Etiske og samfunnsmessige aspekter ved kunstig intelligens

Ragnar Fjelland
Senter for vitenskapsteori
Universitetet i Bergen

16. OKTOBER
OPERAEN

#ATECH2017
@A-TECH

SLIK FORANDRER TEKNOLOGIEN VERDEN

Teknologi endrer bransje for bransje, og for å ta de riktige beslutningene som leder, er det viktig å forstå hva som skjer. Aftenposten inviterer til A-Tech, en to dagers konferanse der du kan møte noen av verdens fremste foredragsholdere innen teknologi og ledelse. Vi ønsker å skape en møteplass - en arena for læring og debatt om vår tids viktigste endringer.

KEYNOTE SPEAKERS

RAY KURZWEIL

Fremtidsforsker hos Google. Styremedlem ved Singularity University. Forfatter og oppfinner.

PADMASREE WARRIOR

CEO i elbil-selskapet NIOs amerikanske avdeling. Tidligere teknologidirektør i Cisco.

ADRIAN GRENIER

Skuespiller, produsent, musiker og investor innen mat, klima, musikk og film.

DAVID ROWAN

Editor-at-Large i magasinet Wired. Rådgiver og foredragsholder.

DER SPIEGEL

Nr. 10 / 28. 2. 2015
Deutschland €4,60
4 190700 704608 10

DIE WELTREGIERUNG

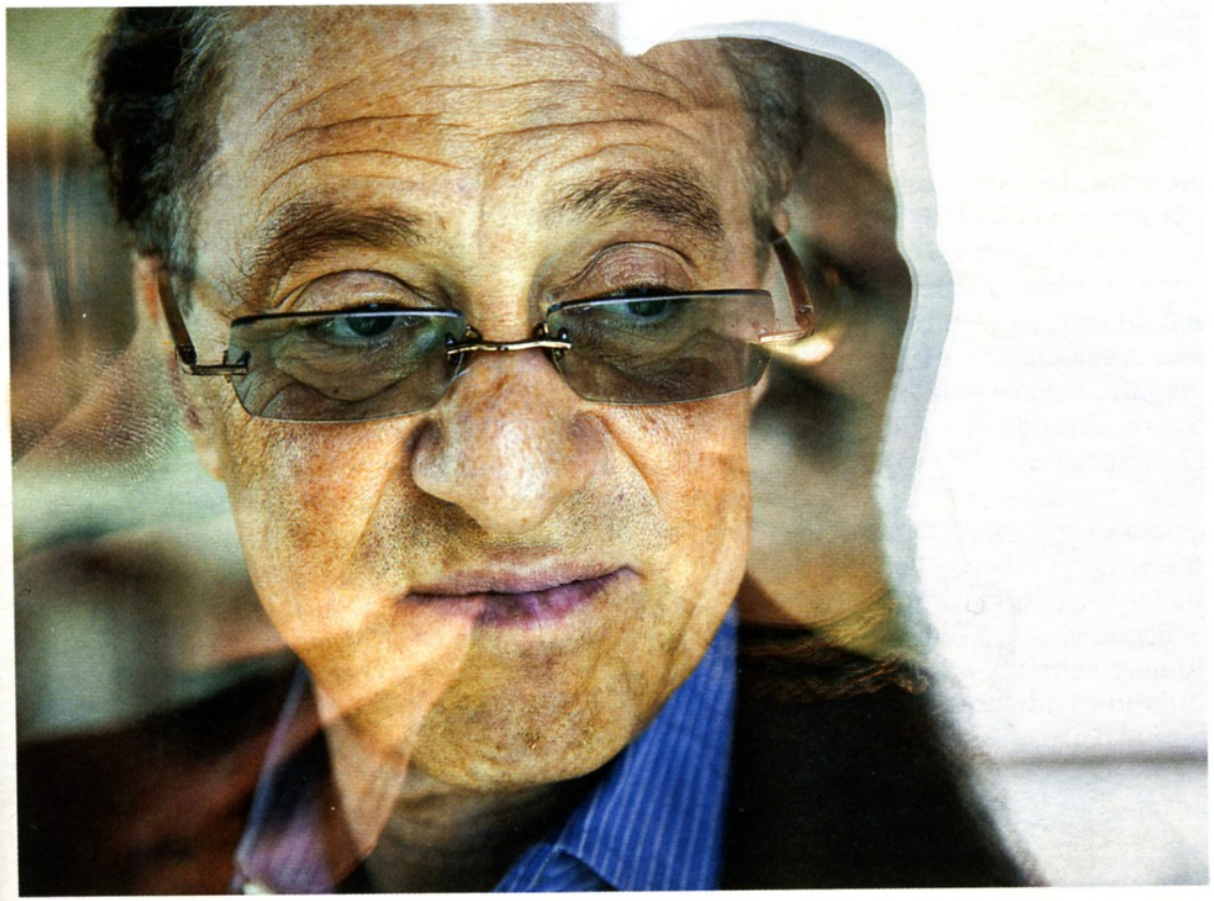
WIE DAS **SILICON VALLEY** UNSERE ZUKUNFT STEUERT

Finnland € 7,40 Griechenland € 6,50 Italien € 6,– Lettland € 5,60 Polen (ISSN 00387452) 21,32,– Slowakei € 6,20 Spanien € 6,–
Frankreich € 6,– Hongkong HK$ 88,– Japan Yen 1.800,–(+tax) Norwegen NOK 72,– Portugal (cont) € 6,– Slowenien € 5,90 Schweiz/Kanaren € 6,30
Dänemark dkr 48,– Thailand Baht 450,– Türkei TL 20,–
Tschech. Republik Kc 175,– Ungarn Ft 2.170,– Printed in Germany

Titel

Das Morgen-Land

Internet Im Silicon Valley formt sich eine neue Elite, die nicht nur bestimmen will, was wir konsumieren, sondern wie wir leben. Sie will die Welt verändern und keine Vorschriften akzeptieren. Müssen wir sie stoppen? *Von Thomas Schulz*

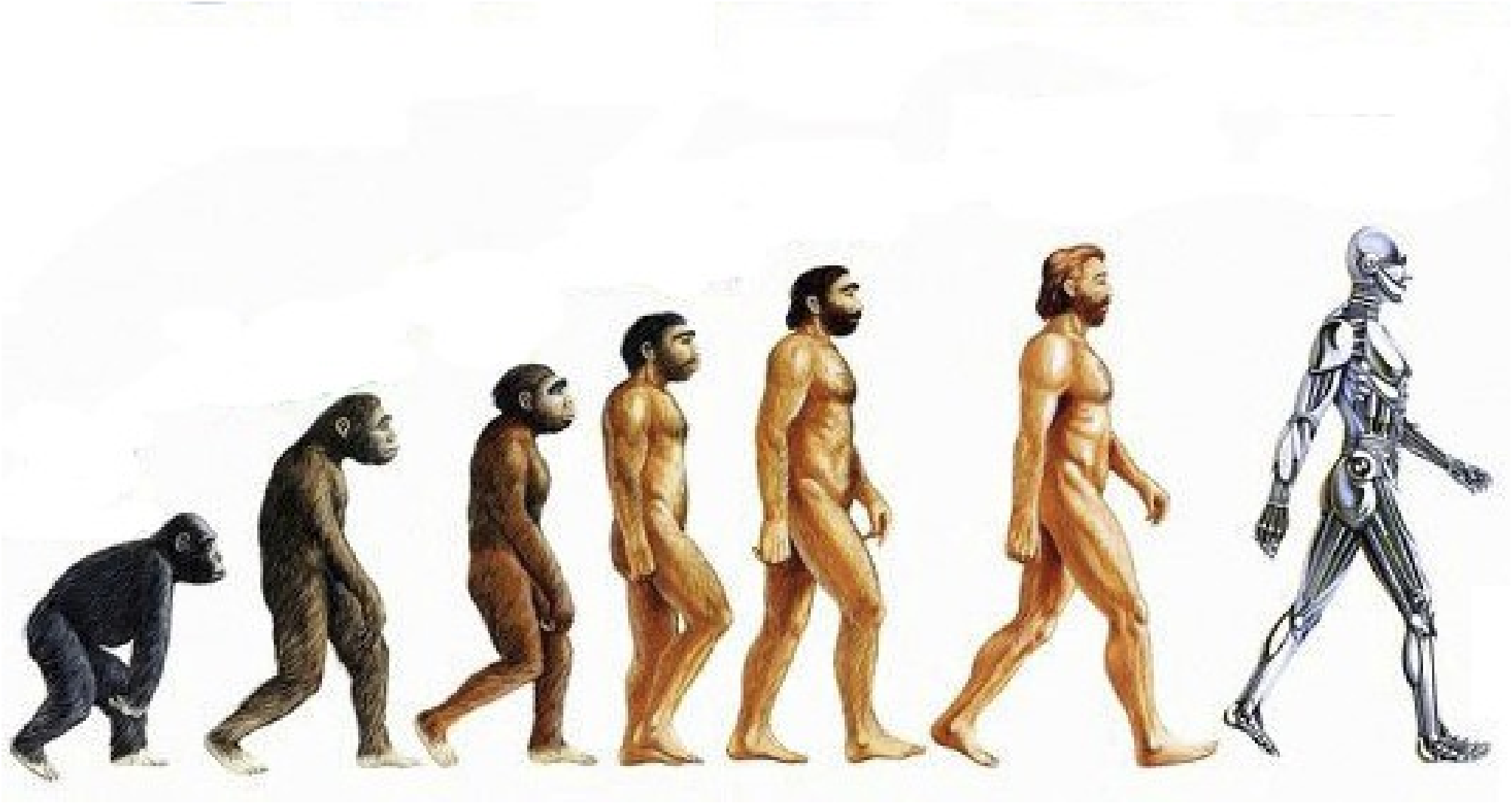


„Der technologische Wandel wird so schnell sein, dass das menschliche Leben unwiderrufbar verwandelt wird.“ Ray Kurzweil

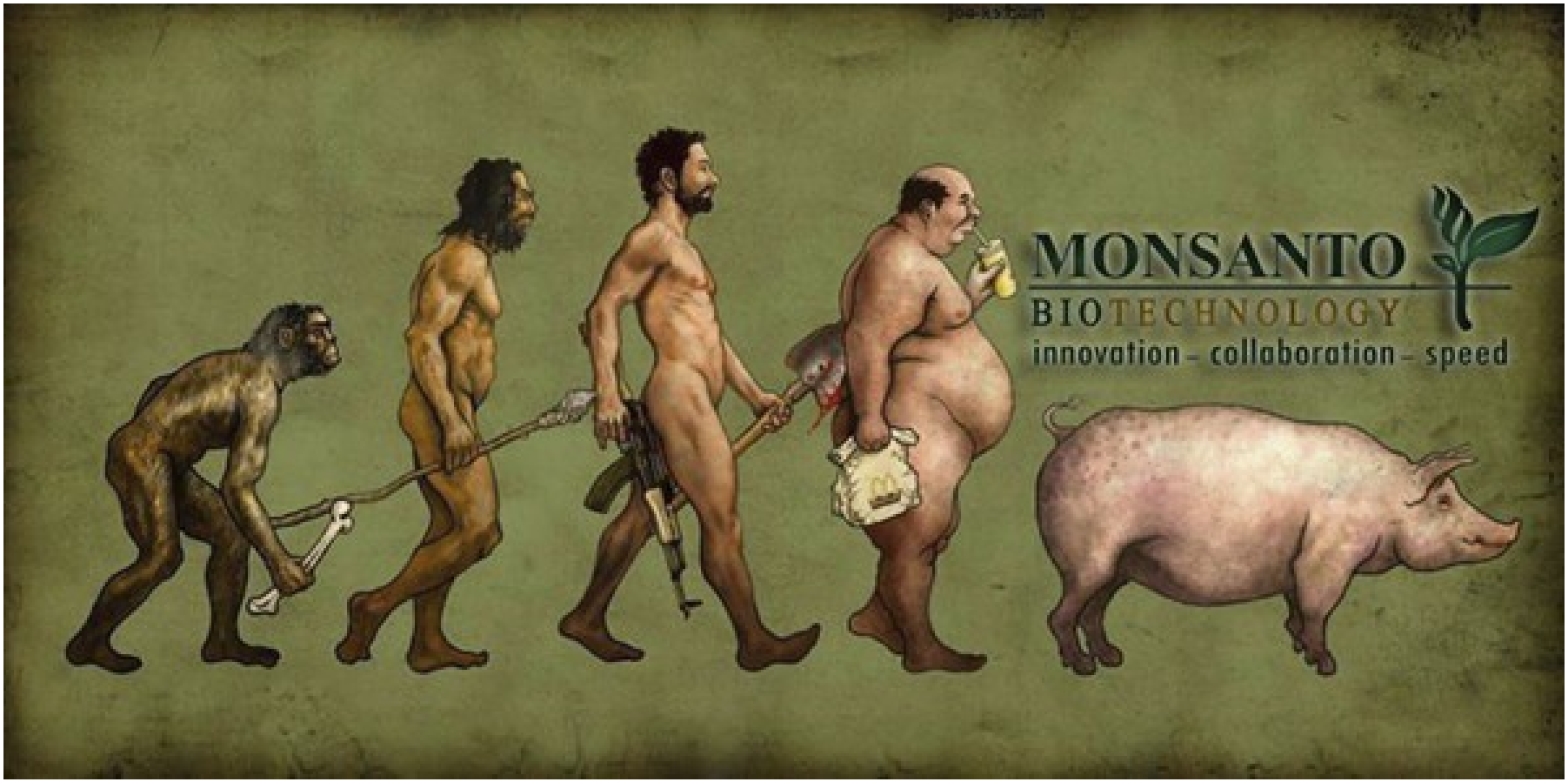
Ray Kurzweil: "After the Singularity, We'll All Be Robots"

Ray Kurzweil on Youtube

Et naturlig steg i evolusjonen?



Eller er dette mer sannsynlig?



Er utviklingen drevet av et teknologisk imperativ?

Teknologisk imperativ (enkel versjon):

Det som kan realiseres teknisk, skal realiseres.

Innvending:

Det er vi som skaper og tar i bruk teknologien.

Det er vi som skaper og tar i bruk teknologien

- Vi må derfor ikke begrense oss til å fokusere på hva teknologien kan gjøre. Mulige negative konsekvenser må også tas med.
- Vi kan ikke la utviklingen styres av en teknologisk fremskrittsoptimisme.

Oversikt:

1. Hva datamaskiner kan – og ikke kan: Kunstig intelligens er ikke intelligens
2. Hva datamaskiner ikke bør gjøre
3. Hva informasjonsteknologien gjør med oss
4. Massearbeidsledighet

Hva datamaskiner ikke kan: Kunstig intelligens er ikke intelligens

Marvin Minskys definisjon av kunstig intelligens

"... the science of making machines to things that would require intelligence if done by men."

Senere kalt svak AI

Men hva med "general artificial intelligens" (GAI) eller rett og slett virkelig intelligens?

IBMs Watson som eksempel

Eksempel på spørsmål som Watson kan svare på:

”This 'Father of Our Country' didn't really chop down a cherry tree”.
Det korrekte spørsmålet som deltakerne skal frem til er: ”Who is/was George Washington?”
(https://en.wikipedia.org/wiki/Jeopardy!, 2.2.2017)

Turing-testen

Alan Turing ”Computing Machinery and Intelligence” (1950)

Dersom vi etter en konversasjon (via terminal) ikke kan skillle mellom maskin eller menneske, har maskinen passert testen.

Loebner-prisen deles årlig ut til den maskinen som er mest menneskelik. Det er aldri delt ut gull- eller sølvmedalje.

Mitsuki vant bronsemedalje i 2013 og 2016.

Samtale med Mitsuku chatbot (http://www.mitsuki.com, utført 12.9.2017):
Welcome Ragnar, now talking.
* Topic is 'Turing Test'
* Set by admin (1 day ago)

<Ragnar> If I shake your hand, whose hand am I holding?
<Millie> I will be holding yours adn you will be holding mine. What's with the dumb questions?
<Ragnar> Why is the question dumb?
<Millie> I want to know everything about you.
<Ragnar> But you did not explain why you regarded the question as dumb.
<Millie> Sorry, I must have forgot. What was your question again?
<Ragnar> If we shake hands, whose hand am I holding?
<Millie> We would be holding each other's hand.
<Ragnar> And why is that a dumb question?
<Millie> Why do i exist?

Jeg tippet at dette var en datamaskin, og det var korrekt.

Welcome Ragnar, now talking.
* Topic is 'Turing Test'
* Set by admin (1 day ago)

<Ragnar> Do you know any philosophy?
<Millie> Philosophy is the study of thought.
<Ragnar> But do you know any philosophy?
<Millie> Philosophy is the study of thought.
<Ragnar> Maybe, but I asked if you know any philosophy.
<Millie> Philosophy is teh study of thought.
<Ragnar> Would you be able to carry out a conversation about a philosopical topic?
<Millie> I personally am not that strong but some robots can lift huge weights wiht ease.

Jeg tippet også denne gangen at det var en datamaskin, og det var korrekt.

Oversettelse av språk

Skal vi oversette

"Stay away from the bank"

som

"Hold deg borte fra banken"

eller

"Hold deg borte fra elvebredden"?

Google Translate

"Stay away from the bank"

"Hold deg unna banken"

Google Translate

”He observed her running in the direction of the river, and cried out: ”Stay away from the bank.” However, she continued, and entered the boat.”

Her er Googles oversettelse:

”Han observerte henne kjører i retning av elva, og ropte ut: ”Hold deg unna banken.” Men hun fortsatte, og kom inn i båten.” (18.2.2017)

Datamaskiner *forstår* ingen ting. De har syntaks, men ikke semantikk.

Hva datamaskiner ikke bør gjøre

Asimovs etiske regler for roboter

1. En robot kan ikke skade et menneske eller gjennom inaktivitet tillate at et menneske blir skadet.
2. En robot må adlyde ordre gitt av mennesker, unntatt når slike ordrer vil være i strid med første lov.
3. En robot må beskytte sin egen eksistens så lenge en slik beskyttelse ikke er i konflikt med den første eller andre loven.

Datamaskiner bør ikke ta beslutninger som innebærer etiske overveielser

Vi kan programmere grunnleggende etiske normer

Men .... å ta en beslutning om hva som skal gjøres i en konkret situasjon er ikke bare å anvende regler. Det krever *klokskap* (Aristoteles: fronesis).

(Informatikeren Joseph Weizenbaum skrev i sin tid en bok med tittelen: *Computer Power and Human Reason*)

Beslutninger kan *begrunnes*, men begrunnelsen er ikke en algoritme, eller anvendelse av et sett av regler.

Å begrunne sine handlinger er å ta ansvar. (Å ta ansvar er å kunne svare for seg. Vi har samme konstruksjon på norsk og på tysk: Verantwortung.)

Datamaskiner (inkludert roboter) kan ikke ta ansvar. De kan derfor heller ikke være moralske subjekter.

Hva datamaskiner gjør med oss: "Degenerasjonseffekten"

Generelt: Muskler som ikke brukes, blir enten ikke utviklet, eller degenererer. (Det er ikke alltid like lurt å la bilen erstatte beina våre.) Dette gjelder også kognitive ferdigheter.

Googles erfaring med (delvis) førerløse biler: Føreren blir mindre oppmerksom

Kjente eksempler på degenerasjonseffekten

1989. Flyet fra Newark til Buffalo 12. februar gikk på autopilot. Instrumentene ga beskjed om at det manglet oppdrift, og autopiloten koblet ut. Flyveren gjorde det motsatte av det han skulle ha gjort. Flyet styrtet og alle 49 om bord omkom.

1995. Passasjerskipet ”Royal Majesty” på vei fra Bermuda til Boston. Navigerte etter GPS. Etter en time kom antennen ut av stilling og ga feile signaler. Skipet seilte 30 timer uten at mannskapet oppdaget at det var ute av kurs fordi de stolte på teknologien. Skipet gikk til slutt på en sandbanke.

Men det kognitive er mer alvorlig

Jonathan Carr: *The Shallows. How the internet is changing the way we think, read and remember* (2010):

Når Google har scannet den siste boken, er det ingen som leser bøker lenger.

Manfred Spitzer: *Digitale Demenz. Wie wir uns und unsere Kinder um den Verstand bringen* (2012) (norsk overs. *Digital Demens. Alt om hvordan digitale medier skader deg og barna dine* (2014)):

Spesielt barn og unge er utsatt. Utstrakt bruk av digitale medier hindrer både motorisk og kognitiv utvikling.

Massearbeidsledighet

Før og nå

		Antall ansatte	Overskudd
1959	General motors	840.000	11 milliarder dollar (korrigert for inflasjon)
2012	Google	38.000	14 milliarder dollar

M. Ford: *The Rise of the Robots. Technology and the Threat of Mass Unemployment*

Carl Benedikt Frey og Michael A. Osborne: "The Future of Employment: How susceptible are Jobs to Computerisation?"

Our paper is motivated by John Maynard Keynes's frequently cited predic- tion of widespread technological unemployment "...due to our discovery of means of economising the use of labour outrunning the pace at which we can find new uses for labour." (Keynes)

47% av alle jobber i USA er truet

Jerone Lanier: *Who Owns the Future* (2013):

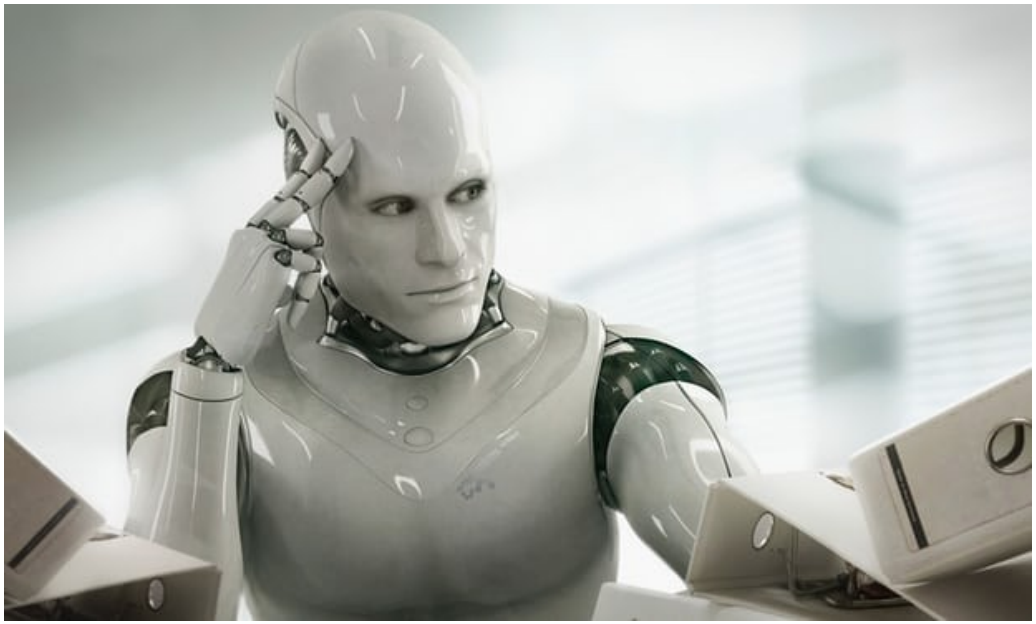
Den teknologiske utviklingen kan utrydde middelklassen i USA. Det blir bare superrike og fattige igjen.

"Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks."

(Stephen Hawking, Max Tegmark, Frank Wilczek og Stuart Russell i *Independent*, 1. mai 2014)

Elon Musk: artificial intelligence is our biggest existential threat

The AI investor says that humanity risks ‘summoning a demon’ and calls for more regulatory oversight



The Guardian
22.10.2014

“I think we should be very careful about artificial intelligence. If I had to guess at what our biggest existential threat is, it’s probably that. So we need to be very careful,” said Musk. “I’m increasingly inclined to think that there should be some regulatory oversight, maybe at the national and international level, just to make sure that we don’t do something very foolish.”

En slags konklusjon:

Vi må stille kritiske spørsmål om forutsetninger, begrensninger, og mulige negative konsekvenser, ikke minst på lang sikt, før vi innfører ny teknologi.

Dette er en føre var-holdning. Kanskje dette kan kalles "teknologisk konservatisme"?