

# MVE420: Nya teknologier, global risk och mänsklighetens framtid

<http://www.math.chalmers.se/Math/Grundutb/CTH/mve420/1819>

Introduktionsföreläsning  
26 mars 2019

Olle Häggström

Ämnet för denna kurs är nya och framtida teknologier som...

- ▶ kan bringa mänskligheten enorma frukter, ekonomiskt eller i form av t.ex. välbefinnande, hälsa och längre liv,
- ▶ för med sig risker av liknande dignitet, i vissa fall till och med mänsklighetens undergång.

Detta för oss långt bortom det ganska korta perspektiv där vi vet med bestämdhet vad som går att göra och vad som är att vänta. Det i sin tur ställer stora krav på oss i att **spekulera med gott omdöme**, vilket bl.a. innebär:

- (a) en välavvägd grad av ödmjukhet, och en förmåga att skilja mellan spekulation och prediktion,
- (b) en vilja att genom studium av argument för och emot olika spekulativa scenarier kunna nyansera, och skilja de välgrundade från de ogrundade, samt de rimliga från de orimliga.

Till de teknikområden som vi har anledning att studera hör

- ▶ artificiell intelligens och robotar,
- ▶ "förbättrandet" av människokroppen, människans kognitiva kapacitet och den mänskliga naturen, på t.ex., farmakologisk eller genetisk väg, eller med elektronik inbyggd innanför skallbenet,
- ▶ radikal nanoteknologi,
- ▶ syntetisk biologi,
- ▶ massövervakning,
- ▶ geoengineeringmetoder för att manipulera jordens klimat,
- ▶ kolonisation av världsrymden,
- ▶ ...

Kursen består av...

- ▶ **föreläsningar**, till stor del inriktade på att ge teoretiska redskap och ramverk att bedriva detta slags futurologi,
- ▶ **projekt**, där studenterna två och två väljer ut ett område (t.ex. ett visst teknikområde, ett visst globalt problem eller en viss katastrofrisk) att studera närmare.

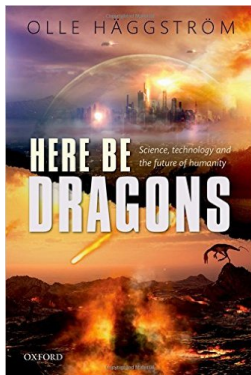
Examinationen består av

- ▶ **projektet**, som redovisas skriftligt och muntligt,
- ▶ **en separat essä** om innehållet i föreläsningarna.

## Viktiga hålltider/deadlines:

- ▶ **29 mars** (nu på fredag) **kl 10-12:** kort-kort muntlig presentation av era preliminära projektidéer.
- ▶ **1 april, deadline P1:** skriftlig anmälan av projektämne.
- ▶ **10 april, deadline P2:** skriftlig redovisning av hur etiska aspekter skall hanteras i projektet.
- ▶ **16 maj, deadline P3 (frivillig):** Preliminär inlämning av projektrapport.
- ▶ **23 och 24 maj:** Muntliga projektpresentationer.
- ▶ **28 maj, deadline P4:** Slutig inlämning av projektrapport.
- ▶ **3 juni, deadline E1:** Inlämning av essä.

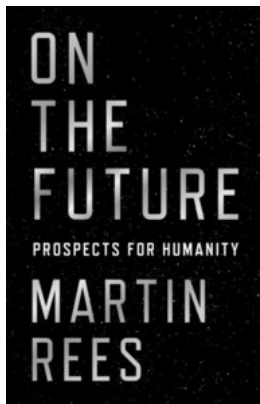
Någon obligatorisk kurslitteratur har vi inte, men en rimlig startpunkt för den som söker lämpligt tema för sitt projekt är min bok **Here Be Dragons: Science, Technology and the Future of Humanity** (Oxford University Press, januari 2016).





## Here Be Dragons: innehållsförteckning

1. Science for good and science for bad
2. Our planet and its biosphere
3. Engineering better humans?
4. Computer revolution
5. Going nano
6. What is science?
7. The fallacious Doomsday Argument
8. Doomsday nevertheless?
9. Space colonization and Fermi's paradox
10. What do we want and what should we do?



En av ambitionerna med kursen är ge förmåga att välgrundat ta ställning till radikala teknikoptimistiska idéer som t.ex. de som uttrycks i citaten på de följande fyra bilderna.

The first ultraintelligent machine is the last invention that man need ever make. **(Jack Good, 1965)**

We have the means right now to live long enough to live forever. Existing knowledge can be aggressively applied to dramatically slow down aging processes so we can still be in vital health when the more radical life extending therapies from biotechnology and nanotechnology become available. But most baby boomers won't make it because they are unaware of the accelerating aging process in their bodies and the opportunity to intervene. **(Ray Kurzweil, 2005)**

Imagine what the world might be like if we were *really* good at making things – better things – cleanly, inexpensively, and on a global scale. What if ultra-efficient solar arrays cost no more to make than cardboard and aluminum foil and laptop computers cost about the same? Now add ultra-efficient vehicles, lighting, and the entire behind-the-scenes infrastructure of an industrial civilization, all made at low cost and delivered and operated with a zero carbon footprint.

If we were *that* good at making things, the global prospect would be, not scarcity, but unprecedented abundance – radical, transformative, and sustainable abundance. We would be able to produce radically more of what people want and at a radically lower cost – in every sense of the word, both economic and environmental. **(Eric Drexler, 2013)**

I see a bright future for the biotechnology industry when it follows the path of the computer industry, [...] becoming small and domesticated rather than big and centralized. [...]

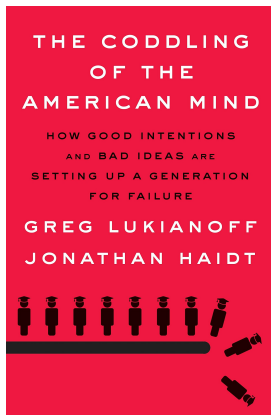
Domesticated biotechnology, once it gets into the hands of housewives and children, will give us an explosion of diversity of new living creatures, rather than the monoculture crops that the big corporations prefer. New lineages will proliferate to replace those that monoculture farming and deforestation have destroyed. Designing genomes will be a personal thing, a new art form as creative as painting or sculpture.

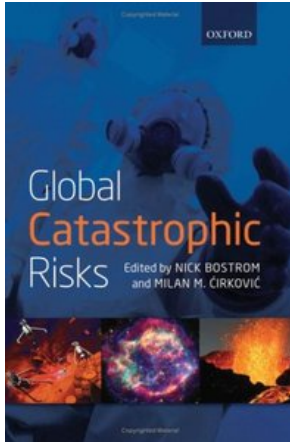
Few of the new creations will be masterpieces, but a great many will bring joy to their creators and variety to our fauna and flora. The final step in the domestication of biotechnology will be biotech games, designed like computer games for children down to kindergarten age but played with real eggs and seeds rather than with images on a screen. Playing such games, kids will acquire an intimate feeling for the organisms that they are growing. The winner could be the kid whose seed grows the prickliest cactus, or the kid whose egg hatches the cutest dinosaur. **(Freeman Dyson, 2007)**

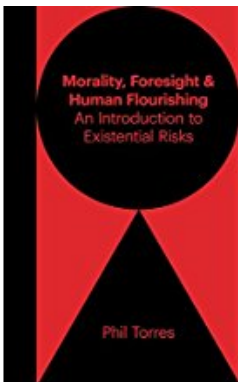
Jämte dessa entusiastiska visioner finns också ett antal undergångsscenarier. Kärnvapenkrig och klimatkatastrof är de två mest kända exemplen, men det finns fler som vi kan ha anledning att beakta och försöka undvika.

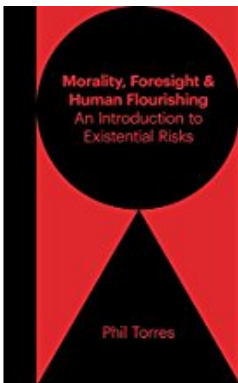


Jämte dessa entusiastiska visioner finns också ett antal undergångsscenarier. Kärnvapenkrig och klimatkatastrof är de två mest kända exemplen, men det finns fler som vi kan ha anledning att beakta och försöka undvika.









Phil Torres har också en väldigt användbar läslista om existentiell risk på <https://www.xrisksinstitute.com/reading-list>



## EXISTENTIAL RISK TO HUMANITY

A cross-disciplinary thematic programme on global catastrophic risk

### THE PROGRAMME

Throughout September and October 2017, researchers in environmental sciences, philosophy, mathematics, physics, astronomy, computer science, political science, economics and other fields will gather in Gothenburg with the goal of advancing our understanding of existential risk to humanity. Such risk emanates both from nature and from ourselves, but what are the biggest risks and how can they be managed and minimized?

The programme is led by Dr Anders Sandberg from the University of Oxford, and will include daily seminars, a workshop, and public lectures.

### GO-CAS

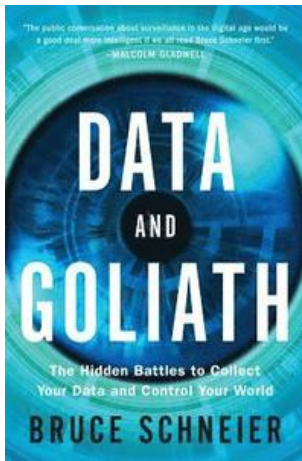
The programme is funded by Go-CAS, the Gothenburg Chair Programme for Advanced Studies.

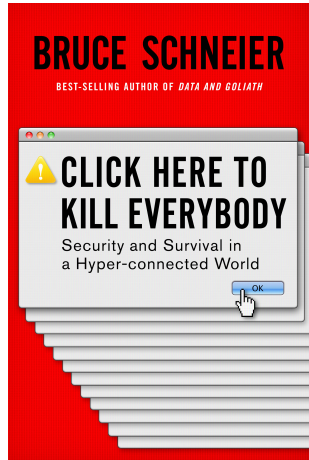
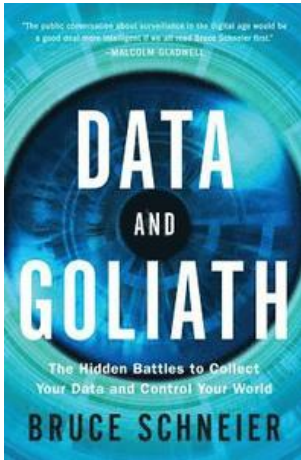
### INFORMATION

Visit <http://www.chalmers.se/en/centres/GoCAS/Events/>

GOETENBURG CHAIR PROGRAMME FOR ADVANCED STUDIES

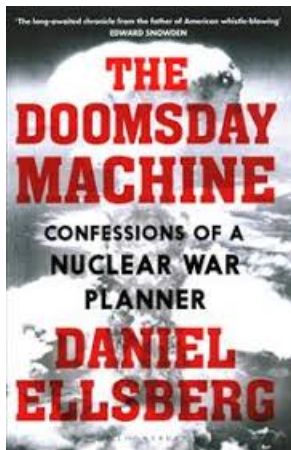
- ▶ Nick Bostrom (2018) The vulnerable world hypothesis, preprint, Future of Humanity Institute.
- ▶ Olle Häggström och Catherine Rhodes (2019) *Existential Risk to Humanity*, specialnummer av tidskriften *Foresight*.
- ▶ Dennis Pamlin och Stuart Armstrong (2015) *Global Challenges – 12 Risks that Threaten Human Civilization*, Global Challenges Foundation.



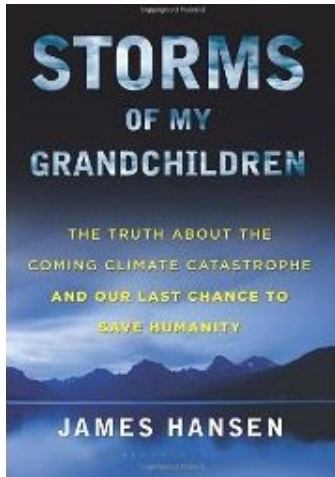


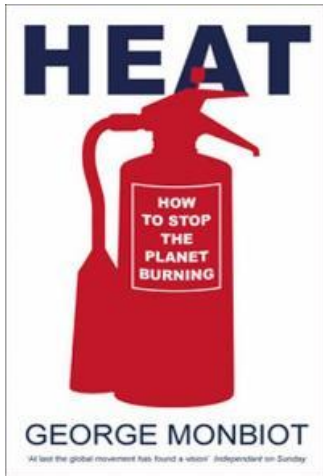


# Kärnvapen



# Klimatförändringar

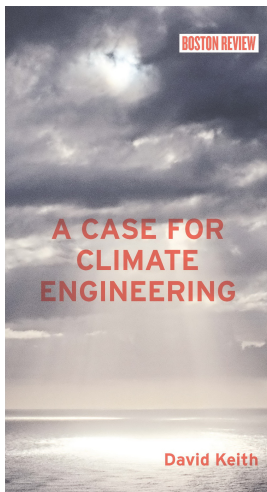






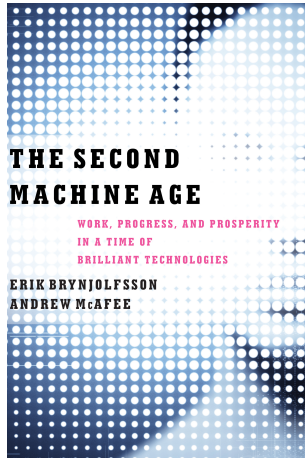
«... 'Makten över klimatet'  
är den bästa klimat-  
boken hittills ...»  
SVENSKA DAGBLADET

«Skall ni läsa  
en enda bok om  
klimatfrågan i år,  
så vet ni nu vilken.»  
SVENSKA NYA TIDNINGEN



# Robotisering och arbetsmarknad







- ▶ Carl Benedikt Frey and Michael Osborne (2013) The future of employment: How susceptible are jobs to computerization?, preprint, Oxford University.
- ▶ Stefan Fölster (2014) Vartannat jobb automatiseras inom 20 år – utmaningar för Sverige, Stiftelsen för Strategisk Forskning.

**AI**  
**SUPER-**  
**POWERS**  
**CHINA,**  
**SILICON VALLEY,**  
AND THE  
**NEW WORLD ORDER**  
**KAI-FU LEE**

## Mer radikala robotscenarier, artificiell intelligens



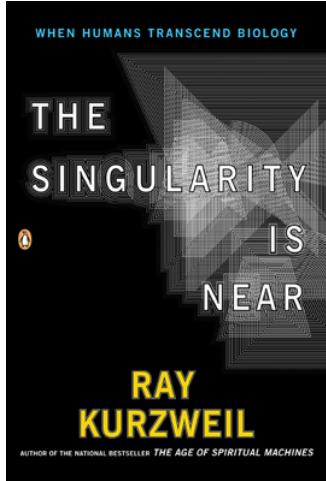
My contention is that machines can be constructed which will simulate the behaviour of the human mind very closely. [...] It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers. There would be no question of the machines dying, and they would be able to converse with each other to sharpen their wits. At some stage therefore we should have to expect the machines to take control. **(Alan Turing, 1951)**

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an intelligence explosion, and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make. **(Jack Good, 1965)**



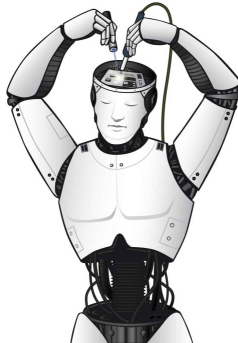
In contrast with our intellect, computers double their performance every 18 months, so the danger is real that they could develop intelligence and take over the world. [...] We should follow [the path of genetic engineering] if we want biological systems to remain superior to electronic ones. **(Stephen Hawking, 2001)**





# THE HANSON-YUDKOWSKY AI-FOOM DEBATE

ROBIN HANSON AND ELIEZER YUDKOWSKY



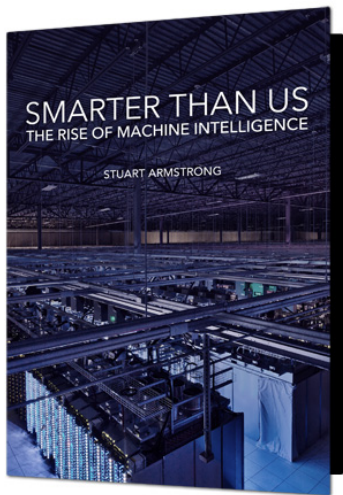
"There are things in this book that could mess with your head."  
—VERNOB VINCE, computer scientist;  
essayist, "The Coming Technological Singularity"

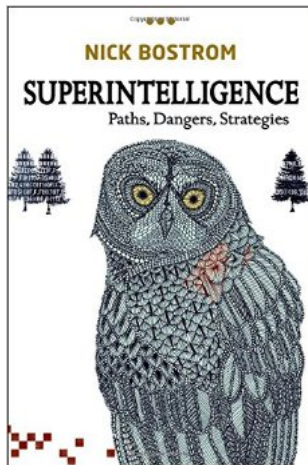
# SINGULARITY RISING

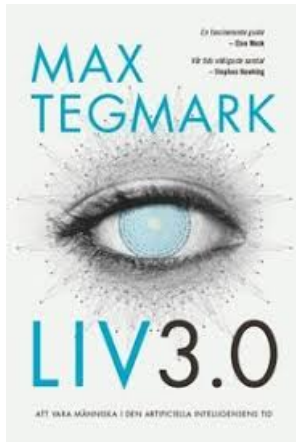


Surviving and Thriving in a Smarter,  
Richer, and More Dangerous World

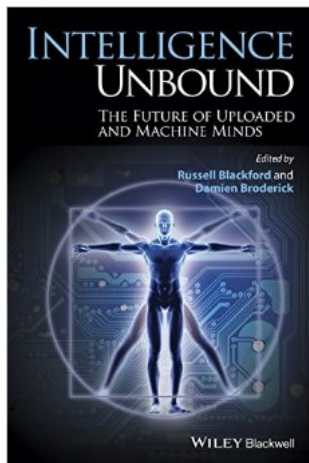
JAMES D. MILLER





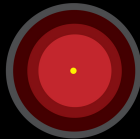






# THE TECHNOLOGICAL SINGULARITY

MURRAY SHANAHAN

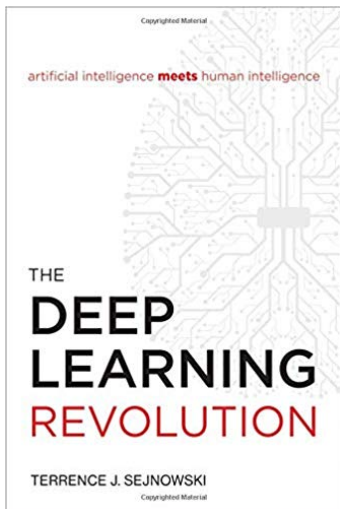


THE MIT PRESS ESSENTIAL KNOWLEDGE SERIES

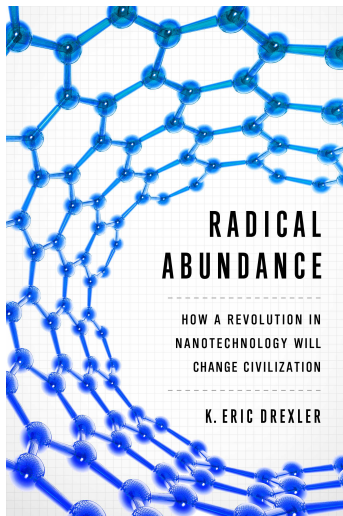
- ▶ Eliezer Yudkowsky (2013) Intelligence explosion microeconomics, Machine Intelligence Research Institute, Berkeley.
- ▶ Kaj Sotala och Roman Yampolskiy (2015) Responses to catastrophic AGI risk: a survey, *Physica Scripta* **90**, 018001.

De flesta bedömare menar att det stora genombrottet för generell AI *inte* är nära förestående, men det kan ändå vara värt att ta en titt på det hetaste området inom AI-utveckling idag: **deep learning**.

- ▶ Yann LeCun, Yoshua Bengio och Geoffrey Hinton (2015) Deep learning, *Nature* **521**, 436–444.
- ▶ Jürgen Schmidhuber (2015) Deep learning in neural networks: an overview, *Neural Networks* **61**, 85–117.
- ▶ David Silver och 12 andra (2017) Mastering chess and shogi by self-play with a general reinforcement learning algorithm, preprint.



# Radikal nanoteknologi

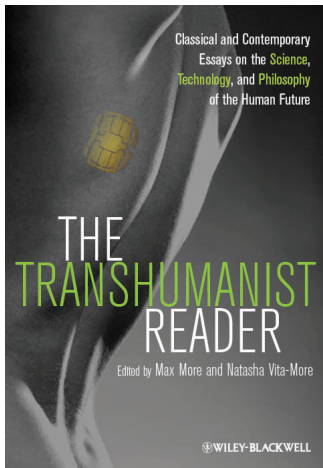


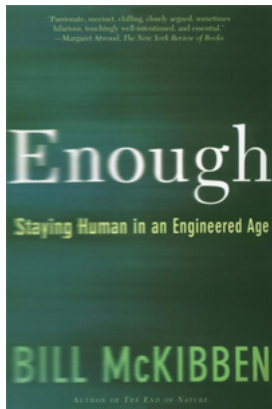
# Rymdkolonisering

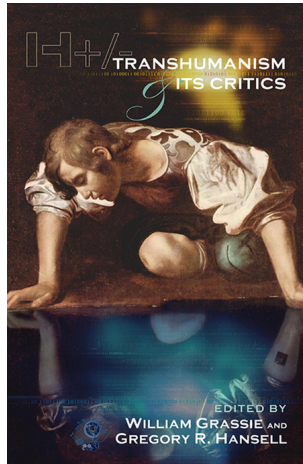


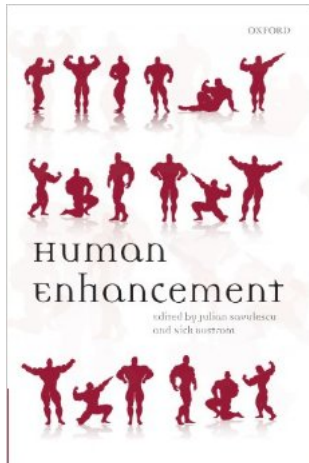
- ▶ Tom Murphy (2011) Stranded resources, *Do the Math*, October 25.
- ▶ Stuart Armstrong and Anders Sandberg (2013) Eternity in six hours: Intergalactic spreading of intelligent life and sharpening the Fermi paradox, *Acta Astronautica* **89**, 1–13.
- ▶ Nick Beckstead (2014) Will we eventually be able to colonize other stars? Notes from a preliminary review, Oxford University.
- ▶ Phil Torres (2018) Space colonization and suffering risks: Reassessing the “maxipok rule”, *Futures* **100**, 74–85.
- ▶ Milan Cirkovic (2018) Space colonization remains the only long-term option for humanity: A reply to Torres, *Futures* **105**, 166–173.

# Modifiering av den mänskliga naturen









ROBIN HANSON

# THE AGE OF EM

*Work, Love,  
and Life when  
Robots Rule  
the Earth*



En annorlunda ingång till ett projekt skulle kunna vara att börja med något relevant stycke skönlitteratur, som någon av följande klassiker. (Observera dock att detta *inte* är en kurs i litteraturhistoria, så projektet kan inte stanna vid en litterär analys; tonvikten behöver ligga på mer vetenskapliga analyser av framtiden.)

