

A mesterséges intelligencia (MI) fejlődésének egyed társadalmi vonatkozásai

Turán György
MTA-SZTE Mesterséges Intelligencia Kutatócsoport

2018. december 14.

Áttekintés

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)
- ▶ a közeljövőben, sőt a jelenben aktuális kérdések

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)
- ▶ a közeljövőben, sőt a jelenben aktuális kérdések
 - ▶ MI hatása a munkaerőpiacra (nem tárgyaljuk)

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)
- ▶ a közeljövőben, sőt a jelenben aktuális kérdések
 - ▶ MI hatása a munkaerőpiacra (nem tárgyaljuk)
 - ▶ **autonóm, önálló döntésre képes** MI rendszerek

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)
- ▶ a közeljövőben, sőt a jelenben aktuális kérdések
 - ▶ MI hatása a munkaerőpiacra (nem tárgyaljuk)
 - ▶ **autonóm, önálló döntésre képes** MI rendszerek
 - ▶ új kérdéskör a technológiai fejlődés korábbi kérdéseihez képest (könyvnyomtatás, gőzgép, atombomba, biotechnika)

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)
- ▶ a közeljövőben, sőt a jelenben aktuális kérdések
 - ▶ MI hatása a munkaerőpiacra (nem tárgyaljuk)
 - ▶ **autonóm, önálló döntésre képes** MI rendszerek
 - ▶ új kérdéskör a technológiai fejlődés korábbi kérdéseihez képest (könyvnyomtatás, gőzgép, atombomba, biotechnika)
 - ▶ jogi, etikai szabályozás

Áttekintés

- ▶ sikerek: sakk, Jeopardy (IBM Watson), Go
- ▶ távolabbi jövőre vonatkozó kérdések
 - ▶ lehet-e a számítógépnek tudata?, mesterséges szuperintelligencia (szingularitás)
- ▶ a közeljövőben, sőt a jelenben aktuális kérdések
 - ▶ MI hatása a munkaerőpiacra (nem tárgyaljuk)
 - ▶ **autonóm, önálló döntésre képes** MI rendszerek
 - ▶ új kérdéskör a technológiai fejlődés korábbi kérdéseihez képest (könyvnyomtatás, gőzgép, atombomba, biotechnika)
 - ▶ jogi, etikai szabályozás
- ▶ az MI közösség közéleti, tudományos válaszadási kísérletei

Önvezető autó

Önvezető autó

- ▶ halálos baleset: 2018. március 18, Uber, Tempe, Arizona, felügyelő ült az autóban, de nem figyelt. A cég az áldozat családjával 10 nap alatt megállapodott a rendezést illetően.

Önvezető autó

- ▶ halálos baleset: 2018. március 18, Uber, Tempe, Arizona, felügyelő ült az autóban, de nem figyelt. A cég az áldozat családjával 10 nap alatt megállapodott a rendezést illetően.
- ▶ engedélyezés: 2018. október 30, Waymo, Kalifornia egy körzetében, először emberi felügyelet nélkül. Felügyelettel 2009 óta 10 millió mérföldet mentek.

Önvezető autó

- ▶ halálos baleset: 2018. március 18, Uber, Tempe, Arizona, felügyelő ült az autóban, de nem figyelt. A cég az áldozat családjával 10 nap alatt megállapodott a rendezést illetően.
- ▶ engedélyezés: 2018. október 30, Waymo, Kalifornia egy körzetében, először emberi felügyelet nélkül. Felügyelettel 2009 óta 10 millió mérföldet mentek.
- ▶ verseny

Önvezető autó

- ▶ halálos baleset: 2018. március 18, Uber, Tempe, Arizona, felügyelő ült az autóban, de nem figyelt. A cég az áldozat családjával 10 nap alatt megállapodott a rendezést illetően.
- ▶ engedélyezés: 2018. október 30, Waymo, Kalifornia egy körzetében, először emberi felügyelet nélkül. Felügyelettel 2009 óta 10 millió mérföldet mentek.
- ▶ verseny
 - ▶ cégek között: Kaliforniában 60 cégnek van teszt engedélye

Önvezető autó

- ▶ halálos baleset: 2018. március 18, Uber, Tempe, Arizona, felügyelő ült az autóban, de nem figyelt. A cég az áldozat családjával 10 nap alatt megállapodott a rendezést illetően.
- ▶ engedélyezés: 2018. október 30, Waymo, Kalifornia egy körzetében, először emberi felügyelet nélkül. Felügyelettel 2009 óta 10 millió mérföldet mentek.
- ▶ verseny
 - ▶ cégek között: Kaliforniában 60 cégnek van teszt engedélye
 - ▶ államok között: engedékenység a cégek odacsábítása céljából

Önvezető autó

- ▶ halálos baleset: 2018. március 18, Uber, Tempe, Arizona, felügyelő ült az autóban, de nem figyelt. A cég az áldozat családjával 10 nap alatt megállapodott a rendezést illetően.
- ▶ engedélyezés: 2018. október 30, Waymo, Kalifornia egy körzetében, először emberi felügyelet nélkül. Felügyelettel 2009 óta 10 millió mérföldet mentek.
- ▶ verseny
 - ▶ cégek között: Kaliforniában 60 cégnek van teszt engedélye
 - ▶ államok között: engedékenység a cégek odacsábítása céljából
- ▶ szövetségi vagy államonkénti szabályozás?

Autonóm fegyver: SGR-AI

Autonóm fegyver: SGR-AI

- ▶ autonóm géppuska a két Korea közti demilitarizált zónában, kb. 10 éve (ld. Youtube)

Autonóm fegyver: SGR-AI

- ▶ autonóm géppuska a két Korea közti demilitarizált zónában, kb. 10 éve (ld. Youtube)
- ▶ döntés: riasztás, gumi vagy éles lövedék

Autonóm fegyver: SGR-AI

- ▶ autonóm géppuska a két Korea közti demilitarizált zónában, kb. 10 éve (ld. Youtube)
- ▶ döntés: riasztás, gumi vagy éles lövedék
- ▶ LAWS (lethal autonomous weapons systems)

Autonóm fegyver: SGR-AI

- ▶ autonóm géppuska a két Korea közti demilitarizált zónában, kb. 10 éve (ld. Youtube)
- ▶ döntés: riasztás, gumi vagy éles lövedék
- ▶ LAWS (lethal autonomous weapons systems)
- ▶ HITL (human-in-the-loop) vagy HOTL (human-on-the-loop)?

Autonóm fegyver: SGR-AI

- ▶ autonóm géppuska a két Korea közti demilitarizált zónában, kb. 10 éve (ld. Youtube)
- ▶ döntés: riasztás, gumi vagy éles lövedék
- ▶ LAWS (lethal autonomous weapons systems)
- ▶ HITL (human-in-the-loop) vagy HOTL (human-on-the-loop)?
- ▶ 2018. március: 50 vezető MI kutató hirdette meg a dél-koreai KAIST egyetem bojkottját a Hanwha céggel tervezett együttműködés miatt

Autonóm fegyver: SGR-AI

- ▶ autonóm géppuska a két Korea közti demilitarizált zónában, kb. 10 éve (ld. Youtube)
- ▶ döntés: riasztás, gumi vagy éles lövedék
- ▶ LAWS (lethal autonomous weapons systems)
- ▶ HITL (human-in-the-loop) vagy HOTL (human-on-the-loop)?
- ▶ 2018. március: 50 vezető MI kutató hirdette meg a dél-koreai KAIST egyetem bojkottját a Hanwha céggel tervezett együttműködés miatt
- ▶ az egyetem elnökének gyors válasza után (elállnak a LAWS fejlesztésben való részvételtől) a bojkott megszűnt

Idősgondozás: Shin-tomi, Care-O-Bot

Idősgondozás: Shin-tomi, Care-O-Bot

- ▶ Shin-tomi: tokiói időotthon, 20 féle robot, pl. “tornatanár”

Idősgondozás: Shin-tomi, Care-O-Bot

- ▶ Shin-tomi: tokiói idősothton, 20 féle robot, pl. “tornatanár”
- ▶ más hozzáállás: Keleten (Japán): Astro Boy, Nyugaton (Európa, USA): Gólem, Frankenstein

Idősgondozás: Shin-tomi, Care-O-Bot

- ▶ Shin-tomi: tokiói idősothton, 20 féle robot, pl. “tornatanár”
- ▶ más hozzáállás: Keleten (Japán): Astro Boy, Nyugaton (Európa, USA): Gólem, Frankenstein
- ▶ Care-O-Bot: bolti asszisztens, idősgondozás (projekt?), pl. tudja, hogy virághoz vázát kell hozni

Idősgondozás: Shin-tomi, Care-O-Bot

- ▶ Shin-tomi: tokiói idősothton, 20 féle robot, pl. “tornatanár”
- ▶ más hozzáállás: Keleten (Japán): Astro Boy, Nyugaton (Európa, USA): Gólem, Frankenstein
- ▶ Care-O-Bot: bolti asszisztens, idősgondozás (projekt?), pl. tudja, hogy virághoz vázát kell hozni
- ▶ Draper, Sorell: *Ethical values and social care robots for older people: and international qualitative study* (Ethics and Information Technology, 2018)

Idősgondozás: Shin-tomi, Care-O-Bot

- ▶ Shin-tomi: tokiói idősothton, 20 féle robot, pl. “tornatanár”
- ▶ más hozzáállás: Keleten (Japán): Astro Boy, Nyugaton (Európa, USA): Gólem, Frankenstein
- ▶ Care-O-Bot: bolti asszisztens, idősgondozás (projekt?), pl. tudja, hogy virághoz vázát kell hozni
- ▶ Draper, Sorell: *Ethical values and social care robots for older people: and international qualitative study* (Ethics and Information Technology, 2018)
- ▶ szempontok: a gondozott önállósága, biztonsága, személyiségi jogai, kommunikáció formái, meggyőzés

Idősgondozás: Shin-tomi, Care-O-Bot

- ▶ Shin-tomi: tokiói idősothton, 20 féle robot, pl. “tornatanár”
- ▶ más hozzáállás: Keleten (Japán): Astro Boy, Nyugaton (Európa, USA): Gólem, Frankenstein
- ▶ Care-O-Bot: bolti asszisztens, idősgondozás (projekt?), pl. tudja, hogy virághoz vázát kell hozni
- ▶ Draper, Sorell: *Ethical values and social care robots for older people: and international qualitative study* (Ethics and Information Technology, 2018)
- ▶ szempontok: a gondozott önállósága, biztonsága, személyiségi jogai, kommunikáció formái, meggyőzés
- ▶ mit csináljon a robot ha a gondozott nem akarja bevenni a gyógyszert?

Robot etika

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**
- ▶ etikai elméletek (utilitarizmus, Kant) - **vagy valami más?**

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**
- ▶ etikai elméletek (utilitarizmus, Kant) - **vagy valami más?**
- ▶ Rawls: *Outline of a **decision procedure** for ethics* (The Philosophical Review, 1951), *A Theory of Justice*, 1971

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**
- ▶ etikai elméletek (utilitarizmus, Kant) - **vagy valami más?**
- ▶ Rawls: *Outline of a **decision procedure** for ethics* (The Philosophical Review, 1951), *A Theory of Justice*, 1971
- ▶ Leben: *A Rawlsian algorithm for autonomous vehicles* (Ethics and Information Technology), 2017

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**
- ▶ etikai elméletek (utilitarizmus, Kant) - **vagy valami más?**
- ▶ Rawls: *Outline of a **decision procedure** for ethics* (The Philosophical Review, 1951), *A Theory of Justice*, 1971
- ▶ Leben: *A Rawlsian algorithm for autonomous vehicles* (Ethics and Information Technology), 2017
- ▶ “villamos dilemmák”, “fairness” elmélet

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**
- ▶ etikai elméletek (utilitarizmus, Kant) - **vagy valami más?**
- ▶ Rawls: *Outline of a **decision procedure** for ethics* (The Philosophical Review, 1951), *A Theory of Justice*, 1971
- ▶ Leben: *A Rawlsian algorithm for autonomous vehicles* (Ethics and Information Technology), 2017
- ▶ “villamos dilemmák”, “fairness” elmélet
- ▶ ha a fék elromlott, és választani kell egy bukósisakos vagy egy bukósisak nélküli biciklista elütése között, mit válasszunk?

Robot etika

- ▶ cél: autonóm robotok etikus viselkedése - **mit jelent?**
- ▶ etikai elméletek (utilitarizmus, Kant) - **vagy valami más?**
- ▶ Rawls: *Outline of a **decision procedure** for ethics* (The Philosophical Review, 1951), *A Theory of Justice*, 1971
- ▶ Leben: *A Rawlsian algorithm for autonomous vehicles* (Ethics and Information Technology), 2017
- ▶ “villamos dilemmák”, “fairness” elmélet
- ▶ ha a fék elromlott, és választani kell egy bukósisakos vagy egy bukósisak nélküli biciklista elütése között, mit válasszunk?
- ▶ a bukósisakosnak nagyobb a túlélési esélye; de járjon rosszabbul azért mert betartotta a szabályt?

Két megoldatlan probléma

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá
 - ▶ “a váza nem fér a bőröndbe, mert túl nagy” - mi túl nagy, a váza vagy a bőrönd?

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá
 - ▶ “a váza nem fér a bőröndbe, mert túl nagy” - mi túl nagy, a váza vagy a bőrönd?
 - ▶ Winograd teszt: számítógépek egyelőre nem tudják megoldani

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá
 - ▶ “a váza nem fér a bőröndbe, mert túl nagy” - mi túl nagy, a váza vagy a bőrönd?
 - ▶ Winograd teszt: számítógépek egyelőre nem tudják megoldani
- ▶ tanult algoritmusok érthetősége, megmagyarázhatósága (ne legyen fekete doboz)

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá
 - ▶ “a váza nem fér a bőröndbe, mert túl nagy” - mi túl nagy, a váza vagy a bőrönd?
 - ▶ Winograd teszt: számítógépek egyelőre nem tudják megoldani
- ▶ tanult algoritmusok érthetősége, megmagyarázhatósága (ne legyen fekete doboz)
 - ▶ EU Általános Adatvédelmi Rendelet (GDPR)

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá
 - ▶ “a váza nem fér a bőröndbe, mert túl nagy” - mi túl nagy, a váza vagy a bőrönd?
 - ▶ Winograd teszt: számítógépek egyelőre nem tudják megoldani
- ▶ tanult algoritmusok érthetősége, megmagyarázhatósága (ne legyen fekete doboz)
 - ▶ EU Általános Adatvédelmi Rendelet (GDPR)
 - ▶ “az algoritmus elutasította hitelfelvételi kérelmemet, miért?”

Két megoldatlan probléma

- ▶ hétköznapi tudás és gondolkodás: mindenki erőfeszítés nélkül képes rá
 - ▶ “a váza nem fér a bőröndbe, mert túl nagy” - mi túl nagy, a váza vagy a bőrönd?
 - ▶ Winograd teszt: számítógépek egyelőre nem tudják megoldani
- ▶ tanult algoritmusok érthetősége, megmagyarázhatósága (ne legyen fekete doboz)
 - ▶ EU Általános Adatvédelmi Rendelet (GDPR)
 - ▶ “az algoritmus elutasította hitelfelvételi kérelmemet, miért?”
- ▶ *A MI matematikai alapjai* (NKP 2018-2021), Rényi, ELTE, SZTAKI, Pázmány, SZTE

Szervezetek, források

Szervezetek, források

- ▶ One Hundred Year Study on Artificial Intelligence: *Artificial Intelligence and Life in 2030* (Stanford University, 2016)

Szervezetek, források

- ▶ One Hundred Year Study on Artificial Intelligence: *Artificial Intelligence and Life in 2030* (Stanford University, 2016)
- ▶ IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2; végső változat: 2019

Szervezetek, források

- ▶ One Hundred Year Study on Artificial Intelligence: *Artificial Intelligence and Life in 2030* (Stanford University, 2016)
- ▶ IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2; végső változat: 2019
- ▶ Future of Life Institute (Hawking, Musk, Stuart Russell, ...)

Szervezetek, források

- ▶ One Hundred Year Study on Artificial Intelligence: *Artificial Intelligence and Life in 2030* (Stanford University, 2016)
- ▶ IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2; végső változat: 2019
- ▶ Future of Life Institute (Hawking, Musk, Stuart Russell, ...)
 - ▶ Boddington: *Towards a Code of Ethics for Artificial Intelligence*, 2017

Szervezetek, források

- ▶ One Hundred Year Study on Artificial Intelligence: *Artificial Intelligence and Life in 2030* (Stanford University, 2016)
- ▶ IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2; végső változat: 2019
- ▶ Future of Life Institute (Hawking, Musk, Stuart Russell, ...)
 - ▶ Boddington: *Towards a Code of Ethics for Artificial Intelligence*, 2017
- ▶ Wendell, Allen: *Moral Machines: Teaching Robots Right from Wrong*, 2009

Szervezetek, források

- ▶ One Hundred Year Study on Artificial Intelligence: *Artificial Intelligence and Life in 2030* (Stanford University, 2016)
- ▶ IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, Version 2; végső változat: 2019
- ▶ Future of Life Institute (Hawking, Musk, Stuart Russell, ...)
 - ▶ Boddington: *Towards a Code of Ethics for Artificial Intelligence*, 2017
- ▶ Wendell, Allen: *Moral Machines: Teaching Robots Right from Wrong*, 2009
- ▶ Lin, Jenkins, Abney: *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, 2017