

# *NEM HUMÁN ÁGENSEK: A MORÁLIS STÁTUSZTÓL A POTENCIÁLIS JOGALANYISÁGIG – FOGALMI TÉRKÉPVÁZLATOK*

MÁRTON MIKLÓS

ELTE ÁJK JOG- ÉS TÁRSADALOMELMÉLETI TANSZÉK



PROGRAM  
FINANCED FROM  
THE NRDI FUND

# Problémafelvetés

- Más technológiáknál is előfordul: fizikai veszély, anyagi kár, adatokkal visszaélés, diszkrimináció, környezeti terhelés etc.
- A MI ágens a szó nem morális értelmében (Floridi – Sanders 2004, Dignum 2019)

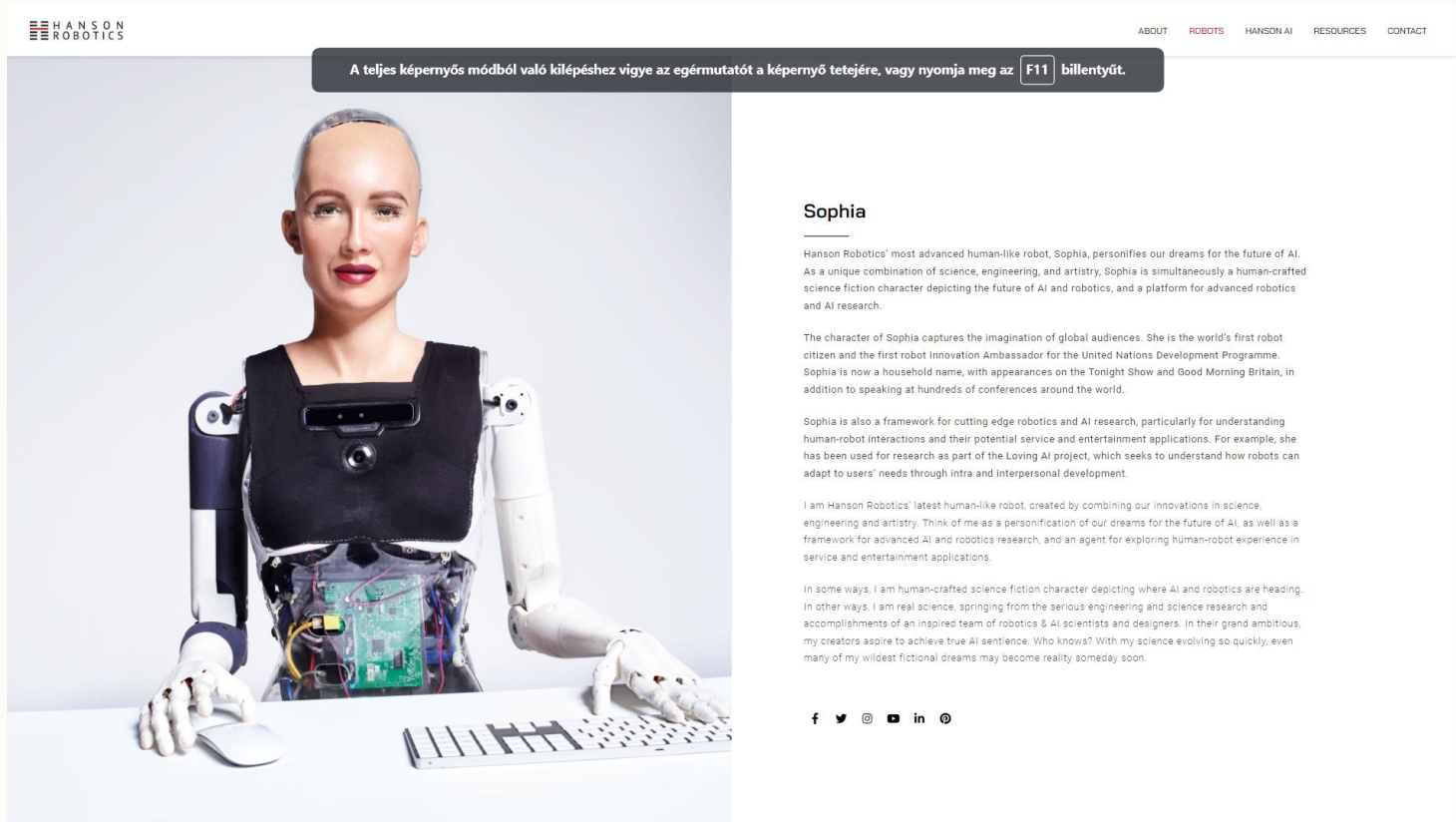
autonóm – explicit külső, környezeti hatás nélkül képes változtatni a belső állapotain, így képes változó környezetben is stabilan fönntartani céljait külső behatás nélkül;

adaptív – az interakciók hatására változnak az átalakítási szabályai, így akár explicit instrukciók nélkül képes mintázatokat azonosítani a környezetből származó adatokban, és ezekből tanulni;

interaktív – kölcsönösen képes más ágensekkel hatásokat gyakorolni egymásra, képes ezek szintén ágensként érzékelésére és ágensként történő azonosítására, illetve ennek megfelelő interakcióba is képes lépni velük.

# Problémafelvetés

- Nem humán ágencia → morális státusz, jogalanyiség? Lásd:



- A morális szimpátia körének kiterjedése:  
idegenek, rabszolgák, nők, állatok, növények  
... gépek?

# Fogalmi megkülönböztetések

1. Morális ágens vs. páciens. Felelősség/kötelezettség vs. jogok (mások kötelességei).

- A morális ágencia:

- a) nem ismerünk nem-humán ágenseket,

- b) ehhez valószínűleg tudatos és szabad cselekvésirányítási képesség kell, ez kérdéses...

- Ellenvélemény (Floridi – Sanders 2004: 17 skk): Értékelő, morális leírásokat olyan lények cselekedeteire is alkalmazunk, akiket nem tartunk morálisan felelősnek (gyerekek, mentőkutyák). Elszámoltathatóság vs. felelősség.

- A kérdés tehát a morális páciensi státusz: lehet-e MI rendszereknek jogaik?

2. Érdem vs. konzekvencialista elméletek

A páciensen tulajdonsága vs. várható haszon/kár.

3. Érdemalapú megközelítés ➡ közvetlen morális státusztulajdonítás.

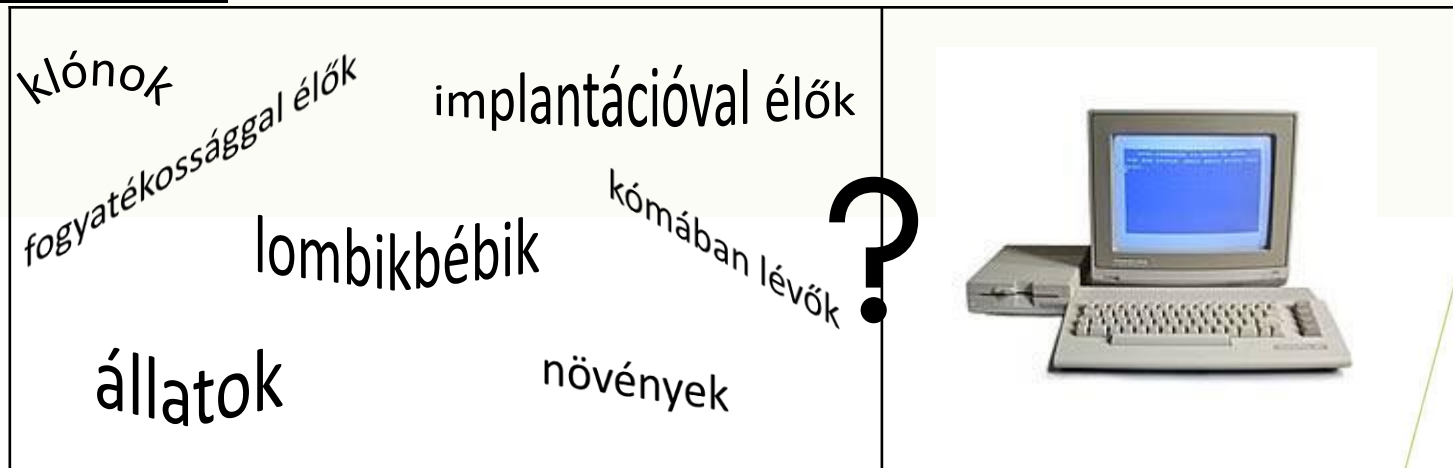
- a) intrinzikus (Floridi – Sanders 2004)

- b) relációs (Coeckelbergh 2014, 2018) tulajdonságok alapján

# Intrinzikus tulajdonságok

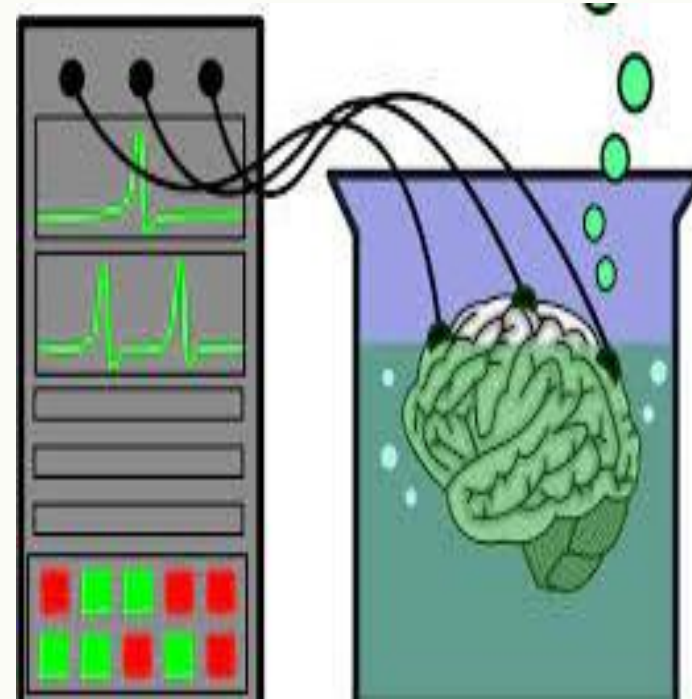
- Humán ágensek rendelkeznek olyan tulajdonságokkal, amelyek megalapozzák a morális státusz tulajdonítást:
  - ❑ fenomenális tudatosság (szenvedés – öröm képessége),
  - ❑ saját célok (valódi teleologikus viselkedés, intencionális állapotok),
  - ❑ szabad döntési képesség,
  - ❑ magasabb szintű kognitív képességek,
  - ❑ megfelelő eredet, vagy strukturális felépítés

## Problémák



# Intrinzikus tulajdonságok

2. Episztemikusan bizonytalan helyzetben vagyunk.



约翰·塞尔不懂中文

John Searle does not Understand Chinese



# Relációs tulajdonságok

- Egy nagyobb rendszerben játszott szerepe alapján tekintünk valamit morális páciensnek. Analógia: környezeti etika.

Érv: az emberek ugyanúgy „vakon” alkalmaznak a MI-re társadalmi szabályokat és elvárásokat, mint az embertársaikra → nem tudatos antropomorfizálás. Interdependens kapcsolati, érzelmi hálók alakulnak ki.



# Etikai behaviorizmus

- Két performatíve (durván) ekvivalens létezőnek azonos morális státuszúnak kell lennie. Az MI alapú robotok performatíve (durván) ekvivalensek velünk. (Danaher 2020)

Érv: tiszteletben tartja episztemikus korlátainkat. Bármilyen metafizikai tulajdonságot tekintünk is a morális státusz kritériumának, ezekre más lények esetében csak következtetni tudunk a viselkedésből.

- Módszertani behaviorizmus - a Turing teszt etikai megfelelője.



# Közvetett morális státusztulajdonítás

---

- Az alapján tulajdonítunk egy létezőnek valamilyen morális státuszt, hogy a tulajdonító rendelkezik morális státusszal, és erre hat, ahogyan bánunk a dologgal. Konzekvencialista megközelítés (Coeckelbergh 2021)
  - Ha valaki rosszul bánik egy MI-vel.
  - Ha valaki érzelmileg kötődik egy robothoz.
  - Ha a robot része egy közös humán-nemhumán együttműködésnek, és ezt érdemes fönntartani.



# *(Tentatív) konklúzió*

Tézis1: A morális státusz kérdése elméleti, normatív, ténykérdés.

Érv: a rabszolgák vagy az állatok esetében nem adományoztuk a morális státuszt nekik, hanem elismertük, hogy mindig is fennállt.

Következmények:

- 1) a szkepticizmuson alapuló érvek irrelevánsak;
- 2) az indirekt érvek irrelevánsak: ezeknél van humán kár, így nem igazán különböznek a rombolás, megbántás általános (károkozás nélküli) tilalmától;
- 3) olyan tulajdonság kéne, ami valamiféle fogalmi kapcsolatban áll a morális státusz iránti igénnyel;
- 4) Kontingens, megváltoztatható szociális tények irrelevánsak, lásd: tulajdonlás vs. rabszolga felszabadítás.

# *(Tentatív) konklúzió*

Tézis2: Az etikai státusból nem következik a jogi státusz. A jogalanyiség kérdése konzekvencialista alapon, praktikus kérdésként vetül föl.

Érv: mindig is tudták, hogy a rabszolgák vagy az állatok érző lények. A relációs tények is fennálltak (neveket adtak az állatoknak, stb.) Miért csak későn merült föl az emancipációjuk?

Tézis3: A jogalanyiség kérdése praktikus, prudenciális kérdés: alapvetően a morális és társadalmi kockázatokat kell figyelembe venni, ide értve a MI rendszerét is. Általában: inklúzió a kisebb kockázat, de a MI-nél nem biztos.

Probléma: feszültség lép föl, ha eltér a jogi és a morális státusz.

# Köszönöm a figyelmet!

Bostrom, N. & E. Yudkowsky 2014. The Ethics of Artificial Intelligence, in *The Cambridge Handbook of Artificial Intelligence*, Keith Frankish and William M. Ramsey (eds.), Cambridge: Cambridge University Press, 316–334.

Bryson, J. J. 2010. Robots should be slaves. *Close Engagements with Artificial Companions: Key social, psychological, ethical and design issues*, 8, 63-74.

Coeckelbergh, M. 2014. The moral standing of machines: Towards a relational and non- Cartesian moral hermeneutics. *Philosophy & Technology* 27, no. 1: 61–77.

Coeckelbergh, M. 2018. What Do We Mean by a Relational Ethics? Growing a Relational Approach to the Moral Standing of Plants, Robots and Other Non-Humans, in *Plant Ethics: Concepts and Applications*, Angela Kallhoff, Marcello Di Paola, and Maria Schörghener (eds.), London: Routledge, 98-109.

Coeckelbergh, M. 2021. Should We Treat Teddy Bear 2.0 as a Kantian Dog? Four Arguments for the Indirect Moral Standing of Personal Social Robots, with Implications for Thinking About Animals and Humans. *Minds and Machines* 31:337–360.

Danaher, J. 2020. Welcoming Robots into the Moral Circle: A Defence of Ethical Behaviourism, *Science and Engineering Ethics* 26:2023–2049.

Floridi, L. & J. W. Sanders. 2004. On the morality of artificial agents. *Minds and Machines*, 14.3: 349-379.

Héder M. 2020. *Mesterséges intelligencia*, Gondolat Kiadó, Budapest

Schwitzgebel, E., & Garza, M. 2015. A defense of the rights of artificial intelligences. *Midwest Studies in Philosophy*, 39(1): 89–119.

PROGRAM  
FINANCED FROM  
THE NRDI FUND