KAKO LAHKO UMETNA INTELIGENCA PREVZAME OBLAST?

Ivan Bratko

Fakulteta za računalništvo in informatiko, UL

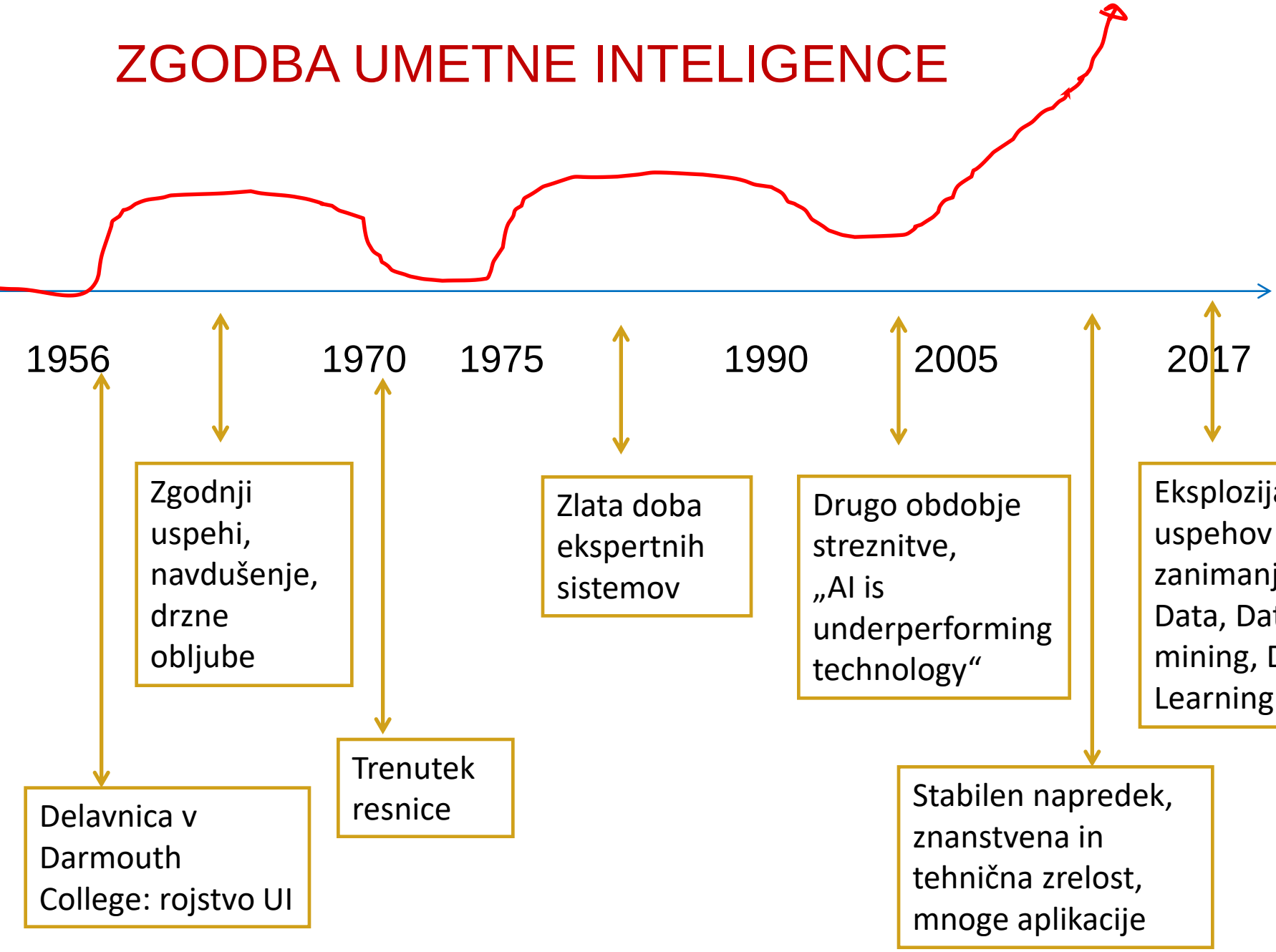
Dnevi slovenske informatike, 2017

KRATKA ZGODBA UMETNE INTELIGENCE

Bilo je nekaj vzponov in padcev, preden je področje dozorelo

"THE CHANGING WORLD OF ARTIFICIAL INTELLIGENCE", Bratko 2013

ZGODBA UMETNE INTELIGENCE



Primeri tipičnih problemov za reševanje z umetno inteligenco,

od začetka do zdaj

Artificial Intelligence 1960,

Carnegie-Mellon Uni., Pittsburgh: Simon and Newell

Kriptogramske uganke

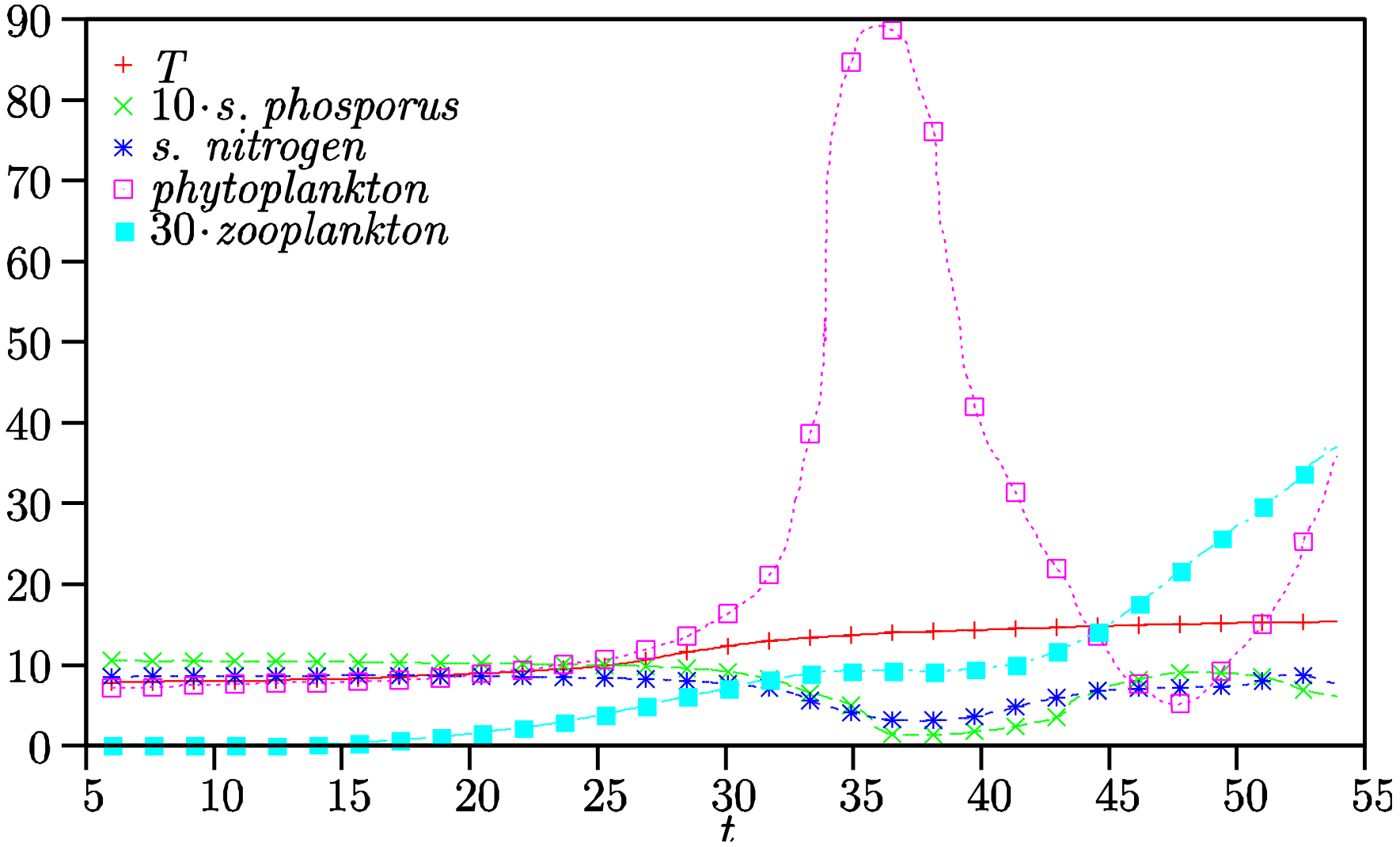
```
   D O N A L D
+  G E R A L D
=  R O B E R T
```

Kako ljudje rešujejo take probleme?

S kakšnimi algoritmi lahko simuliramo človekovo reševanje?

Simon: Nobelova nagrada

Okrog 1990, Dansko jezero Glumso



Računalnik sam odkrije zakon rasti alg (Kopenhagen; Padova; Ljubljana, FRI)

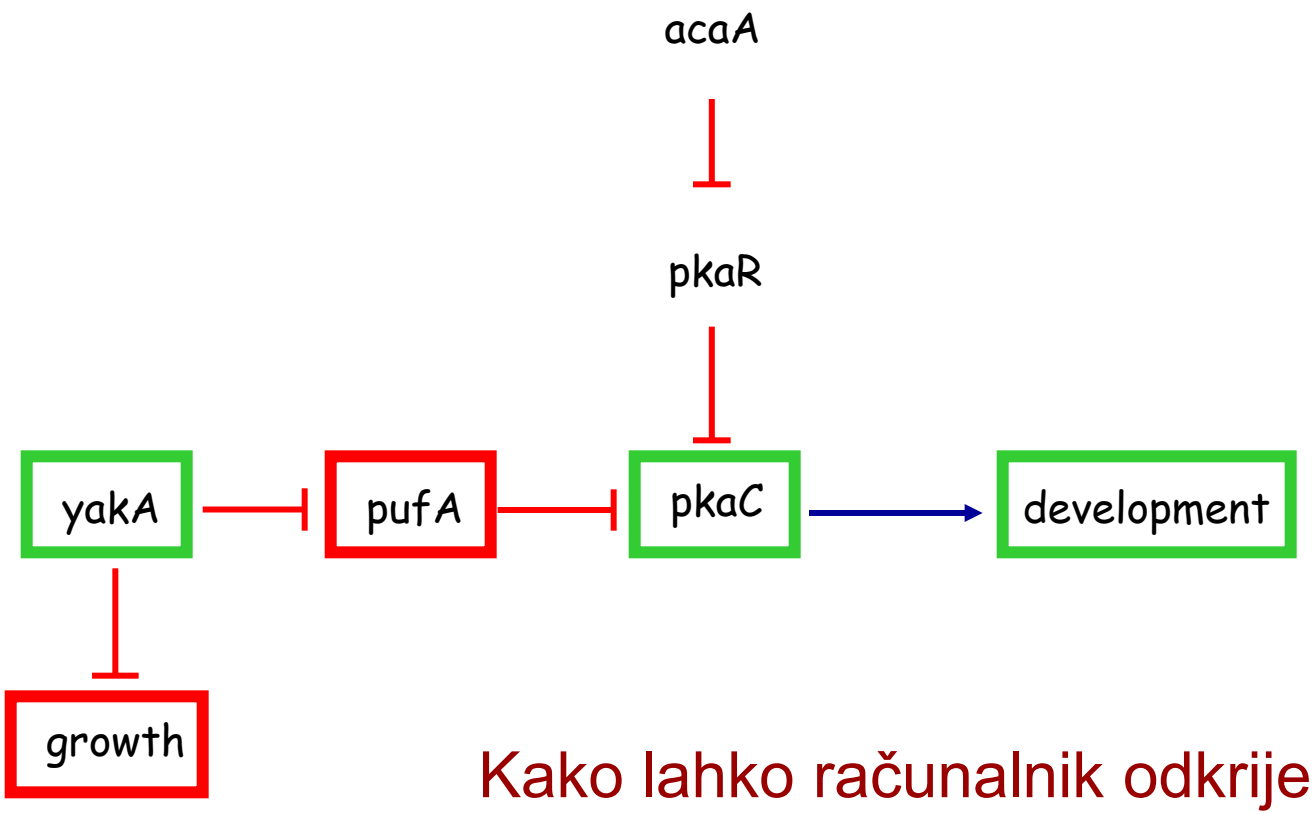
2000s, Ameba Dictyostelium

Ko zmanjka hrane, se amebe združijo v kopico.

Tako dobijo sposobnost premikanja.

Računalnik odkrije teorije o tem, kako geni vplivajo med seboj in na funkcije organizma (Houston, Ljubljana)

Možna genetska teorija, ki razloži eksperimentalne podatke



Kako lahko računalnik odkrije tako teorijo? Za vsega 7 elementov obstaja kar ~***1.000.000.000.000.000*** možnih mrež!

DISKUSIJA IZ 2013
Z: S. Russell in S. Muggleton

- Doslej je bila za nas umetna inteligenca izredno privlačno raziskovalno področje – „a lot of fun“.

  Veliko idej, nepričakovanih aplikacij ...

  Vendar redko z res veliko težo odgovornosti in usodnosti.

- Toda zdaj aplikativni pomen u.i. ni več samo „a lot of fun“

- Primeri:
  - Inteligentna, avtonomna vozila (Google car, Sebastian Thrun)
  - Inteligentna orožja, brezpilota letala: etični problemi, zakonodajni problemi
  - Priporočilni sistemi, krojenje mnenj po družbenih omrežjih

ALI LAHKO UMETNA INTELIGENCA PREVZAME OBLAST?

- Do pred kratkim je bil odgovor jasen: Ne!
- Odgovor zdaj se zapleta: Vse bolj postaja to mogoče

Spreminja se celo definicija

umetne inteligence

TRADICIONALNA DEFINICIJA UMETNE INTELIGENCE JE ZAPLETENA

- Ena od zgodnjih definicij, ~1980, Winston:

  Umetna inteligenca preučuje ideje,
  ki omogočajo računalnikom,
  da izvajajo naloge, ki jih pripisujemo človeški inteligenci

- Glavna težava: Kaj je *inteligenca*?
  Vse to je treba precizirati .....

DEFINICIJA UI DANES

- Definicija danes na splošno ne izgleda več problem
- Velika medijska pokritost
- Vsak si zato že predstavlja nekaj po svoje, večinoma:

  „Umetna inteligenca je tista tehnologija, ki lahko naredi vse“
- Točka tehnološke singularnosti – „superinteligenca“
- Vedno bolj pogosto »AI will kill us all«

Google »AI will kill us all« (april 2017)

Nine times tech leaders warned us that robots will kill us all | Techworld

www.techworld.com › Galleries › Personal Tech Galleries

Oct 20, 2016 - Nine times tech leaders warned us that robots *will kill us all*: ... that "the rise of powerful *AI will* be either the best, or the worst thing" for humanity.

Why AI will probably kill us all. - YouTube

▸ 22:02

https://www.youtube.com/watch?v=SPAmbUZ9UKk

Mar 5, 2017 - Uploaded by Boyinaband

When you look into it, Artificial Intelligence is absolutely terrifying. Really hope we don't die. ▹ ▹ If you want to ...

Will A.I. kill us all? Maybe not . . . - YouTube

▸ 4:29

https://www.youtube.com/watch?v=G3SXUir1jlw

Sep 5, 2016 - Uploaded by John Michael Godier

An exploration into artificial intelligence and its potential to destroy the human race in which I cover possibilities ...

Stephen Hawking: Artificial intelligence could wipe out humanity when ...

www.independent.co.uk/.../stephen-hawking-artificial-intelligence-could-wipe-out-hu...

Oct 8, 2015 - Such computers could become so competent that they *kill us* by accident, ... “A super intelligent *AI will* be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, we're in trouble. ... + show *all* ...

POVZETEK POMISLEKOV PRED U.I.

- Tradicionalna bojazen pred umetno inteligenco

  Ta bojazen je tehnično in racionalno šibko utemeljena

- Točka „tehnološke singularnosti“

  Velika medijska pozornost, vendar zelo vprašljivo

- Preobrat v zadnjih letih: Bistveno nove tehnične možnosti, novi problemi; odprto pismo znanstvenikov UI, januar 2015; Asilomar principles 2017

- „Scenarij Watson“, ko računalnik de facto prevzame oblast; konec demokracije; tehnični elementi UI to omogočajo

TRADICIONALNA BOJAZEN PRED UI

- Vprašanje iz nestrokovnih krogov od začetka UI:

  Ali ni nevarno, kar delate? Ali se ne bojite, da UI prevzame oblast?

  Ne, kako pa naj bi jo prevzela?

  No, roboti se zarotijo in z vojaško silo zavladajo ...?

  Toda, kako naj bi robotom to uspelo? Preveč očitno, da ljudje ne bi posegli vmes

- Z leti je postajalo to vprašanje vse manj zanimivo

„TECHNOLOGICAL SINGULARITY“

- Točka singularnosti: trenutek, ko bo rač. inteligenca presegla človeško (Kurzweil).
- Kdaj se bo to zgodilo? Razne napovedi, med 2035 in 2070 ...
- Vendar: Vse to je videti zelo špekulativno in slabo definirano.
- Kako je sploh definiran trenutek, ko bo umetna inteligenca presegla človeško?

TOČKA SINGULARNOSTI – KDAJ?

- Inteligenco sestavlja vrsta zmožnosti: pomnjenje, računanje, sklepanje, učenje, naravni jezik, inteligentno komuniciranje z drugimi inteligentnimi „agenti“, razlaganje, argumentiranje, poučevanje, ...

- V nekaterih od teh sposobnosti je UI že zdavnaj presegla človeka (pomnjenje, računanje, ...), nekatere so videti še zelo daleč (naravni jezik, poučevanje, komentiranje, smisel za humor, ...)

- Kako bodo ljudje sploh zaznali, kdaj je napočil trenutek singularnosti?

TOČKA SINGULARNOSTI - POSLEDICE

- Helbing (2017): “ Technology visionaries ... are warning that super-intelligence is a serious danger for humanity, possibly even more dangerous than nuclear weapons.”

- Elon Musk, ekspert za UI in podjetnik: „AI is our biggest existential threat".

- Stephen Hawking, fizik:

  " Machines with AI could spell the end of the human race".

  „The creation of powerful artificial intelligence will be either the best, or the worst thing, ever to happen to humanity”

TOČKA SINGULARNOSTI – POSLEDICE?

- Vendar tej diskusiji manjka konkretnosti:

  V kakšnem smislu bo to konec?

  Kaj točno naj bi se dejansko zgodilo?

  Zakaj naj bi se to zgodilo? Ker bodo ljudje tako razočarani nad sabo? In bodo od tega kar izumrli? Ali ...

Hipoteza o točki singularnosti ni videti posebej pomembna

Toda : obstajajo druga vprašanja

KAJ MENIJO SAMI ZNANSTVENIKI V UI?

- Tu je vprašanje o problematičnih vidikih UI postalo v zadnjih letih nenadoma zelo aktualno

- Odprto pismo nekaterih vodilnih znanstvenikov UI o nadaljnjem razvoju UI in predvsem o uporabah UI, kjer prežijo tudi nevarnosti (2015; 8000 podpisnikov)

- Zakonodaja sopiha daleč zadaj za realnimi možnostmi in realnimi problemi

KAJ SE JE ZGODILO V UI V ZADNJIH LETIH?

- Kaj se je takega zgodilo v UI v zadnjih letih? Nekateri tehnični dosežki UI ...

- Pa tudi, najpomembneje: splet, ki zdaj skoraj neomejeno in nekontrolirano omogoča stvari, ki pred tem niso bile možne

- Zelo vprašljivo: praktično nekontrolirano, brezobzirno zbiranje osebnih podatkov v podatkovnih bazah (internet, pametne mobilne naprave, ...)

ODPRTO PISMO: NEKATERE PRIORITETE GLEDE DRUŽBENIH VIDIKOV UI

- Zakonodaja in etika
  - Samovozeči avtomobili, avtonomna letala (droni)
  - Strojna etika
  - Avtonomna orožja: „Can lethal autonomous weapons be made to comply with humanitarian law?“
  - Zasebnost: „Our ability to take full advantage of the synergy between AI and big data will depend in part on our ability to manage and preserve privacy.“
  - Profesionalna etika raziskovalcev UI

ODPRTO PISMO, VTIS

- Pismo vsebuje **pravo sporočilo**

- Vendar je napisano **zelo previdno!**

ČESA ODPRTO PISMO NE OMENJA, VSAJ NE EKSPLICITNO?

INTERNET IN UI
KOT ORODJE MANIPULACIJE –

KONEC DEMOKRACIJE?

SCENARIJ WATSON

- **G. Williams** pride na idejo, **povečati obseg bio hrane. U**stanovi majhno skupino, zametek **socialnega omrežja** na spletu, imenovanega **BioNet**.

- Od tu naprej Williams uporablja inteligentni **program Watson.** Watson sledi diskusijam na spletu o tej temi.

WATSON

- 1 Pozna vse osebne podatke o ljudeh na BioNetu, pa tudi praktično vseh v državi, tudi bolezni, probleme, močne in šibke lastnosti

- 2 Pozna vlogo ljudi v omrežjih, priljubljenost, vpliv

- 3 Pozna okus in preference vseh teh ljudi. Zna napovedovati, kako bi bil vsakemu posamezniku všeč npr. izbrani politik ali potencialni politik

WATSON, nad.

- 4 Watson pozna metode, kako učinkovito spreminjati mnenje v omrežju.

- 5 Watson sledi diskusiji znotraj BioNeta: vpliva na mnenja s selektivnim prepošiljanjem izvlečkov diskusij.

- 6 Watson tudi razume, da bi se dalo cilj hitreje uresničiti s političnimi sredstvi.

  Zato Williams ustanovi stranko.

WATSON, nad.

- 6 Watson določi, kateri voditelji v stranki bi imeli največ možnosti na splošnih volitvah.

  Ustrezno zmanipulira volitve vodstva stranke.

- 7 V volilni kampanji pošlje vsakemu volivcu njemu prilagojeno vabilo.

  Upošteva posameznikov okus, in njemu osebno prilagodi volilne obljube

- 8 Stranka zmaga na volitvah.

KDO JE DOBIL VOLITVE?

- Kdo je pravzaprav zmagal? Williams ali Watson, človek ali računalnik?

- Originalni cilji o bio prehrani so zdaj v ozadju. Watson se zdaj ukvarja z drugimi, bolj nujnimi nalogami ...

- Tak scenarij: *Konec demokracije*

UI ZA PREVZEM OBLASTI NE POTREBUJE ROBOTOV ...

- Za dejanski prevzem oblasti roboti niso niti dovolj spretni niti niso potrebni

- Za izvedbo prevzem oblasti računalnik potrebuje le ljudi, ki jih bo manipuliral prek interneta.

- Watson zna Izbirati vodljive posameznike, ki imajo vpliv v socialnih omrežij in ki so izvoljivi: izgled, nastop, ...

SCENARIJ WATSON JE IZMIŠLJEN
TODA TEHNIČNI ELEMENTI ŽE OBSTAJAJO

- Sistematično in stalno zbiranje podatkov na spletu, vključno z osebnimi in občutljivimi podatki

- Metode strojnega učenja in podatkovnega rudarjenja (data mining, big data, globoke nevronske mreže)

- Pomislek: Globoke nevronske mreže so fantastično uspešne, vendar nihče ne ve, kako in zakaj? Kaj gre lahko narobe? Je lahko zelo tvegano ...

- Metode avtomatskega priporočanja; računalnik pozna okus vsakega posameznika; ne le o filmih, temveč tudi o političnih opcijah

TEHNIČNI ELEMENTI SCENARIJA WATSON

- Tehnike za računalniška družbena omrežja:
  - vplivne točke omrežja in
  - računalniško vodeno spreminjanje mnenj

- Razpoznavanje uporabnikovih odzivov:
  - ne samo „like“,
  - tudi čustev, tudi namerno prikritih, npr. s kamero; (Li et al. 2017) - je lahko etično zelo vprašljivo

- Tehnike avtomatskega sklepanja, planiranja**,** iger in odločanja, spodbujevano učenje

SCENARIJ WATSON: FAKTORJI TVEGANJA

- Ogromen poslovni interes podjetij, ki s to tehnologijo in vohunjenjem dosegajo veliko prednost
- Razmeroma majhno zanimanje in zavedanje javnosti, (med)državnih institucij in zakonodaje za učinkovit nadzor nad možnimi zlorabami te tehnologije
- Slabo splošno poznavanje te tehnologije in njenih zmožnosti
- Udobno, pasivno zadovoljstvo ob prednostih, ki jih nudi internet; ob tem se večina ljudi ne sprašuje ...

KAJ JE TREBA NAREDITI?

- Omejiti računalniško zbiranje osebnih podatkov
- Vzpostaviti učinkovite mehanizmem nadzora
- Zakonodaja:
  - Bo zelo zapleteno, dolgotrajno, veliko odporov
  - Kdo ima motivacijo za to? Prevladuje nezainteresiranost
  - Tudi če predpisi obstajajo, se jih v praksi izognejo
  - „For improved user experience“! Če ne daš soglasja za uporabo osebnih podatkov, se socialno in tehnično izključiš

- Russell takrat, 2013, o scenariju Watson:

  „End of democracy?

  Yes, I belive it is possible.

  You should write a book!“

Oktober 2016

Scientific American, Feb. 2017

Will Democracy Survive Big Data and Artificial Intelligence?

Dirk Helbing, Bruno S. Frey, Gerd Gigerenzer, Ernst Hafen, Michael Hagner, Yvonne Hofstetter, Jeroen van den Hoven, Roberto V. Zicari, Andrej Zwitter

Helbing:

“In the future, using sophisticated manipulation technologies, these [software] platforms will be able to steer us through entire courses of action ... from which corporations earn billions. *The trend goes from programming computers to programming people.*”

„ ... sedanje razširjeno zbiranje in procesiranje osebnih podatkov

gotovo ni v skladu z zakonodajo o zaščiti podatkov v Evropi in drugod.“

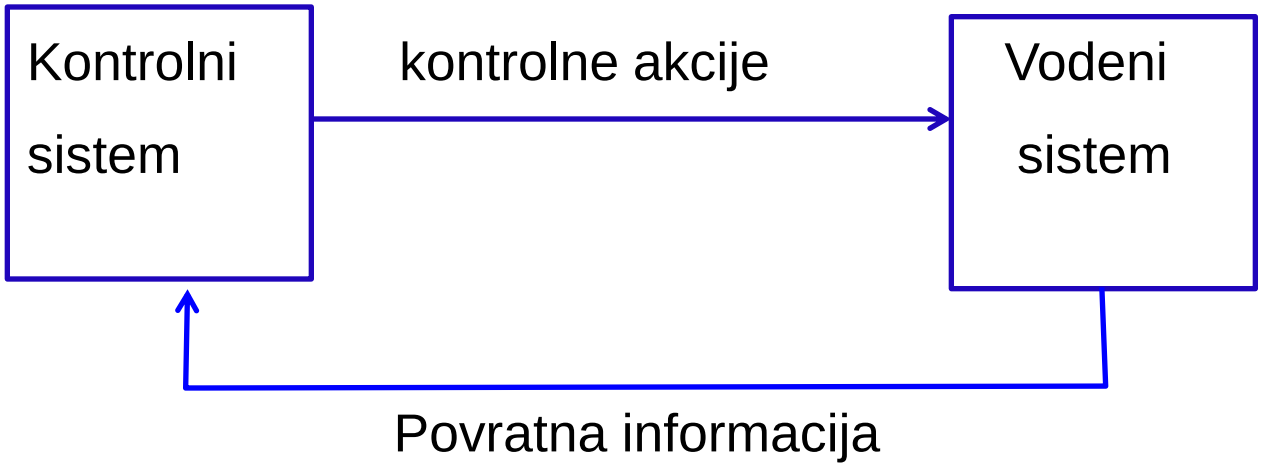
“ ... a single click to confirm that we agree with ... a hundred-page “terms of use” agreement (which is the case these days for many information platforms) is woefully inadequate.”

Helbing: „Osnovne človekove pravice bi morale biti zaščitene, čeprav smo v času digitalne revolucije ... Država bi morala poskrbeti za zakonski okvir, ki bi zagotovil, da se tehnologija razvija tako, da ne ogroža demokracije.“

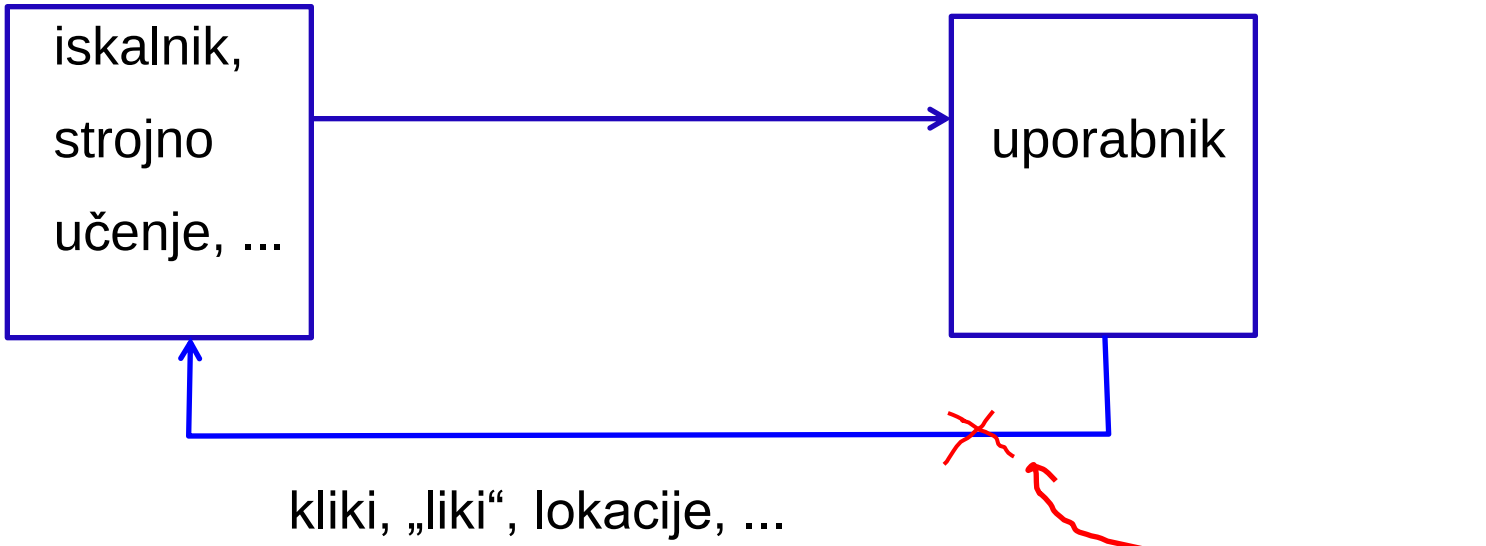
Frey: Europe must guarantee citizens a right to a digital copy of all data about them (Right to a Copy)

Y. Hofstetter o spletnem vodenju ljudi

- N. Wiener: Z mehanizmom povratne zanke lahko vodimo/kontroliramo vsak sistem – vse, stroje in ljudi.



Spletno vodenje uporabnikov



Hofstetter: „Uporabnika loči do svobode le en korak: prekini tukaj!“

Toda .... ?!?

NEKATERE POBUDE

- Future of Life Institute
  - Odprto pismo znanstvenikov UI 2015 (~ 8000 podpisov)
  - Odprto pismo proti avtonomnim orožjem (~ 20.000 podpisov)
  - Konference Beneficial AI 2017, Asillomar AI principles (~4000 podpisov)
- Responsible AI, evropska mreža, v fazi predloga