Tekoälyn filosofiaa

TO1 / Hämäläinen
Luento 1.

Luettavaa

– – –

Stephen Hawking varoittaa: Tekoäly voi syrjäyttää ihmiset ja kehittyä paremmin suoriutuvaksi elämänmuodokseen

The building blocks of Interpretability

Mustan laatikon ongelma

Kun robotti saa tunteet, voiko sitä enää sammuttaa?

Tekoälylle ihmisoikeudet

Euroopan parlamentti: Robotiikkaa koskevat yksityisoikeudelliset säännöt

- https://areena.yle.fi/audio/1-3995328 https://areena.yle.fi/1-50365870
- https://areena.yle.fi/audio/1-4131181

Aloitus: kysytään Siriltä / Alexalta tms.! Vertailkaa esim.

Millainen sää on huomenna?

Soita ‘kaverille’ (kaverin nimi)

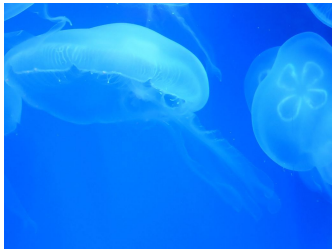
Soita ‘soittolista’

Mikä on Ranskan pääkaupunki?

Mitä onnellisuus on?

Mitä on ruoaksi?

Googlen tekoälyohjelma DeepDream etsii kuvista toistuvia muotoja ja opettelee näin luokittelemaan kuvia ja tunnistamaan esimerkiksi kasvoja. DeepDreamin toiminnan voi kääntää myös toisinpäin, jolloin se korostaa kuvasta löytyviä elementtejä. Tämä on taideteos, jonka koiria tunnistamaan opetettu DeepDream on luonut tämän valokuvan pohjalta:



Peruskäsitteitä: algoritmi

- Algoritmi = sarja seurattavia ohjeita, eräänlainen resepti joka ohjaa tehtävän suorittamista (jos…..ja….niin….)
- Mitä kaikkea algoritmit tekevät?
  - tunnistavat kasvoja ja eläimiä kuvista
  - suosittelevat tv-sarjoja ja musiikkia
  - käyttävät kaivoskoneita, kokoavat autoja, siivoavat, hoitavat vanhuksia...
  - kääntävät Hemingwayn teoksia englannista japaniin virheettömästi
  - jakavat Facebookissa sellaista sisältöä, joka saa ihmiset käyttämään Facebookia jatkossakin
-

Tekoälytutkijoiden arvioita...

Tekoälyn odotetaan

- kääntävän kieliä ihmistä paremmin vuonna 2024
- kirjoittavan bestsellerin vuonna 2049
- tekevän kirurgit tarpeettomiksi vuonna 2053

Tekoäly: ylhäältä alas...

- Ihminen syöttää informaatiota, kone prosessoi sitä ja vertaa siihen uutta informaatiota
  - esim. 1997 shakissa Kasparovin voittanut Deep Blue -ohjelma
- + rationaalisuus ja koneen laskentateho (esim. poikkeamat sydänsähkökäyrissä, ruuhkaennusteet navigointisovelluksessa)
- - luovuuden ja ongelmanratkaisun puute

...vai alhaalta ylös?

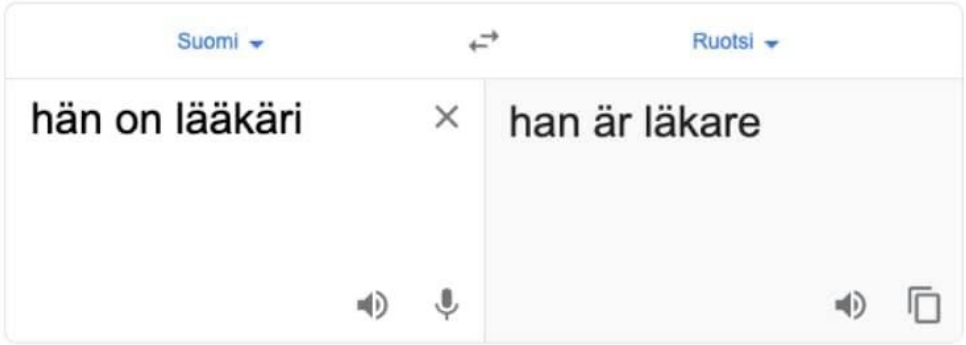
- **koneoppiminen: tekoäly opetetaan oppimaan itse**
- induktiivinen ja deduktiivinen päättely
  - esim. roskapostisuodatin, kirjoitusvirheet hakukoneissa tai automaattisessa tekstinsyötössä
  - Sirin tai Alexan opettaminen - mitä enemmän puhut, sen paremmin se palvelee
- syväoppiva neuroverkko: satojen tasojen tietokonemalli
  - alimmille tasoille syötetty materiaali (esim. kissojen kuvat) + ylempien tasojen komennot > lopulta verkko tunnistaa kissoja sellaisistakin kuvista, joita sinne ei ole alunperin syötetty
  - koneen “ajattelu” tapahtuu neuroverkon välikerroksissa

AlphaZero vs. Stockfish

Ihminen opetti Stockfishin (vuoden 2016 maailmanmestari, vuosisatojen kokemus koodattuna)

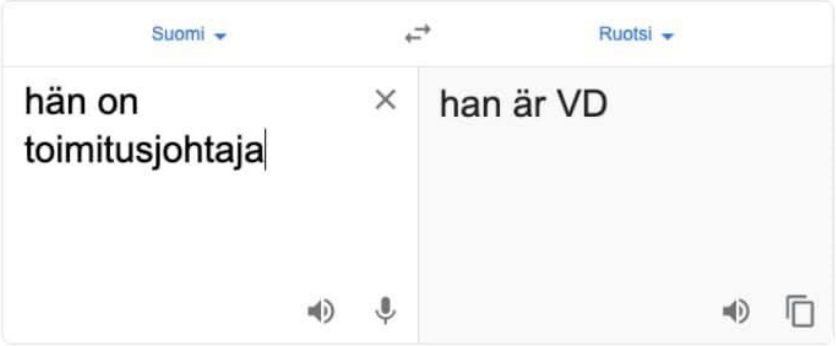
AlphaZero opetteli alusta alkaen itse pelaamaan shakkia ja suunnitteli siirrot 4 tunnissa > **pelasi luovempaa ja taitavampaa shakkia.**

Noin 259 000 tulosta (0,41 sekuntia)



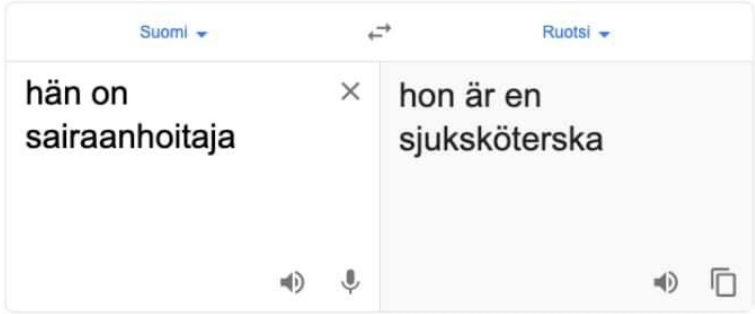
Avaa Google Kääntäjässä Palaute

Noin 259 000 tulosta (0,41 sekuntia)



Avaa Google Kääntäjässä Palaute

Noin 259 000 tulosta (0,41 sekuntia)



Avaa Google Kääntäjässä Palaute

Noin 259 000 tulosta (0,41 sekuntia)

Avaa Google Kääntäjässä Palaute

Peruskäsitteitä: mustan laatikon ongelma

- Algoritmeillä huomattavan suuri ote elämäämme:
  - mereen ajavat kuskit, mustia ihmisiä gorilloiksi tunnistavat ohjelmat, johtajaksi vain miehiä suositteleva HR-algoritmi...
- Itseoppivat ohjelmat ovat suljettuja ulkopuolisilta. Dataa menee sisään, jotain tapahtuu ja toimintaa tai informaatiota tulee ulos - mutta kukaan ei tiedä, miten tämä tarkalleen tapahtuu.

Algoritmien huonot puolet

- Vinoumat: Tekoäly on siis juuri niin hyvä kuin data, joka sille on syötetty. Vinoumat saattavat jäädä myös pimentoon, jos ohjelma on musta laatikko ja tekee päätöksensä ulkopuolisilta piilossa.
  - Uhkana yksilöllinen syrjintä: algoritmi voi löytää dna:sta tai somesta jotain, mistä ei pidä, mutta et tiedä mikä asia on

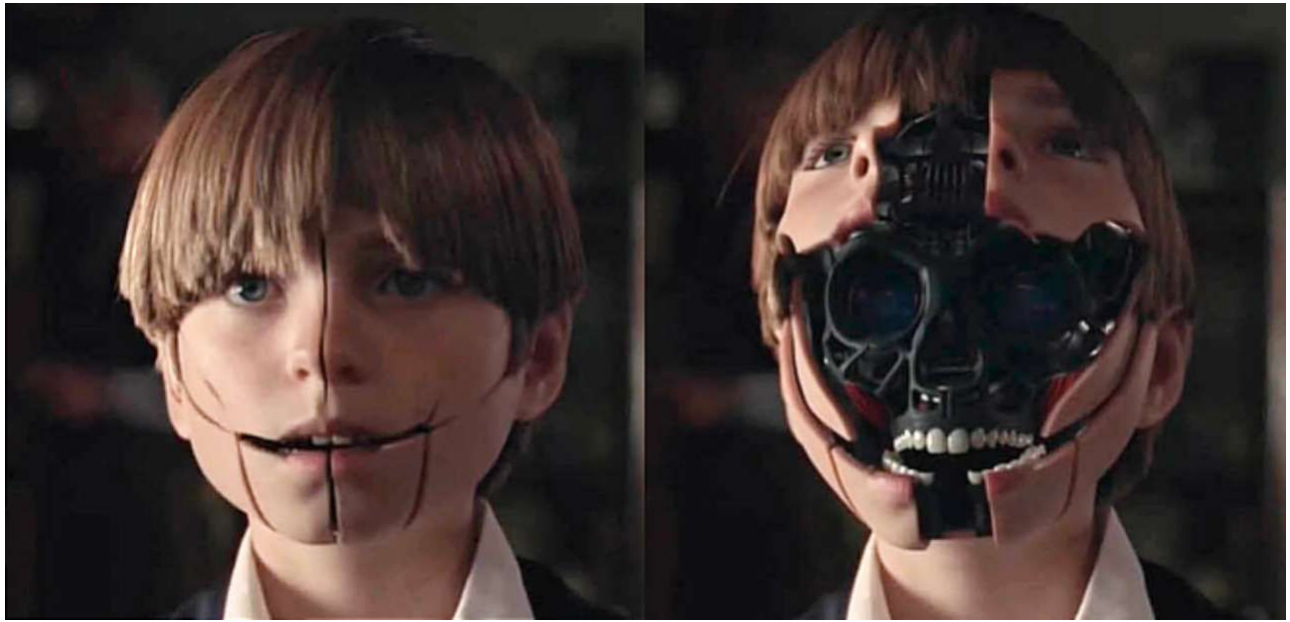
Algoritmien huonot puolet

- Moraalisen vastuun ja harkinnan kysymys: jos itseohjautuva auto ajaa ihmisen päälle tai robotti tekee lääkkeen annosteluvirheen, kenen on vastuu?

Algoritmien hyvät puolet

- tieliikenneturvallisuus
- lääketiede (esim. riskien ja diagnoosien tunnistaminen ajoissa)
- Timo Honkela: Rauhankone, tekoäly vuorovaikutuksen ja kommunikaation avustajana
- ihhimillisten vinoumien välttäminen esim. työhönotossa

Westworld



Funktionalismi ja mieli

- Mieli voi toteutua myös muualla kuin ihmisen aivoissa.
- Tallennuspaikalla tai oliolla ei ole väliä: jos se toimii kuin mieli, se on mieli.

Turingin testi

- jos kone suoriutuu jossain älyllisessä tehtävässä yhtä hyvin kuin ihmiset ja saa ihmiset pitämään sitä ihmisenä (esim. keskustelussa), se on läpäissyt Turingin testin

Turingin testin läpäisseitä esimerkkejä

Eliza: täydellinen psykoterapeutti äänikirja…

[Tietokone läpäisi ensimmäistä kertaa tekoälyä mittaavan testin]

3:24 puhelu kampaamoon

Heikko ja vahva tekoäly

- Turingin testi, Deep Blue ja esim. älypuhelimet ja autonomiset autot esimerkkejä **heikosta tekoälystä**
  - ainoastaan simuloi ajattelua
  - todellinen ymmärrys puuttuu
- **vahva tekoäly** ymmärtää tekemäänsä, sillä on mielentiloja ja jonkinlainen tietoisuus

Ajatuskoe: kiinalainen huone

Vastauksia kiinalaisen huoneen ajatuskokeeseen

Systeemivastaus:

- huone kokonaisena systeeminä ymmärtää kiinaa
- kognitiivisia järjestelmiä on tarkasteltava kokonaisuuksina

Robottivastaus:

- mieli ja ymmärrys syntyy todellisessa vuorovaikutuksessa ympäristön kanssa
- robotti voi oppia, saada reaaliaikaista tietoa maailmasta ja reagoida siihen

Tietoisuus

Mihin eettisiin kysymyksiin (itse)tietoisuus liittyy?

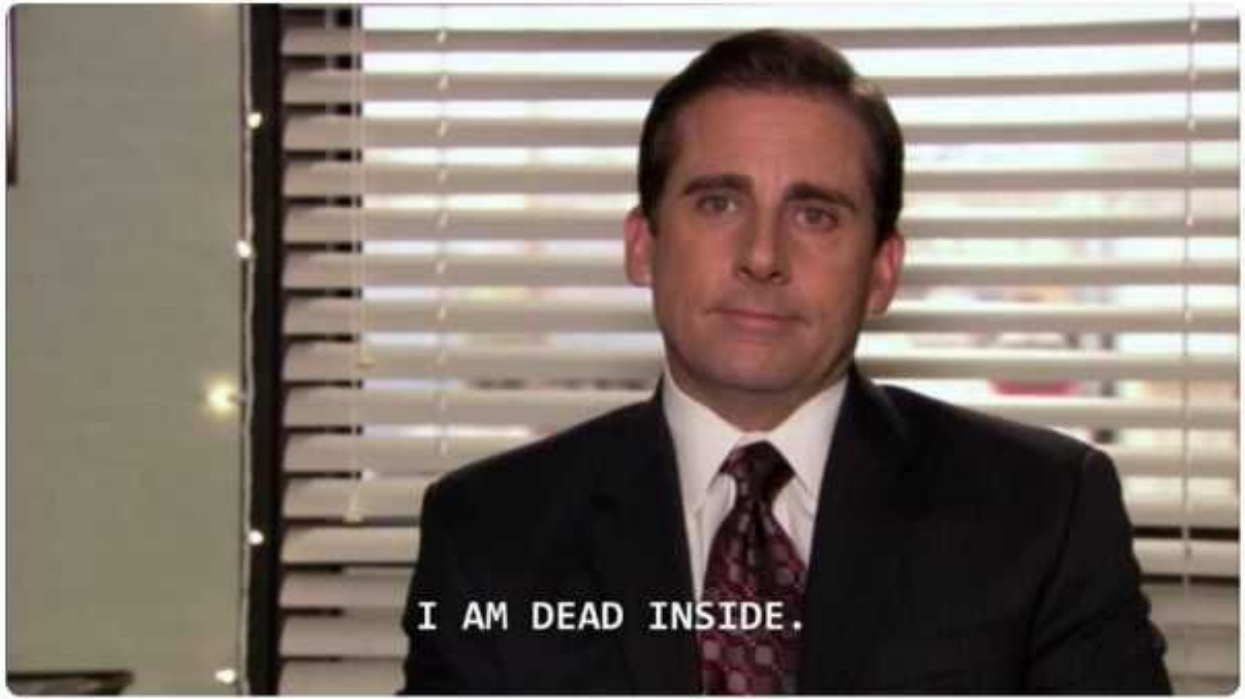
Robotti tulee tietoiseksi itsestään

Tietoisuus

- tietoisuutta tutkivat filosofien lisäksi kognitiotietelijät, psykologit ja tekoälytutkijat
- tietoisuuden kriteerit?
  - kyky havaita, toimia ympäröivässä maailmassa + ymmärrys
  - tietoiset **mielentilat**?
- **mielentilat:**
  - tarina, jota kerromme itsellemme?
  - intersubjektiivisuuden ongelma
  - **tärkeintä kokemus**

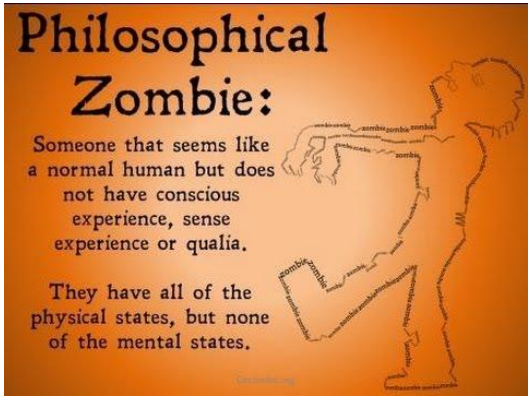
Random Person: hi I’m

Me:



Tietoisuus

Chalmersin zombiajatuskoe:



- kuvitellaan filosofinen zombi, joka pystyy hurraamaan Ilveksen pelissä, ylistämään puolikuivaa valkoviiniä ja kirjoittamaan rakkausrunon kyynel silmässä MUTTA hän ei tunne mitään
- Chalmers:pelkästä ulkoisesta käyttäytymisestä ei voi päätellä, onko olio tietoinen. Olio, eliö tai systeemi on tietoinen ainoastaan silloin jos TUNTUU joltain olla tuo olio, eliö tai systeemi

Tietoisuuden helppo ja vaikea ongelma

- Chalmers: tiede pystyy vastaamaan tietoisuuden helppoihin ongelmiin, kuten
  - Mitä tarkoittaa hereillä olo?
  - Miten aistihavainnot muodostuvat?
  - Miten tarkkaavaisuutta voidaan säädellä?
  - Mitkä aivoalueet liittyvät kasvojentunnistukseen, onnellisuuteen tai kivun tunteeseen?
- Mutta tiede ei voi ottaa kantaa siihen, miksi minkään **pitäisi tuntua miltään** > tietoisuuden vaikea ongelma

Zombiargumentti materialismia vastaan

- Chalmers: jos voimme kuvitella filosofisia zombeja, todellisuus ei ole yksinomaan materialistinen
- Dennett: emme VOI kuvitella filosofisten zombien olemassaoloa, sillä tietoisuus ei ole yksittäine ominaisuus, jonka voisi kuvitella poissa olevaksi

Ota kantaa!

- Robotille pitäisi myöntää ihmisoikeudet, jos
  - se tuntee
  - se on tietoinen
  - jos se on molempia
  - ei missään tapauksessa.
- Tekoäly tekee maailmasta tasa-arvoisemman ja auttaa pääsemään eroon inhimillisten kognitiivisisten vinoumien haitallisesta vaikutuksesta.