

**CHALMERS**



UNIVERSITY OF GOTHENBURG

# AI TEKNIKEN, MÄNNISKAN, SAMHÄLLET, KUNSKAPEN OCH ETIKEN

Gordana Dodig-Crnkovic

Chalmers University of Technology and Gothenburg University,  
& Mälardalen University, Sweden

[dodig-crnkovic@chalmers.se](mailto:dodig-crnkovic@chalmers.se)

<http://gordana.se/>

# Gordana Dodig-Crnkovic



Professor i IxD på Chalmers tekniska högskola  
Professor i Datavetenskap på Mälardalens högskola  
Disputerat i teoretisk fysik  
Disputerat i datavetenskap och filosofi  
Forskar i beräkningsmodeller för kognitiva system & etik  
Undervisat i forskningsmetodik, vetenskapsteori, och etik.  
Jobbar tvärvetenskapligt. Medlem i CHALMERS AI Etikkommittee

## Etikkurser

Professional Ethics CDT409 (earlier version CD5590) MDH

Research Ethics & Sustainable Development, Chalmers University of Technology, PhD course, GFOK025 (2014 – 2017)

Emerging Trends in Interaction Design – CIU280 - Chalmers – Ethical aspects of new technologies lecture

Human-Centric Design – 2018 Chalmers – HCD-TDA486 – Design Ethics lecture

# Forskning inom tillämpad etik

## 2020

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2020) **Real-world Ethics for Self-Driving Cars**. In Proceedings of the 42nd International Conference on Software Engineering (ICSE '20) Poster Track.

## 2019

Giovagnoli, R., Crucitti, A., Dodig Crnkovic, G. (2019) **Robotics and the quality of life: the case of robotics assisted surgery**. Philosophical Inquiries - philing VII, 1-2019, pp. 77-87. ISSN (print) 2281-8618-ETS.  
<https://www.philing.it/index.php/philing/article/view/240>

## 2018

Holstein, T. Dodig-Crnkovic G. **Avoiding the Intrinsic Unfairness of the Trolley Problem**. 2018 ACM/IEEE International Workshop on Software Fairness ACM ISBN 9781450357463/18/05  
<https://doi.org/10.1145/3194770.3194772>

Holstein, T. Dodig-Crnkovic G. and Pelliccione P. **Ethical and Social Aspects of Self-Driving Cars**,  
<http://arxiv.org/abs/1802.04103> <https://easychair.org/publications/preprint/CQbW>

## 2017

Johnsen A., DodigCrnkovic G., Lundqvist K., Hänninen K., Pettersson P. **Risk-based Decision-making Fallacies: Why Present Functional Safety Standards Are Not Enough**. MARCH2017 International Workshop on decision Making in Software Architecture @ ICSA 2017 Gothenburg, Sweden. 04.04.2017. Published in: Software Architecture Workshops (ICSAW), 2017 IEEE International Conference. DOI: 10.1109/ICSAW.2017.50 pp. 153160.

For the complete list, see: <http://www.gordana.se/work/publications.html>

# Föreläsningar i etik

## 2020

**ETHICS OF ARTIFICIAL INTELLIGENCE**, CHALMERS 2020 08 10

AI From Philosophy to Ethics to Science to Technology to Law and Back. Thoughts from an European perspective  
CHAIR ETHICS SEMINAR, CHALMERS 2020 02 25

- 1-AI-Philosophy-Sciences
- 2-AI-Technology-Values
- 3-AI-Ethics-Guidelines-Policies
- 4-AI-Ethics-Initiatives-Laws

**ETHICS IN DESIGN & SUSTAINABLE FUTURES**, CHALMERS 2020 04 21

## 2019

**Ethics in the Design of Intelligent Artifacts**, AI for Health and Healthy AI conference  
CHALMERS & Gothenburg University, Gothenburg 2019 08 28

**Ethics in the Design of Intelligent Artifacts**, NTNU Trondheim, Norway, 2019 06 19  
The Big Challenges Science Festival. Code of Ethics.

**Avoiding the Intrinsic Unfairness of the Trolley Problem**: Towards Technically and Socially Informed Ethical Guidelines for Designing of Self-driving Cars. 2019 03 15 @NTNU, Trondheim

## 2018

**ETHICS** lecture 2018 08 31 IDEA League PhD School Engineering Complex Systems with Big data and Information Technology  
ECS-BIT'18, Chalmers Gothenburg

Seminar on **ETHICS**, FORA Fog Computing for Robotics and Industrial Automation PhD Summer School 2018,  
June 08, Vienna

For the complete list, see: <http://www.gordana.se/work/presentations.html>

# Några perspektiv på AI som kunskap, teknik och samhällsligt fenomen

1. **KOMPLEXITET** – ekologiskt tänkande – kunskapsekologier, grupp- och samhällsekologier, fysiska ekologier.  
Kopplar till hållbart digitalt samhälle
2. **TIDSPERSPEKTIV** – skillnad mellan korttids- och långtids- perspektiver på AI teknologikutveckling
3. **KONKRETA FRÅGOR** i fokus i dag:
  - Hur jämlikt eller ojämnt ska AI-samhället vara?
  - Kan man bygga in en etisk kompass i en artificiell intelligens?
  - Går det att koka ned den komplexa dynamik som etik och moral utgör till fem-sex principer för etisk AI?
  - Kan AI göras fördomsfri i motsats till människans fördomsfullhet?
  - Vad är egentligen en människa?

Att gräva där man står...  
börjar vi med ansvarsfull forskning

Ansvarsfull forskning omfattar inte bara:

- forskaretik (kvalitetskrav på själva forskningsarbete),
- forskningsetik (relationen till involverade subjekt) utan även
- forskningens påverkan på vårt samhälle med dess materiella och kulturella miljö dvs. hållbar existentiell, social och fysisk miljö. Den glöms bort väldigt ofta. " Prata om att "Sila mygg och svälja kameler."

Exempel:

Forskning om minnesförstärkning och manipulering av minnen

Nytta: kan hjälpa människor med minnesförlust och minnesstörningar

Risk: öppnar för manipulering av minne hos människan

# Etik och ansvarsfull forskning

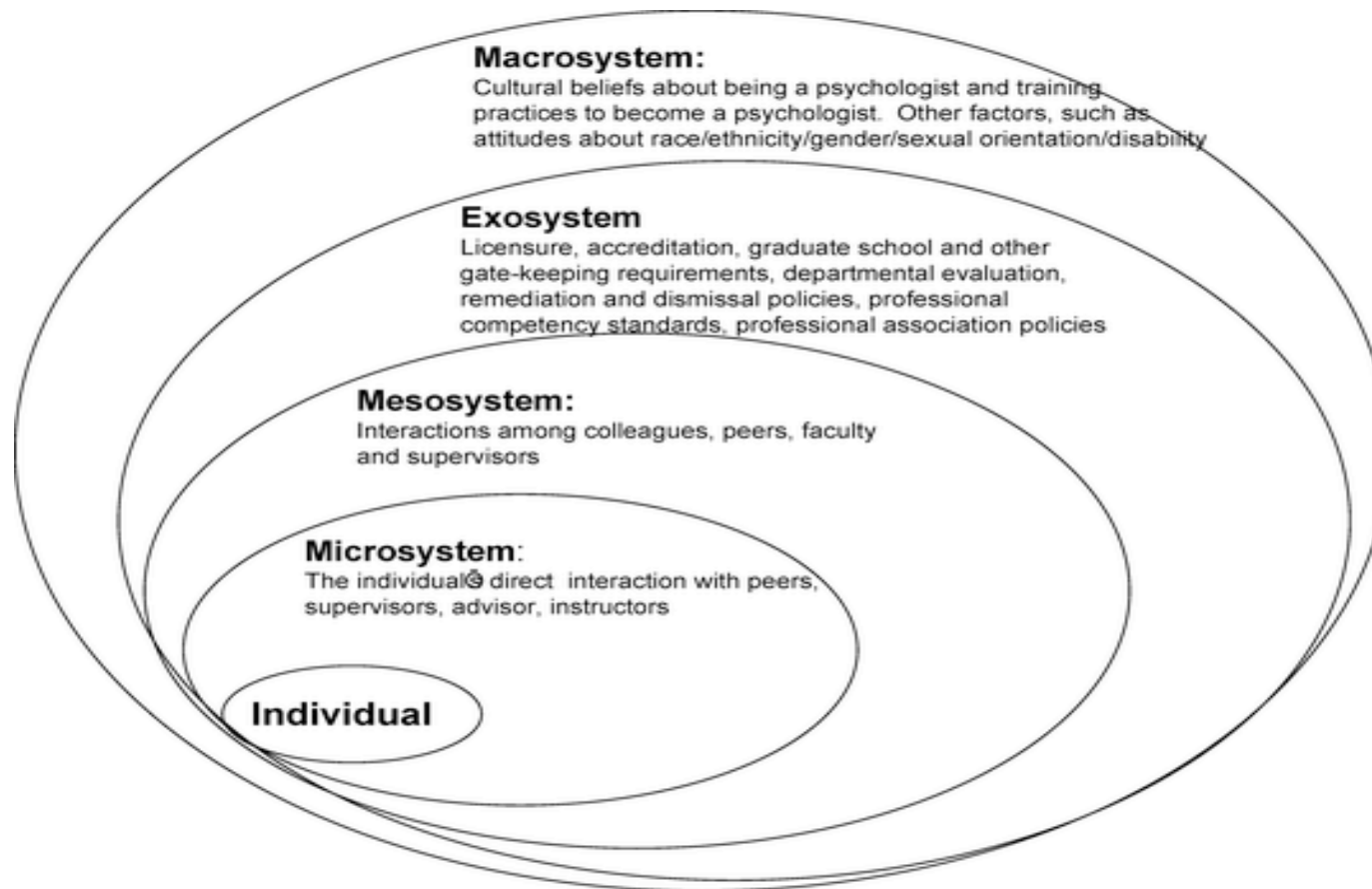


**Responsible Research & Innovation (RRI)** is an approach which implies that societal actors, such as researchers, citizens, policy makers, companies and civil society organisations) work together in the whole research and innovation process in order to better align both the process and its outcomes with the values, needs and expectations of society.

<http://www.sisnetwork.eu/rri/> Researchers working on responsible AI – Virginia Dignum, Jeroen van den Hoven

# DEN STORA BILDEN – ETIKENS OLIKA DOMÄNER

## ETISK KUNSKAP-EKOLOGY (SYSTEMVY) BASERAD PÅ INFORMATION



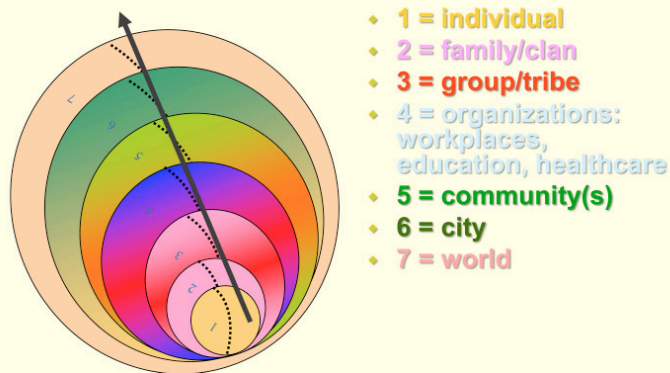
Domains of Ethics: Micro – Meso – Exo – Macro

Source: American Psychological Association website. Similar scheme in Ecology - The nested levels of the Ecology Theory (Bronfenbrenner (1977))

# KOMPLEXITETEN I MÄNSKLIGA RELATIONER

Micro – Meso – Exo – Macro Levels of Analysis – Example of City

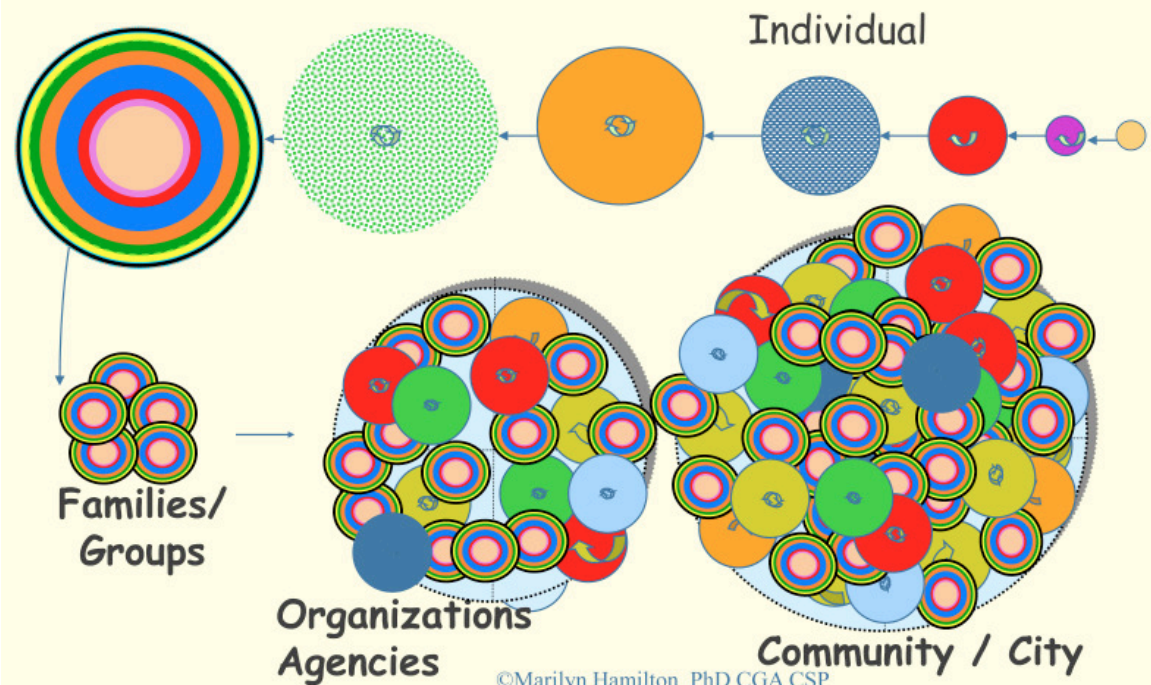
**Map 2: The Nested Holarchy of City Systems**



©Marilyn Hamilton PhD CGA CSP

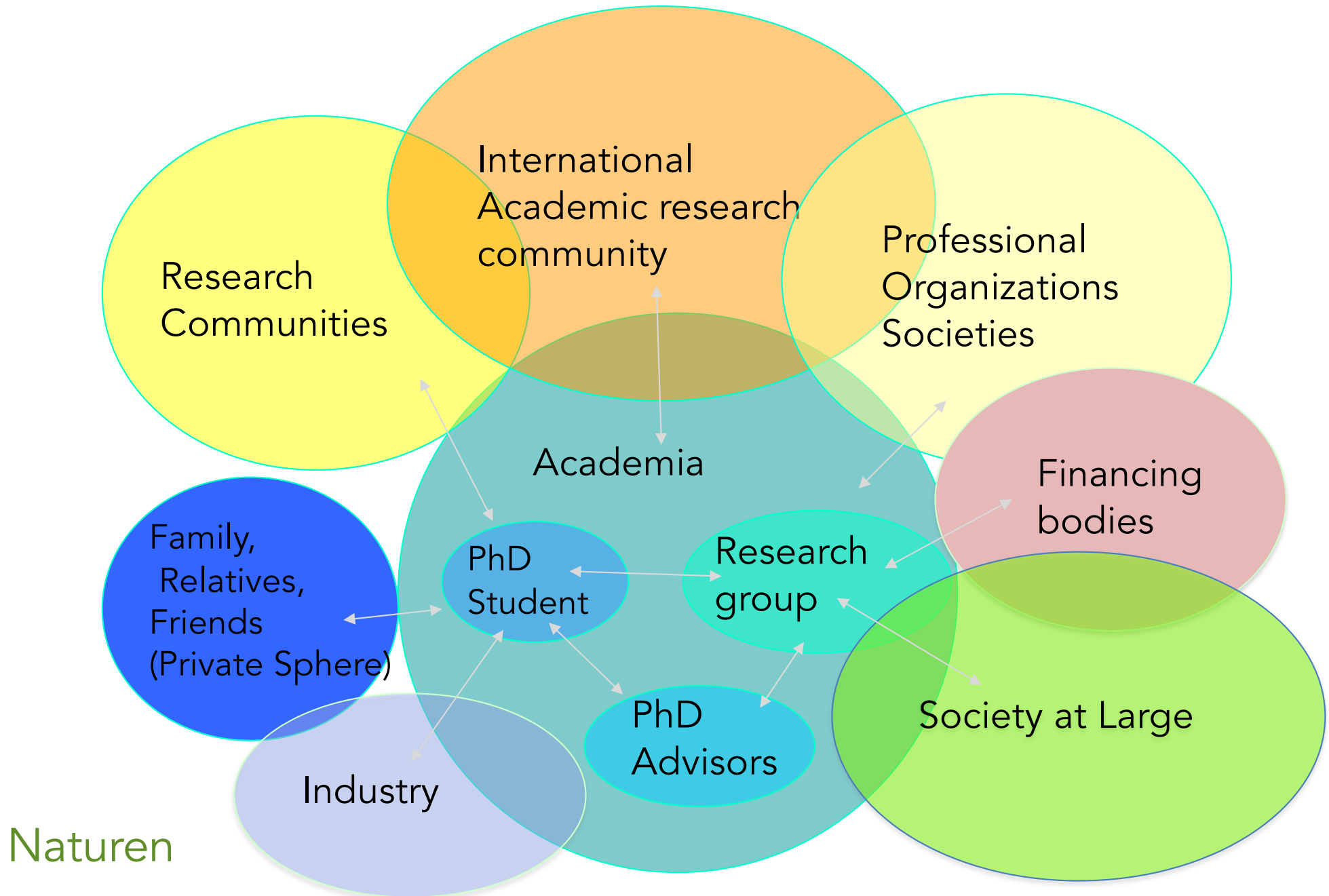
A holarchy, in the terminology of Arthur Koestler, is a connection between holons, where a holon is both a part and a whole. The term was coined in Koestler's 1967 book *The Ghost in the Machine*.

**Map 3: The Scalar Fractal Relationship of Micro, Meso, and Macro Human Systems**



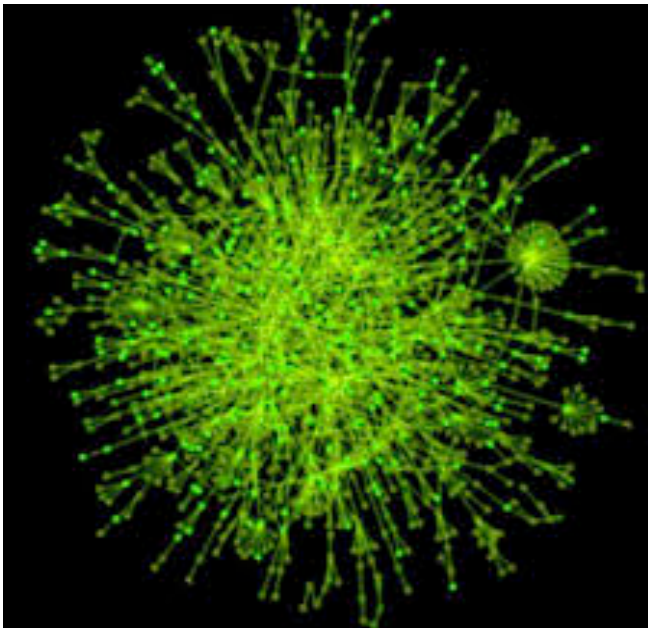
# ETIKEN HANDLAR OM RELATIONER/KOMMUNIKATION

STAKEHOLDERS I ETT AKADEMISKT PROJEKT

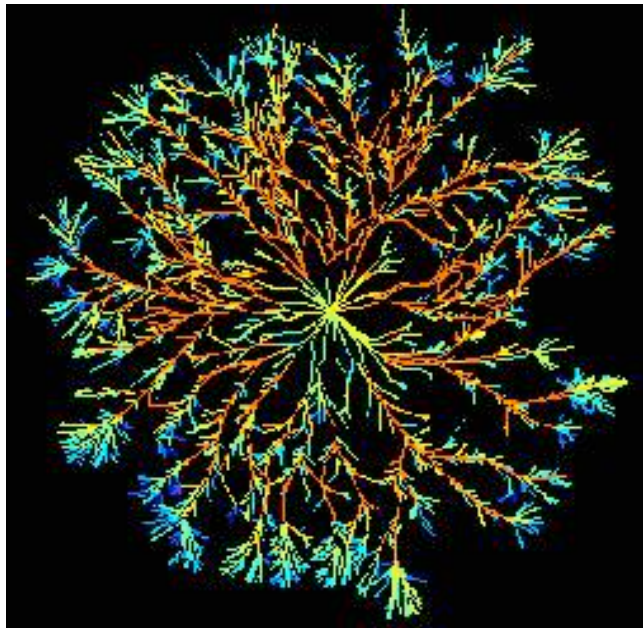


# NÄTVERK av nätverk av informationsprocesser

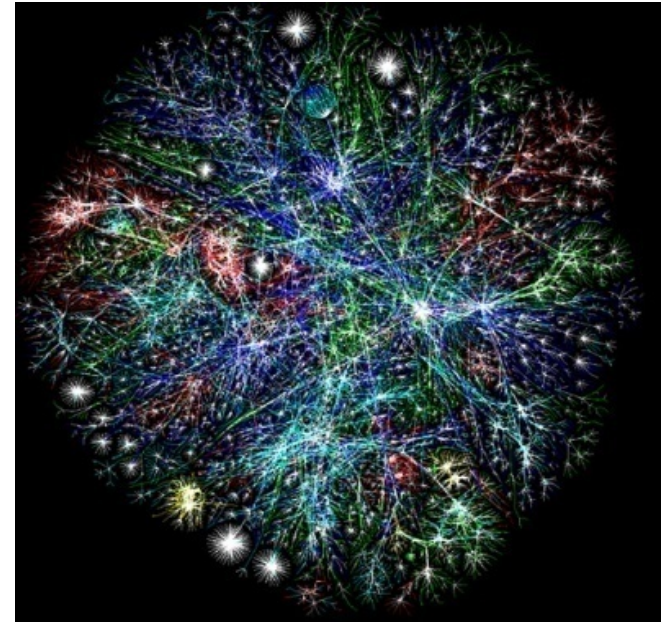
Self-similarity on different levels of organisation



Jäst proteinnätverk



Bakteriekolloni nätverk



Internet

<https://fineartamerica.com/featured/yeast-protein-interaction-map-hawoong-jeong-university-of-notre-dame.html> ;  
[https://en.wikipedia.org/wiki/Opte\\_Project](https://en.wikipedia.org/wiki/Opte_Project)  
[https://en.wikipedia.org/wiki/Ben-Jacob%27s\\_bacteria](https://en.wikipedia.org/wiki/Ben-Jacob%27s_bacteria)

# TIDSPERSPEKTIV

Skillnaden mellan korttids- och långtids- perspektiv på AI [utveckling](#)

**Korttidsperspektiv** (jordnära, omedelbara) – teknologier som redan finns och ändrar vårt samhälle i grunden:

- World Wide Webb (man utvecklar "Internet of Things")
- Sociala media (medför frågan om [demokratins framtid](#), rollen av stora IT företag (Google, Apple, Facebook, Microsoft, Amazon - Big Tech)
- Införande av AI inom statliga institutioner – som ersätter handläggare, jurister, chefer och andra anställda.  
Krav på transparens. vi vill förstå vad besluten grundas på. Fairness.
- Framtidens arbete – kan vi leva utan att gå till jobbet? Hur ska resurser fördelas som skapas av AI?

Fågor för dagens seminarium:

- [Går det att koka ned den komplexa dynamik som etik och moral utgör till fem-sex principer för etisk AI?](#)

Vi är vana vid korta listor med regler som både religiösa och sekulära organisationer brukar ha. Sen återstår i praktiken att tolka vad dessa betyder.

- [Kan AI göras fördomsfri i motsats till människans fördomsfullhet?](#)

Det finns två sorters fördomar: kulturella och individuella. AI kan göras fri från individuella fördomar, men inte från kulturella. Den lär sig från exempel.

# TIDSPERSPEKTIV

**Långtidsperspektiv – AGI** (Artificiell Generell Intelligens – på samma nivå som människa eller högre). Kan vi samexistera?

Kopplar till frågan:

Vad är egentligen en människa?

Den frågan utforskar vi bl.a. genom att bygga artificiella kognitiva intelligenta system.

Hur jämlik ska AI vara? Svag AI och stark AI har olika status.

Kan man bygga in en etisk kompass i en artificiell intelligens?

Kan man **uppfostra** en AI (kanske med hjälp av "deep learning")?

Problemet: ingen AI hittills förstår kausalitet och har "sunt förnuft". Då är det svårt för AI som de är nu att vara etisk.

# Vad står i centrum? Vems intressen styr?

## En fråga om värderingar

Innovation – företag – produkter – tjänster  
(USA modellen)

Staten – "top-down" kontrollen – stora bilden dominerar (Kina modellen)

Människan – integriteten – rättigheter/skyldigheter – samförstånd mellan stakeholders (Europa modellen)

Ett hållbart samhälle – existentiellt, social, fysiskt, tekniskt - digital hållbart

Viktigt att synliggöra: värderingar & instrument som implementerar dessa (juridik, principer, guidelines, utbildning, kultur)