágens **nem autonóm**, ha a tervezői által beépített tudásra épít, és nem a saját megfigyeléseire
    - egy racionális ágensnek autonómnak kell lenni -> így függetlenedhet a készítői által belerakott kezdeti tudástól
- **intelligens ágens:** racionális módon választja meg a cselekvéseit és a célállapotait sikeresen éri el

## A Környezetek természete

### A környezet meghatározása

- **feladatkörnyezetek** (*task environments*): ezek a problémák, amikre a racionális ágensek jelentik a megoldásokat
  - TKBÉ
    - **Teljesítmény** (*Performance*): pl helyes célállomás elérése, üzemanyag fogyasztás minimalizálása, út költségének vagy idejének minimalizálása, közlekedési szabályok betartása, stb
    - **Környezet** (*Environment*): pl vidéki és városi út, forgalom, gyalogosok, kóbor állatok, útkarbantartás, időjárási viszonyok
    - **Beavatkozók** (*Actuators*): pl mint az ember számára, azaz motor és gázpedál vezérlése, kormányzás és fékezés, képernyő és egyéb járművekkel és az utassal kommunikálás

- **Érzékelők** (*Sensors*): pl tv-kamerák, sebességmérő, kilométeróra, gyorsulásmérő, GPS, hangradar, infra-érezkelő
- ágens első lépése a **feladatkörnyezet lehető legteljesebb meghatározása**

## A környezet tulajdonságai

- Megfigyelhetőség
  - **Teljesen megfigyelhető** (*fully observable*): ha az ágens szektorai minden pillanatban hozzáférést nyújtanak a környezet teljes állapotához
  - **Részlegesen megfigyelhető** (*partially observable*): zaj vagy pontatlan szenzorok miatt, vagy ha az állapot egyes részei nem szerepelnek a szenzorok adatai között
- Determinisztikusság
  - **Determinisztikus** (*deterministic*): Ha a környezet következő állapotát jelenlegi állapota és az ágens által végrehajtott cselekvés teljesen meghatározza
  - **Sztokasztikus** (*stochastic*): Ha ez nem áll fent.
  - **Stratégiai** (*strategic*): Ha a környezet más ágensek cselekvéseit leszámítva determinisztikus
- Lefolyása szerint
  - **Epizód szerű** (*episodi*): Ágens tapasztalata epizódokra bontható, minden epizód az ágens észleléseiből és a cselekvéséből áll, és a következő epizód nem függ az előzőekben végrehajtott cselekvésektől
  - **Sorozatszerű** (*sequential*): Sorozatszerű környezetben az aktív döntés befolyásolhat minden továbbit
- Környezet változásának szempontjából
  - **Statikus** (*static*): Ha a környezet nem változhat meg, amíg az ágens gondolkodik
  - **Dinamikus** (*dinamic*): Ha a környezet megváltozhat, amíg az ágens gondolkodik, ezek a környezetek állandóan azt kérdezik az ágenstől, hogy mit akar tenni
  - **Szemidinamikus** (*semidynamic*): Ha a környezet nem változik az idő előrehaladtával, de az ágens teljesítménymértéke igen. (Azaz ha sokat gondolkodik, nem lesz elég hatékony!)
- Diszkrét/folytonos felosztás: alkalmazható a környezet állapotára, az időkezelés módjára, az ágens észleléseire, valamint cselekvéseire
  - **Diszkrét** (*discrete*): Véges számú külön álló állapottal rendelkezik (pl sakkjáték, ahol az akciók és cselekvések száma is diszkrét)
  - **Folytonos** (*continuous*): A taxivezetés pl folytonos állapotú és idejű probléma (sebesség, más járművek helye), és az akciói is folytonosak (kanyarodás szöge)
- Ágensek száma alapján: könnyűnek tűnik, de bonyolult is lehet -> Egy A ágensnek egy B objektumot ágensnek kell tekintenie, ha B viselkedése legjobban egy A viselkedésétől függő teljesítménymérték maximalizálásával írható-e le.
  - **Egyágenses** (*single agent*): pl keresztrejtvény
  - **Többágenses** (*multiagent*): pl sakk
    - **versengő** (*competetive*): pl a sakk, ahol a két ágens célja egymás legyőzése
    - **kooperatív** (*cooperative*): pl taxivezetésben a többi autó célja is, hogy ne karambolozzanak (de pl a parkolóhelyekért már versengenek) **környezetosztály** (*environment class*): egy ágens lehetséges környezeteinek gyűjteménye (pl teszteléshez)
- Legnehezebb: nem hozzáférhető, nem epizódszerű, dinamikus, nem determinisztikus és folytonos, többágenses környezet.
- Ágens "ellenségei"

1. véges erőforrásai (rendelkezésre álló időt is beleértve)
2. információhiány érzékeléskor
3. a környezet változékonyága

## Az intelligens ágensek struktúrája

- Hogyan működik belülről egy ágens
- Mesterséges intelligencia feladata -> ágensprogram megtervezése: egy függvényé, ami megvalósítja az észlelések és a cselekvések közti leképztést.
- Feltételezés: a program egy megfelelő érzékelőkkel és beavatkozókkaal ellátott eszközön fut -> ez az **architektúra** (*architecture*)
- **ágens = architektúra + program**

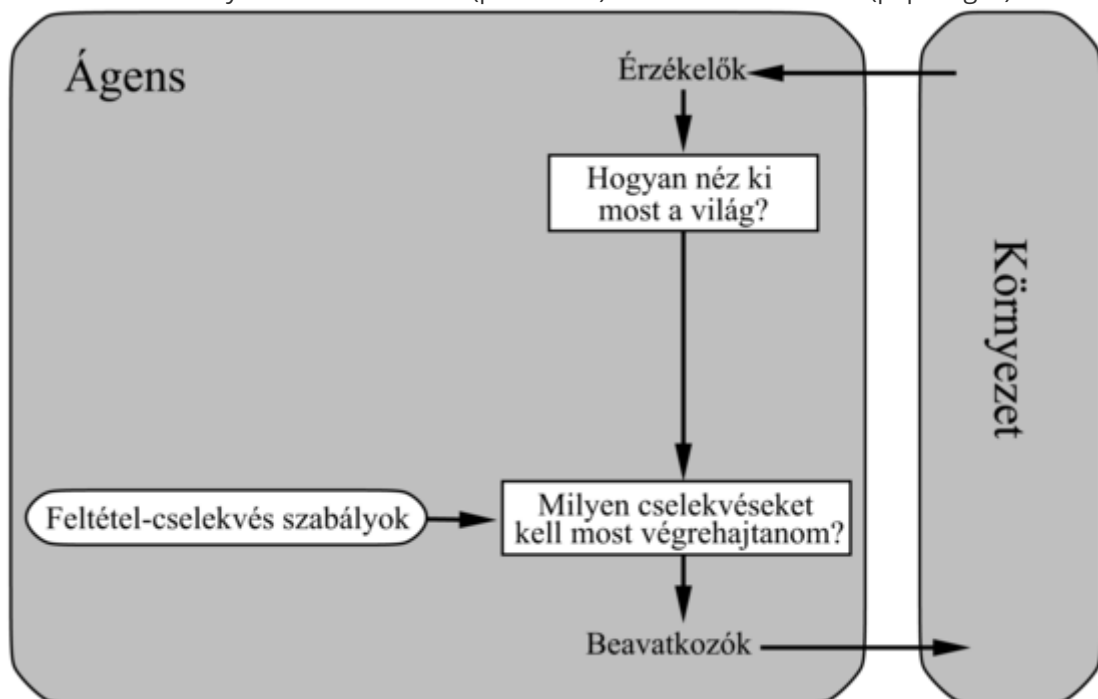
## Ágensprogramok

- Ágensprogram -> aktuális észlelést veszi bemenetként
  - ha teljes észleléstől függene -> ágensnek emlékeznie is kell tudni
- Ágensfüggvény -> teljes észlelési történetet fogadja
- **Táblázat-vezérlésű-ágens**
  - minden észlelésre meghívódik, és egy cselekvést ad vissza.
  - Rendelkezik egy táblázattal (ágensfüggvény), amiből kikeresi az adott észleléshez tartozó cselekvést, majd azzal tér vissza
  - `function TÁBLÁZAT_VEZÉRLÉSŰ_ÁGENS(észlelés)` returns egy cselekvés  
 static: észlelése, egy sorozat, kezdetben üres  
         táblázat, egy táblázat, az észlelési sorozat indexeli,  
 kezdetben teljesen feltöltött  
  
         csatold az észlelést az észlelések végére  
         cselekvés <- KIKERESÉS(észlelések, táblázat)  
         return cselekvés
  - Táblázat vezérelt megközelítés negatívumai:
    - P - lehetséges észlelések halmaza
    - T - az ágens élettartama
    - Táblázat elemszáma:  $\sum_{t=1}^T |P|^t$
  - Ez nagyon nagyon sok...
  - **Terv**: váltsuk fel a nagy táblázatokat olyan rövid programmal, ami ugyan azt adja (mint a Newton-módszer a négyzetgyök táblázatokra)
- Ezek alapján -> alapvető ágensprogram típusok
  - **Egyszerű reflexszerű ágensek**
  - **Modellalapú reflexszerű ágensek**
  - **Célorientált ágensek**
  - **Hasznosságorientált ágensek**

## Egyszerű reflexszerű ágensek (*simple reflex agent*)

- aktuális észlelés alapján döntenek → figyelmen kívül hagyják az észlelési történet többi részét
- pl.: porszívóágens ilyen volt (csak az számít, hogy adott mezőn van e piszok)
- kezelhető méretű táblázat okai:
  - észlelési történet figyelmen kívül hagyása →  $4^T$  helyett 4 eset
  - ha az aktuális négyzet piszkos, nem számít a helyszín
- ```
function REFLEXSZERŰ_PORSZÍVÓ_ÁGENS(helyszín, állapot) returns egy cselekvés

    if állapot = Piszos then return Felszívás
    else if helyszín = A then return Jobbra
    else if helyszín = B then return Balra
```
- feltétel-cselekvés szabály** (*condition-action rule*): ha [az-előző-autó-fékez] akkor [kezdj-fékezni]
  - emberekben ilyen a **tanult válasz** (pl vezetés) és a **feltétlen reflex** (pl pislogás)

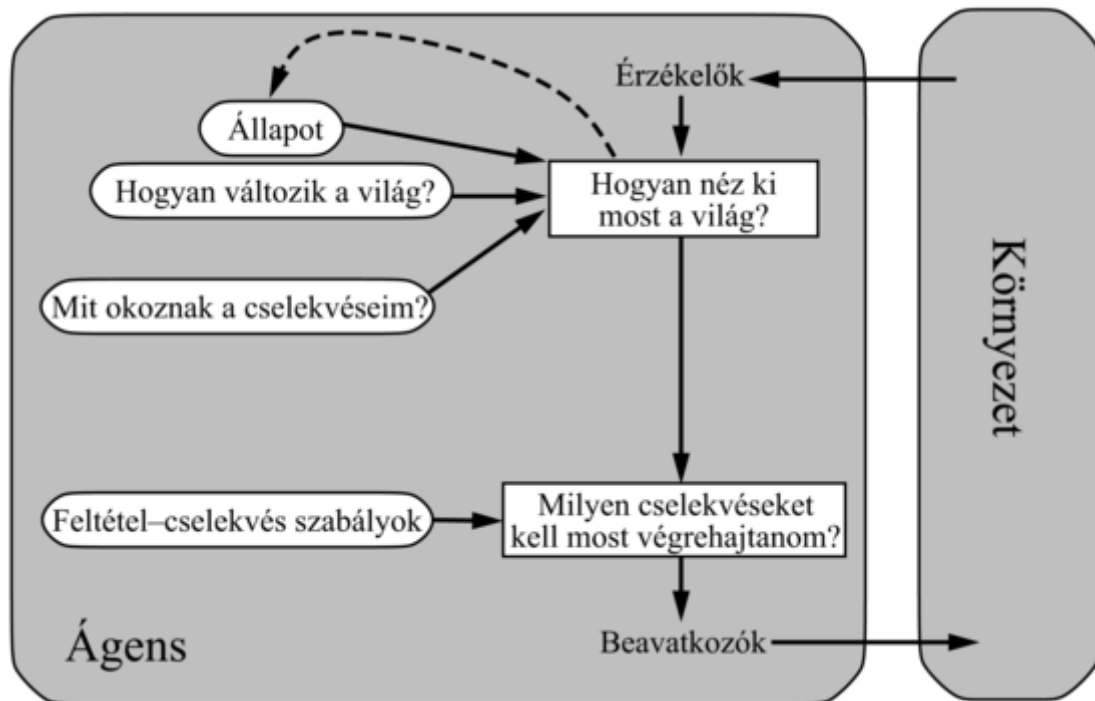


- ```
function EGYSZERŰ_REFLEXSZERŰ_ÁGENS(észlelés) returns egy cselekvés
    static: szabályok, feltétel-cselekvés szabályok halmaza

    állapot <- BEMENET_FELDOLGOZÁS(észlelés) //bemenetből állapotot csinál
    szabály <- SZABÁLY_ILLESZTÉS(állapot, szabályok) //első, adott állapotra
    illő szabály
    cselekvés <- SZABÁLY_CSELEKVÉS[szabály] //cselekszik
    return cselekvés
```
- Összegzés: egyszerű, de korlátozott intelligenciájú. Kritikusan fontos neki a környezet pontos megfigyelhetősége, különben könnyen végtelen ciklusba kerül!

## Modellalapú reflexszerű ágensek

- részletes megfigyelhetőség kezelése → a világ jelenleg nem látható részének nyomon követése
  - **belső állapot** segítségével (*internal state*)

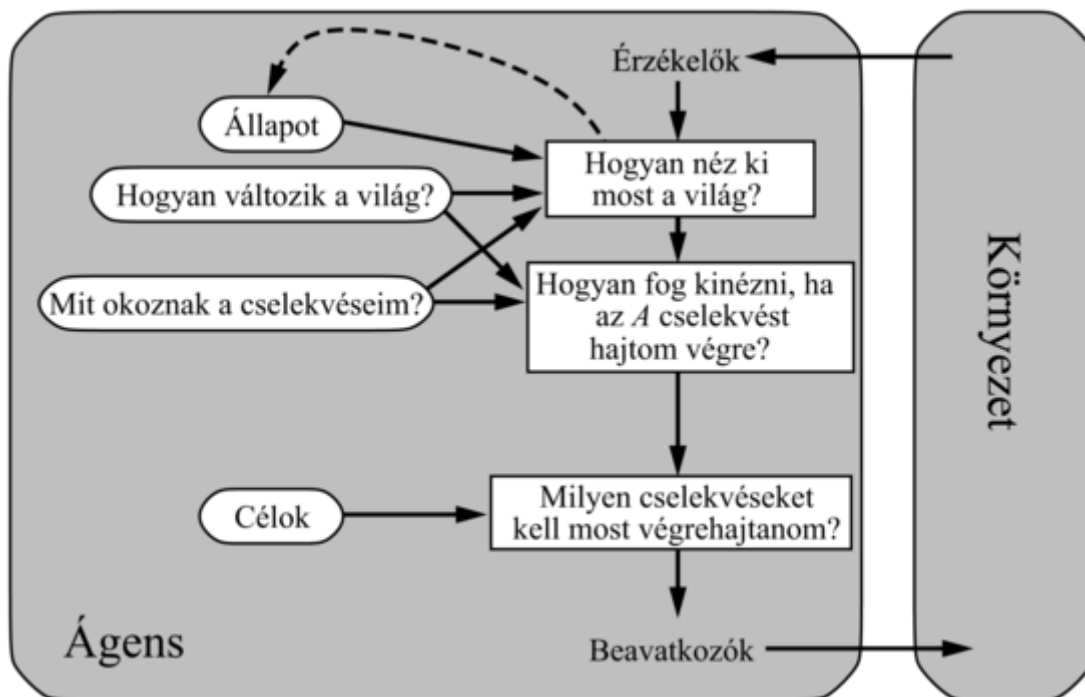


- `function REFLEXSZERŰ_ÁGENS_ÁLLAPOT(észlelés)` returns egy cselekvés  
 static: állapot, a világ jelenlegi állapotának leírása  
 szabályok, feltétel-cselekvés szabályok halmaza  
 cselekvés, a legutolsó cselekvés, kezdetben semmi  
  
`állapot <- ÁLLAPOT_FRISSÍTÉS(állapot, cselekvés, észlelés) //belső`  
`állapot leírásáért felel`  
`szabály <- SZABÁLY_ILLESZTÉS(állapot, szabályok)`  
`cselekvés <- SZABÁLY_CSELEKVÉS[szabály]`  
`return cselekvés`

- **modellalapú ágens** (*model-based agent*): olyan ágens, ami magában tárolja a világ működésének mikéntjéről szóló tudást (vagyis egy **modellt**)

## Célorientált ágensek

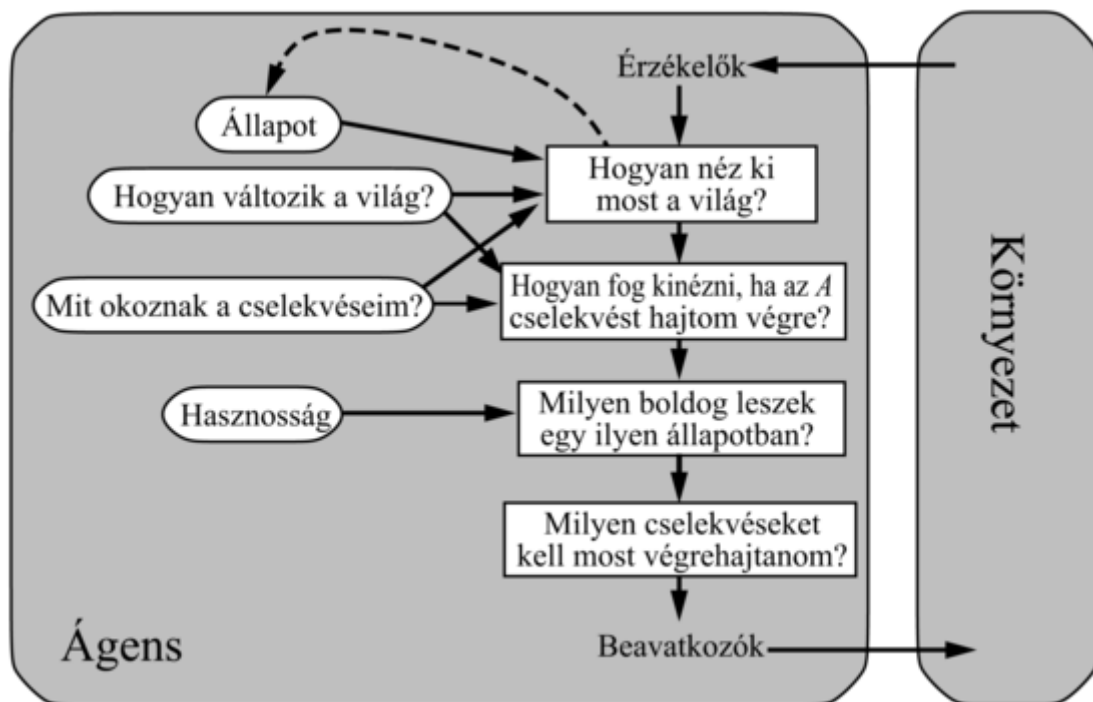
- sokszor szükség van valamilyen **célra** (*goal*), pl taxi a kereszteződésben a cél fele tartó utat válassza
-



- célorientált cselekvés választás lehet egyszerű (következő lépés a cél), és nagyon bonyolult is
  - cél eléréséhez megfelelő cselekvéssorozat megtalálása: **keresés** (*search*) és **tervkészítés** (*planning*)
- Összegzés: a Cél elérése leváltotta a Feltételes-cselekvés szabályokat, mivel ez már nem reflexszerű megközelítés, hanem a jövő figyelembe vételével is foglalkozik! Ezekből következik, hogy a **célorientált** megközelítés sokkal rugalmasabb (esőben nem fog blokkolva fékezni)

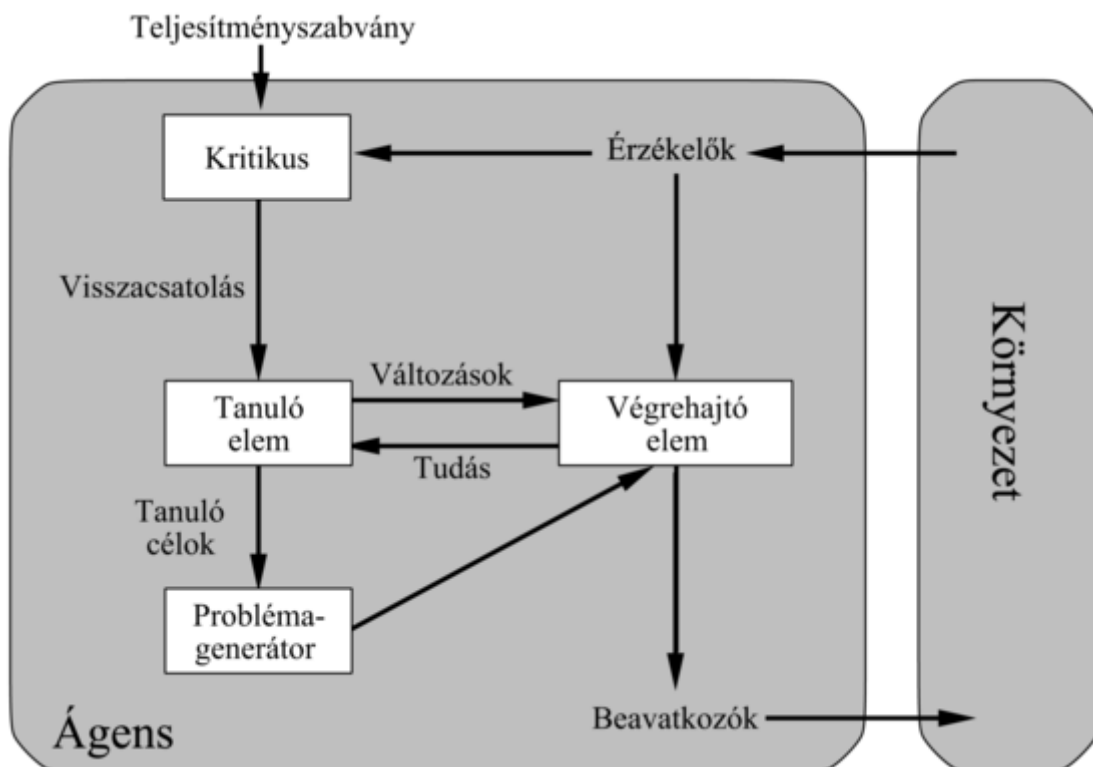
### Hasznosságorientált ágensek

- cél elérése nem feltétlenül elég, fontos a **teljesítménymérték** is
- **hasznosság** (*utility*): a világ egyik állapota előnyösebb egy másikhoz képest, ha nagyobb a **hasznossága** az ágens számára
- **hasznosságfüggvény** (*utility function*): állapotot egy valós számra képez le, ami a hozzá rendelt boldogság fokát jelzi
  - ha egymásnak ellentmondó célokat kéne teljesíteni → ekkor helyes kompromisszumot határoz meg a hasznosságfüggvény
  - ha több cél van, de egyikük sem érhető el teljes bizonyossággal → olyan módszert ad, amivel a siker valószínűsége a célok fontosságához mérhető.



### Tanuló ágensek

- 4 koncepcionális komponenssel rendelkezik
  - **tanuló elem** (*learning element*): javításokért felel
  - **végrehajtó elem** (*performance element*): külső cselekvések kiválasztásáért felel
    - észleléseket végzi, cselekvésről dönt
  - **kritikus** (*critic*): ágens működéséről szóló visszajelzést ad, milyen jól működik az ágens egy rögzített teljesítményszabványhoz viszonyítva
  - **problémagenerátor** (*problem generator*): olyan cselekvések javaslása, ami új és informatív tapasztalatokhoz vezet. Ő felel e kísérletezésért, felfedezésért!



- érdekes eset: **hasznosságinformáció megtanulása** → a teljesítményszabvány a beérkező észlelés egy részét elkönyveli **jutalomként** (*reward*) vagy **büntetésként** (*penalty*)

## Összefoglalás

- **ágens** (*agent*): olyan valami, amely észlel és cselekszik egy környezetbe.
- **ágensfüggvény** (*agent function*): meghatározza, hogy egy tetszőleges észlelési sorozathoz milyen cselekvést hajtson végre az ágens.
- **teljesítménymérték** (*performance measure*): értékeli az ágens viselkedését egy környezetben
- **racionális ágens** (*rational agent*): úgy cselekszik, hogy maximalizálja a teljesítménymérték várható értékét az addigi észlelési sorozat alapján
- **feladatkörnyezet** (*task environment*): leírása tartalmazza a **teljesítménymérték**, a **külső környezet**, a **beavatkozók** és a **szenzorok** meghatározását. Egy ágens tervezése során az első lépés mindig a feladatkörnyezet minél teljesebb leírása.
- feladatkörnyezet lehet:
  - teljesen vagy részben megfigyelhető
  - determinisztikus vagy sztochasztikus
  - epizódszerű vagy sorozatszerű
  - statikus vagy dinamikus
  - diszkrét vagy folytonos
  - egyágenses vagy többágenses
- **ágensprogram** (*agent program*): implementálja az ágensfüggvényt. Többféle alapvető ágensfelépítés van annak függvényében, hogy milyen információ jelenik meg és használható fel a döntési folyamatban. Ezen megközelítések különböznek: hatékonyságukban, kompaktságukban és rugalmasságukban. A megfelelő konstrukció kiválasztása a környezet természetétől függ.
- **egyszerű reflexszerű ágensek** (*simple reflex agent*): azonnal válaszolnak az észlelésekre
- **modellalapú reflexszerű ágensek** (*model-based reflex agent*): világ aktuális észleléseiből nem nyilvánvaló részeket belső állapotukban tartják nyilván (pl mi merre változhatott amit épp nem látnak)
- **célorientált ágensek** (*goal-based agent*): céljaik elére érdekében cselekszenek
- **hasznosságorientált ágensek** (*utility-based agent*): saját "boldogságuk" igyekeznek maximalizálni
- minden ágens javíthatja a teljesítményét **tanulás** (*learning*) segítségével