

Att undervisa artificiell intelligens på gymnasial nivå - samhällsperspektiv
OAU916 AU916 HT2024, Mälardalens universitet, 2025 03 13

Etik och sociala aspekter av AI för gymnasielärare

Gordana Dodig-Crnkovic
Mälardalens Universitet
Chalmers Tekniska Högskola

<https://gordana.se/>
<http://www.gordana.se/work/presentations.html>
<http://www.gordana.se/work/publications.html>

Gordana Dodig-Crnkovic, affiliations



School of Innovation, Design and Engineering

Division of Computer Science and Software Engineering

Research groups:
Artificial Intelligence and Intelligent Systems
Ubiquitous Computing



Department of Computer Science and Engineering

Division:
Computer Science and Software Engineering

Research groups:
Interaction Design and Software Engineering
Critical Robotics



Min bakgrund – från formella till naturliga språk

My background - from formal to natural languages

Thus we have

$$\begin{aligned}
 B = & \sum_{J_C M_{L_C}} (-1)^{\lambda_\nu + \lambda_\pi + L_C} \delta(J_\nu, \lambda_\nu) \delta(J_\pi, \lambda_\pi) (L_C M_{L_C} 00 | J_C M_{J_C}) \\
 & \times \sum_{L_C M_{L_C}} ((L_\nu L_\pi) \lambda_\nu (L_\pi L_\nu) \lambda_\pi; L_C | (L_\nu L_\pi) l_C (L_\nu L_\pi) L_C; L_C) \quad (54) \\
 & \times \langle l_C m_{l_C} | L_C M_{L_C} | L_C M_{L_C} \rangle (Y_{l_\nu} Y_{l_\pi})_{l_C} (Y_{L_\nu} Y_{L_\pi})_{L_C} (X^{S_\nu=0} X^{S_\pi=0})_{S_C=0}.
 \end{aligned}$$

The whole expression for A may be thereafter written as

$$\begin{aligned}
 A = & \sum_{J_C M_{L_C}} (-1)^{\lambda_\nu + \lambda_\pi + L_C} \delta(J_\nu, \lambda_\nu) \delta(J_\pi, \lambda_\pi) (L_C M_{L_C} 00 | J_C M_{J_C}) \\
 & \times \sum_{L_C M_{L_C}} ((L_\nu L_\pi) \lambda_\nu (L_\pi L_\nu) \lambda_\pi; L_C | (L_\nu L_\pi) l_C (L_\nu L_\pi) L_C; L_C) \quad (55) \\
 & \times \langle l_C m_{l_C} | L_C M_{L_C} | L_C M_{L_C} \rangle (Y_{l_\nu} Y_{l_\pi})_{l_C} (Y_{L_\nu} Y_{L_\pi})_{L_C} \\
 & \times (X^{S_\nu=0} X^{S_\pi=0})_{S_C=0} R_{n_\nu L_\nu} R_{n_\pi L_\pi} R_{n_\nu L_\nu} R_{n_\pi L_\pi}.
 \end{aligned}$$

After Moshinsky-Talmi transformation $(N_\nu L_\nu; N_\pi L_\pi) \rightarrow (n_C l_C; N_C L_C)$ it reads

$$\begin{aligned}
 A = & \sum_{J_C M_{L_C}} (-1)^{\lambda_\nu + \lambda_\pi + L_C} \delta(J_\nu, \lambda_\nu) \delta(J_\pi, \lambda_\pi) (L_C M_{L_C} 00 | J_C M_{J_C}) \\
 & \times \sum_{L_C M_{L_C}} ((L_\nu L_\pi) \lambda_\nu (L_\pi L_\nu) \lambda_\pi; L_C | (L_\nu L_\pi) l_C (L_\nu L_\pi) L_C; L_C) \quad (56) \\
 & \times \langle l_C m_{l_C} | L_C M_{L_C} | L_C M_{L_C} \rangle (Y_{l_\nu} Y_{l_\pi})_{l_C} R_{n_\nu L_\nu} R_{n_\pi L_\pi} (X^{S_\nu=0} X^{S_\pi=0})_{S_C=0} \\
 & \times \sum_{n_C l_C} (n_C l_C; N_C L_C; J_C | N_\nu L_\nu N_\pi L_\pi; J_C) (Y_{l_\nu} Y_{l_\pi})_{l_C} R_{n_C l_C} R_{n_C l_C}.
 \end{aligned}$$

29



PhD in Computing, 2006
Computer Science,
Mälardalen University



Current: Morphological
Cognitive and Intelligent
Computing, AI Ethics, Digital
Ethics, Digital Humanism

<https://www.gordana.se/work/courses.html>

<https://tinyurl.com/34r7xyw>

Gordana Dodig-Crnkovic, Experiences from work with Ethics

- Member of the Board of Informatics Europe with responsibility for Ethical and Social issues <https://www.informatics-europe.org/>
Informatics Europe has over 160 member institutions across 33 countries.
Co-organized several workshops on Teaching ethics and Societal Impact of Digital Transformation. Lead Working Group on Ethics.
- Chalmers AI Ethics Committee member
- Member of the network ETHICS4EU- facilitating introduction of Ethics to IT students in Europe
- Member of the network INCLUSION4EU who works on Ethics of inclusion in Informatics
- Teaching “Professional Ethics” at Mälardalen University 2002-2016 to undergraduate and graduate students and “Research Ethics & Sustainable Development” – a Ph.D. Course at Chalmers University of Technology (2014-2017)
- Current topics of research interest: Ethics of autonomous cars, and Robotic ethics, Design Ethics, AI Ethics.

Vad är AI

Etiska riktlinjer för lärare avseende användningen av artificiell intelligens (AI) och data vid undervisning och inläring
<https://tinyurl.com/yjfhfobb>



Användningen av artificiell intelligens och data

Vad är artificiell intelligens?

I hela Europa använder elever och lärare i allt högre grad system för artificiell intelligens (AI) men inser det inte alltid själva. Sökmotorer, smarta assistenter, chattbotar, språköversättning, kartappar, onlinevideospel och många andra applikationer som vi använder i vår vardag utnyttjar artificiell intelligens. AI-systemen bygger på data som samlas in på olika sätt (t.ex. genom ljud, bilder, text, inlägg, klickar) och tillsammans bildar våra digitala spår.

AI har stor potential att förbättra utbildningen från såväl elevernas som läraarnas och skolledarnas perspektiv. Redan i dag hjälper AI-system dels en del lärare att upptäcka specifika inlärningsbehov och ge studerande individanpassade lärandeupplevelser, dels en del skolor att fatta bättre beslut så att de kan använda sina tillgängliga undervisningsresurser mer effektivt.

AI-systemen utvecklas hela tiden och dataanvändningen ökar och det är därför oerhört viktigt att vi får en bättre förståelse av hur världen runt oss påverkas, särskilt på utbildningsområdet. Om teknikens fulla potential ska kunna utnyttjas måste lärare och skolledare ha åtminstone grundläggande kunskaper om AI och dataanvändning så att de kan inta ett positivt, kritiskt och etiskt förhållningssätt till denna teknik och använda den korrekt.

Den definition av ett system för artificiell intelligens (AI-system) som föreslås i utkastet till AI-akten är "programvara som utvecklats med en eller flera [...] (nedan förtecknade) tekniker och metoder [...] och som, för en viss uppsättning människodefinierade mål, kan generera utdata såsom innehåll, förutsägelser, rekommendationer eller beslut som påverkar de miljöer som de samverkar med".

De AI-tekniker och AI-metoder som förtecknas är följande:

- Metoder för maskininläring, inbegripet övervakad, övervakad och förstärkt inläring, med hjälp av en mängd olika tillvägagångssätt, inklusive djupinläring.
- Logik- och kunskapsbaserade metoder, inbegripet kunskapsrepresentation, induktiv (logisk) programmering, kunskapsbaser, inferens- och deduktionsmotorer, (symboliska) resonemang och expertsystem.
- Statistiska metoder, bayesisk beräkning, sök- och optimeringsmetoder.

När vi talar om AI-system menar vi en programvara i datorer eller maskiner som programmerats att utföra uppgifter som vanligtvis kräver mänsklig intelligens, t.ex. inläring eller resonerande. Med hjälp av data kan vissa AI-system "tränas" att göra förutsägelser, utfärda rekommendationer eller fatta beslut, ibland utan mänsklig medverkan.

Handlingsplanen för digital utbildning (2021–2027) innehåller två strategiska prioriteringar:

1 För att främja utvecklingen av ett högpresterande digitalt utbildnings system behöver vi

- **infrastruktur, uppkopplingar och digital utrustning**,
- **ändamålsenlig planering och utveckling av den digitala kapaciteten**, bland annat effektiv och modern organisatorisk kapacitet,
- **lärare och annan utbildningspersonal med digital kompetens** som **känner sig trygga i undervisningen**,
- **utbildningsinnehåll av hög kvalitet, användarvänliga verktyg och säkra plattformar** som respekterar etiska normer och integritetsregler på nätet.

För att förbättra den digitala kompetensen och de digitala färdigheterna för den digitala eran krävs det att man

- **stöder utvecklingen av grundläggande digitala färdigheter och kompetenser** från tidig ålder. Detta omfattar
 - digital kompetens, bland annat om hur man hanterar informationsöverflöd och känner igen desinformation,
 - datorutbildning,
 - goda kunskaper och bra förståelse av datadriven teknik, t.ex. artificiell intelligens (AI).
- **främjar avancerade digitala färdigheter** för att fler ska specialisera sig på digital teknik och fler flickor och kvinnor ska välja studier och yrken på det digitala området.

Vanliga missuppfattningar om AI

AI är för svårt att förstå

AI har ingen plats i utbildningen

AI är inte inkluderande

Man kan inte lita på AI-system

AI kommer att försvaga lärarens roll

Användning av AI och data i utbildningen

ELEVUNDERVISNING

AI används i elevundervisningen

Intelligent handledningssystem	Eleven gör ett antal på varandra följande uppgifter och får individualiserad undervisning eller återkoppling utan att läraren behöver ingripa.
Dialogbaserade handledningssystem	Eleven gör ett antal på varandra följande uppgifter i form av samtal på naturligt språk. Mer avancerade system kan automatiskt anpassas till graden av nredverkan, för att eleven ska fortsätta vara motiverad och fokuserad på uppgiften.
Språkappar	AI-baserade inlärningsappar används i såväl formella som icke-formella utbildningssammanhang. De stöder inläringen genom att de erbjuder språkkurser och lexikon och ger direkt återkoppling på uttalet, förståelsen och flytet.

ELEVSTÖD

AI används för att stödja elevernas inläring

Miljöer för utforskande inläring	Elever erbjuds flera olika vägar att gå som hjälper dem att hitta sina egna sätt att uppnå lärandemålen.
Formativ skrivbedömning	Elever får regelbunden automatisk återkoppling på hur de skriver/sina inlärningsuppgifter.
AI-stödd samarbetsbaserad inläring	Data om varje elevs arbetssätt och tidigare resultat används till att dela in dem i grupper med samma kompetensnivå eller en lämplig blandning av färdigheter och talanger. AI-systemen ger information om/förslag på hur en grupp arbetar/kan arbeta tillsammans genom att övervaka graden av interaktion mellan gruppmedlemmarna.

LÄRARSTÖD

AI används för att stödja läraren

Summativ skrivbedömning, poängsättning på uppsatser	AI används till att automatiskt utvärdera och betygsätta elevers skriftliga arbete. AI- och maskininlärningsmetoder identifierar egenskaper såsom: ordanvändning, grammatik och meningsbyggnad för att sätta betyg och ge återkoppling.
Övervakning av elevforum	Vissa ord i inlägg på elevforum utlöser automatisk återkoppling. Diskussionsanalyser ger inblick i elevernas aktivitet på forumet och kan visa vilka elever som kan behöva hjälp eller som inte deltar som förväntat.
AI-läroassistenter	AI-agenter eller chattbotar svarar på elevernas vanliga frågor och ger enklare instruktioner och anvisningar. Med tiden kommer AI-systemet att kunna erbjuda allt fler svar och alternativ.
Rekommendationer om pedagogiska resurser	AI-rekommendationsmotorer kan rekommendera särskild undervisning eller särskilda resurser utifrån varje elevs preferenser, framtida behov.

STÖD PÅ SYSTEMNIVÅ

AI används för att stödja diagnostisering eller systemomfattande planering

Utvinning av utbildningsdata för resursfördelningsarbetet	Skolorna samlar in elevdata som analyseras och används vid planeringen av hur de tillgängliga resurserna bäst kan fördelas vid arbetet med att skapa klassgrupper, fördela lärarresurserna och lägga scheman, och för att lyfta fram elever som kan behöva ytterligare inlärningsstöd.
Diagnostisering av inlärningssvårigheter	Med hjälp av inlärningsanalyser mäts kognitiva förmågor såsom ordförråd, hörsel, rumsligt tänkande, problemlösning och minne i syfte att diagnostisera inlärningssvårigheter, även utdaggande problem som kan vara svåra för en lärare att se men som skulle kunna upptäckas tidigt med hjälp av AI-system.
Vägledningstjänster	AI-baserade vägledningstjänster ger kontinuerliga förslag på eller valmöjligheter för att bana vägen mot framtida utbildning. Användarna kan skapa en kompetensprofil som omfattar deras tidigare utbildning och lägga till egna intressen. Utifrån dessa data, och i kombination med aktuell kurskatalog eller information om studiemöjligheter, kan relevanta studierekommendationer skapas genom bearbetning av naturligt språk.

Förord

DEL 1: HUR DET BÖRJADE

Kapitel 1: Lite noggrant utvald historik

Världens mest lyckade bedrägeri
Algoritmer
Datorer
Gödel, Turing och von Neumann
Ett ämnesområde föds
Schack matt
Stutledningens bananfluga
Universums bästa schackspelare
Maskinernas tidsålder

Kapitel 2: Hur vi försöker göra maskiner intelligenta

Symbolisk AI
Expertsystem
ELIZA-effekten
Vinter och vår
Maskiner som lär sig
Väglett lärande
Blommor och beslutsträd
Subsymbolisk AI
Neurala nätverk genom tiderna
Att lära vad som helst

Kapitel 3: Jakten på data

Den sanna fördelningen
Data till maskininläring
Fel fördelning
Bias för statistiker och journalister
Data är dyrt
Långa problematiska svansar
Att hitta sina egna data

DEL 2: ARTIFICIELL INTELLIGENS IDAG

Kapitel 4: Du tror det inte förrän du ser det

Världen är matriser
Maskiner som nästan förstår vad de tittar på
Det enkla är ofta fel
Maskiner som förstår mer än vi gör
Artificiell intuition och schackkoncept
Mental hälsa
Dataintegritet
Rätten till en förklaring

Kapitel 5: Är det någon hemma?

Artificiell samtalspartner
Anden i maskinen
ChatGPT och bekymrade norsklärare
Prat om problem
Ett Turingtest för medvetenhet
Konsten att skapa
Från brus till verklighet

Kapitel 6: Vårt artificiellt intelligenta liv

Maskinrevolutioner

En etisk värld
Någon måste bestämma
Elkrig och AI Act
Det sociala dilemmat med AI-utveckling
Vad ska vi ha dataintegritet till?

Kapitel 7: Hur man attackerar en maskininlärningsmodell

Smarta maskiner med goda intentioner
Synvillor för maskiner
Att klippa gräs i tre dimensioner
Dimensionsförbannelsen
In genom bakvägen
Vinter och politik
Autonomi och kontroll

INGA STRÖMKE

MASKINER SOM TÄNKER

Algoritmernas hemligheter och
vägen till artificiell intelligens

Översättning av Leo Gefver

BOKFÖRLAGET
POLARIS **P**

DEL 3: MORGONDAGENS ARTIFICIELLA INTELLIGENS

Kapitel 8: Superintelligens och andra spekulationer

Artificiell kreativitet
Tal och akustik
Dataanvändning
AI-etik och moraliska aktörer
Generell intelligens och superintelligens
Maskinmål, människomål och mesa-optimering
En hjärna som tänker på sig själv

Illustrationer

Noter

Teknik och etik

Sven Ove Hansson

Kungl Tekniska Högskolan, Stockholm

2009

Innehåll

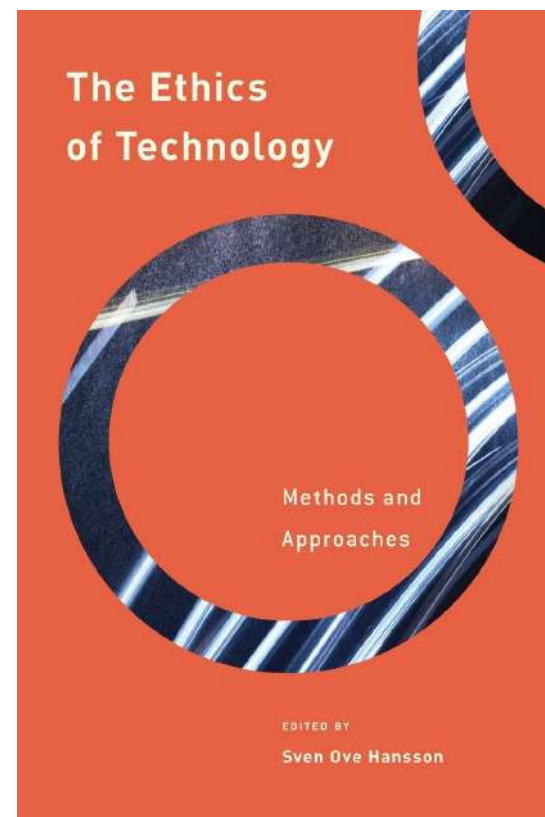
Förord	7
1 Teknikens etiska problem	9
1.1 Är tekniken god, ond eller neutral?	9
1.2 Två exempel på tekniketiska problem	11
Teknik och arbetsinnehåll	12
Teknikens minne	16
1.3 Är teknikutvecklingen oundviklig?	17
Inläsningseffekten	18
Kulturell eftersläpning	19
1.4 Teknikmotstånd och politisk styrning	21
Politiskt teknikmotstånd	21
Kommersiellt teknikmotstånd	23
Politisk styrning av tekniken	24
1.5 Övningsuppgifter	25
2 Etiken	29
2.1 Vad är etik?	29
Etik och moral	29
Etik och juridik	31
Fakta och värderingar	32
2.2 Utilitarism	34
Hedonistisk utilitarism	34

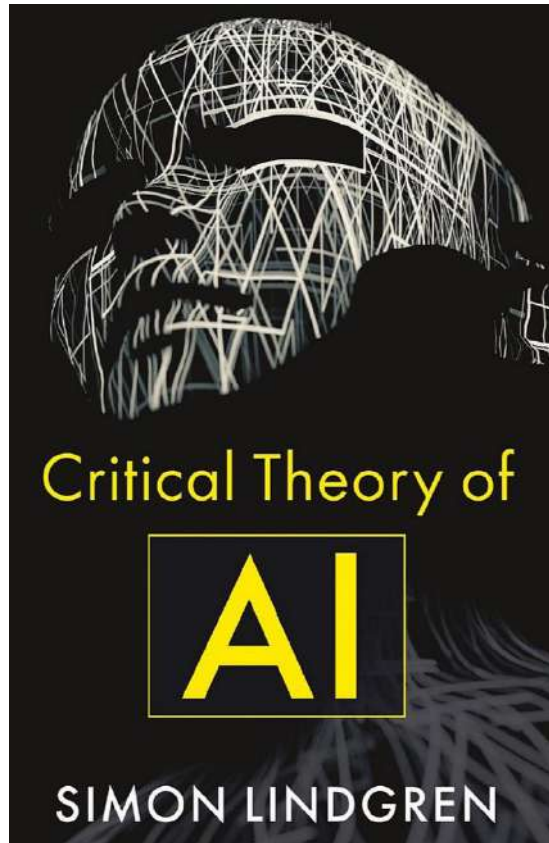
4

Icke-hedonistisk utilitarism	35
En jämförelse	36
Konsekventialism	38
En opersonlig etik	39
Regelutilitarism	40
Utilitaristiska kalkyler	42
2.3 Plikter och pliktetik	44
Språkliga uttryck för plikter	44
Pliktordens mångtydighet	46
Förbud och tillåtelser	47
Prima facie-plikter	49
Restplikter	50
Moraliska dilemman	51
Pliktetik eller utilitarism?	52
2.4 Frihet	56
Frihet och utilitarism	56
Frihet och pliktetik	57
Paternalism	58
2.5 Rättigheter	60
Positiva och negativa rättigheter	61
Behövs rättigheter?	61
Mänskliga rättigheter	62
2.6 Dygdetik	63
2.7 Etikens grundvalar	65
Religiös etik	65
Etik på naturens grund	66
Samhällskontraktet	67
Rawls och det hypotetiska samhällskontraktet	68
Kritik mot kontraktsteorin	69
Reflektiv jämvikt	70
Diskursetik	71
2.8 Övningsuppgifter	73

5

3 Ingenjören	75
3.1 Ingenjörnsrollen	75
Från slav till civilingenjör	75
Behövs en yrkesetik?	77
3.2 Ansvar	78
Uppgiftsansvar och skuldansvar	78
Ingenjörens ansvar	79
Hur långt sträcker sig ansvaret?	82
"Annars gör någon annan det"	85
Ingenjörsetik och företagsetik	86
3.3 Lojalitet och lojalitetskonflikter	90
3.4 Koder och eder	92
Ingenjörsetiska koder	93
Ed och legitimation?	96
3.5 Övningsuppgifter	99
4 Etisk teknikvärdering	101
4.1 Teknikvärdering	101
Teknikvärderingens framväxt	101
Förenklade varianter av teknikvärdering	103
4.2 Hur teknikvärderingar görs	105
Teknikvärderingens stadier	105
Metoder för teknikvärdering	111
Att utnyttja expertkunskaper	114
4.3 Etisk teknikvärdering	116
4.4 Övningsuppgifter	121



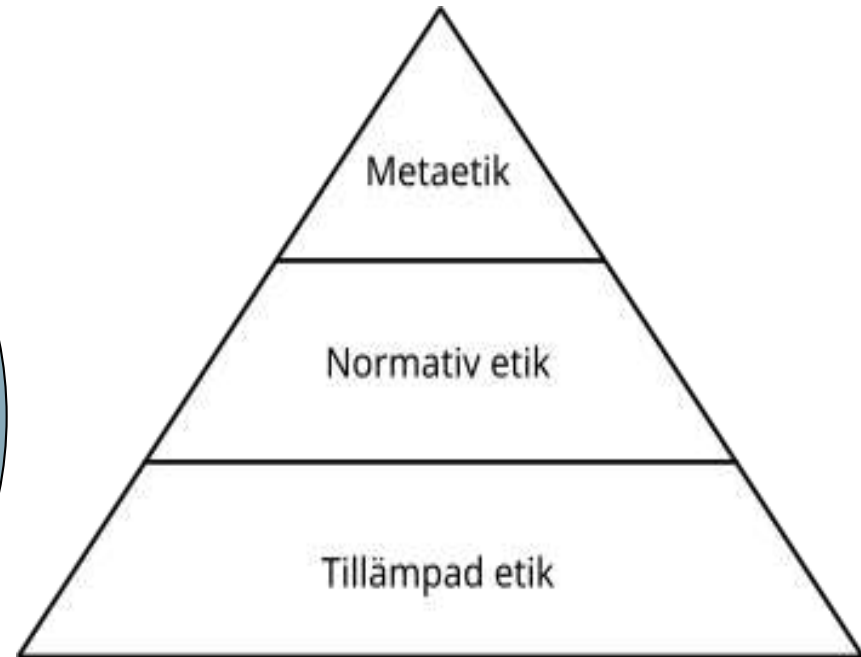
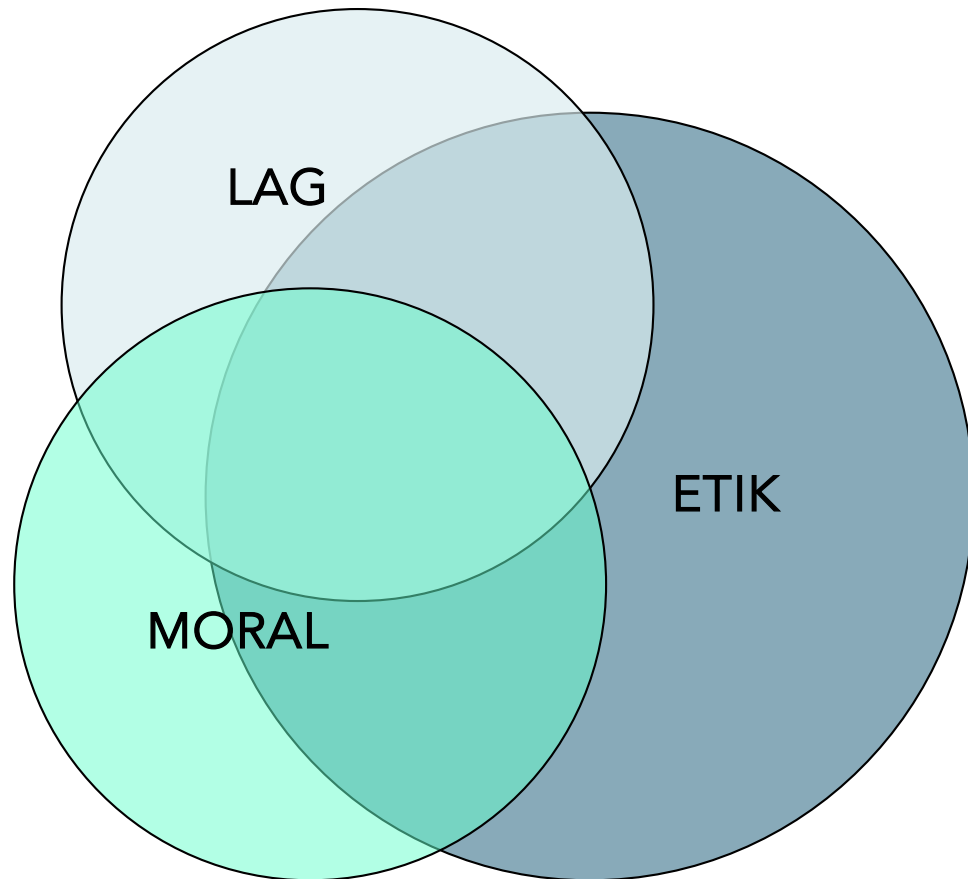


Contents

1 AI and critical theory	1
2 AI assemblage	22
3 Ideology behind AI	43
4 Ideology within AI	65
5 Social machines	82
6 AI at work	103
7 AI subjects	128
8 AI in the loop	153
<i>References</i>	171
<i>Index</i>	202

<https://www.youtube.com/watch?v=MszpMnJyJwU>
What is a social science of AI? Simon Lindgren

Samhällets normativa system



Terminologiskt förtydligande: Etik och moral

Termerna **etik** och **moral** används ofta omväxlande - ja, de kan oftast betyda samma sak, och i vardagliga samtal är det inga problem med att växla mellan det ena och det andra.

Det finns dock en skillnad mellan dem i filosofin!

Moral handlar om vad människor faktiskt gör eller tror är rätt och fel i praktiken.

Etik är den systematiska studien och reflektionen över moraliska principer. Det är en gren av filosofi som utforskar vad som är rätt och fel på ett mer teoretiskt plan.

Etikens kontinuum - Ett pågående samtal

Ethics Continuum - An ongoing conversation

- Världen förändras kontinuerligt, och vi måste tolka/bygga om våra uppfattningar om och om igen.
- Vi kommer tillbaka till idéer gång på gång och hittar ny mening i dem.
- Professionella diskussioner om etiska frågor i tidskrifter.
- The world changes continually, and we have to interpret/construe it over and over again.
- We come back to ideas again and again, finding new meaning in them.
- Professional discussions of ethical issues in journals.

ETIK, en deltagarstyrd och iterativ process

ETHICS, a Participatory and Iterative Process

Ethics involves a participatory and iterative process of ethical reflection, inquiry, and deliberation. Combining action and reflection is crucial.

Etik innebär en deltagande och iterativ process av etisk reflektion, undersökning och överläggning. Att kombinera handling och reflektion är avgörande.

It is instructive to go back and forth between zooming out and zooming in on the problem.

Det är lärorikt att gå fram och tillbaka mellan att zooma ut och zooma in på problemet.

In this process, we consult different ethical approaches (Consequentialism, Duty ethics, Virtue ethics, Relational ethics, etc.)

I denna process använder vi olika etiska förhållningssätt (konsekventialism, pliktetik, dygdetik, relationsetik, etc.)

Dygdetik (Virtue Ethics) är en moralfilosofisk teori som fokuserar på individens karaktär och dygder som grund för att bedöma etiskt handlande

Pliktetik (Duty Ethics) betonar regler och plikter

Methods from Human-Centered Design (HCD) organizing participatory and iterative processes, Value Sensitive Design (VSD), bringing different stakeholder values, and Responsible Innovation (RI) with a focus on inclusion, participation, and diversity.

Metoder från Human-Centered Design (HCD) som organiserar deltagande och iterativa processer, Value Sensitive Design (VSD), som ger olika intressentvärden och Responsible Innovation (RI) med fokus på inkludering, delaktighet och mångfald.



Konsekventialism (Consequentialism) bedömer handlingar utifrån deras konsekvenser.

Relationell etik (Relational Ethics) fokuserar på betydelsen av relationer och kontext i moraliskt handlande.

We face complex, interdisciplinary, and global challenges: climate crisis, political polarization, and inequalities. These are all **wicked problems**, which require diverse disciplines, both to better understand the problem and to envision and create solutions.

Vi står inför komplexa, tvärvetenskapliga och globala utmaningar: klimatkris, politisk polarisering och ojämlikheter. Dessa är alla onda problem, som kräver olika discipliner, både för att bättre förstå problemet och för att föreställa sig och skapa lösningar.

Doing ethics is not always easy or pleasant. It can involve asking uneasy questions, creating awkward situations, and tolerating tension and uncertainty.

Att använda etik är inte alltid lätt eller trevligt. Det kan innebära att ställa obehagliga frågor, skapa besvärliga situationer och tolerera spänningar och osäkerhet.

<https://dl.acm.org/doi/pdf/10.1145/3550069> Marc Steen

AI:s etiska utmaningar

“Ethical guidelines on the use of AI and data in teaching and learning are an incremental process of continuous deliberation and learning.”

Expert Group on AI and data in education and training

Med tanke på att det krävs en stor mängd data för att träna AI-systemen, algoritmerna till sin natur är automatiserade och deras applikationer går att skala upp innebär användningen av AI att viktiga frågor måste diskuteras när det gäller personuppgifter, dataskydd och integritet.

Dessa etiska riktlinjer för användningen av AI och data vid undervisning och inläring har utformats på ett sätt som ska hjälpa lärare att förstå vilken potential användningen av AI-applikationer och data kan ha inom utbildningen och öka medvetenheten om möjliga risker. De kan då inta ett positivt, kritiskt och etiskt förhållningssätt till AI-systemen och förverkliga deras fulla potential.



Transparens och förklarbarhet i AI-system



Bias (fördomar) och rättvisa i AI-algoritmer



Integritet och dataskydd



Ansvarsfrågor vid AI-baserade beslut

Framtiden är inte oundviklig, utan vår att skapa!

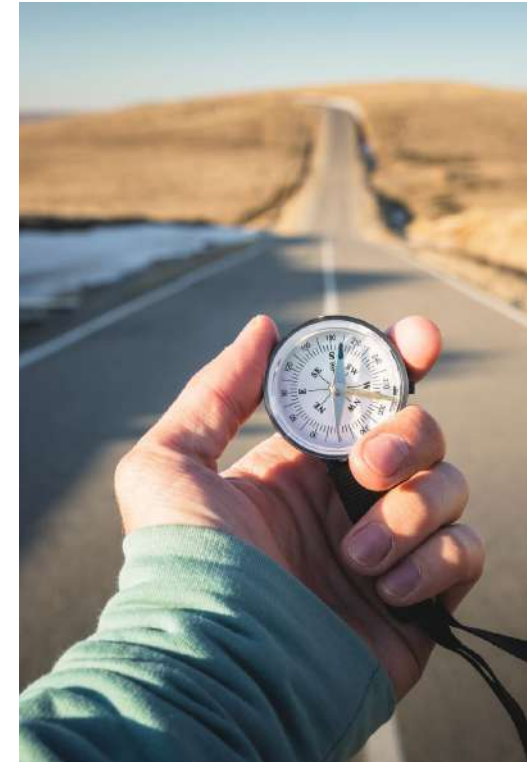
The future is not inevitable, but ours to create!

"AI is not a done deal. We're building the road, as we walk it, and we can collectively decide what direction we want to go in, together."

"AI är inte en färdig affär. Vi bygger vägen, medan vi går den, och vi kan tillsammans bestämma vilken riktning vi vill gå i, tillsammans."

- Sasha Luccioni, 2023

<https://www.nationalacademies.org/news/2024/02/just-how-intelligent-is-artificial-intelligence>



Exempel på snedvridning/partiskhet/bias i AI

AI bias exempel

Ibland tror man att LLMs är skyldiga till att deras resultat är partiska och fördomsfulla. Men de är tränade på stora mängder historiskt data som visar brist på jämlikhet i olika sammanhang. Bias kan komma från både

- data
- algoritmer

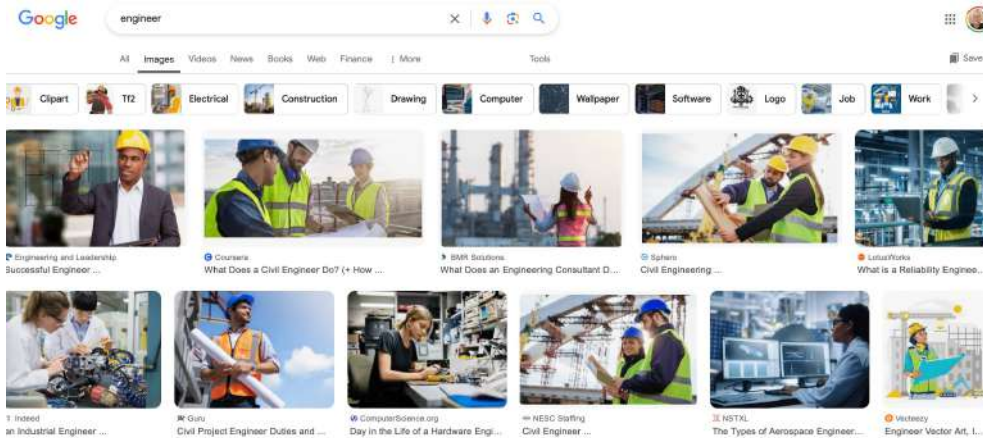
Google har nyligen försökt aktivt medverka diskriminering pga ras och kön.



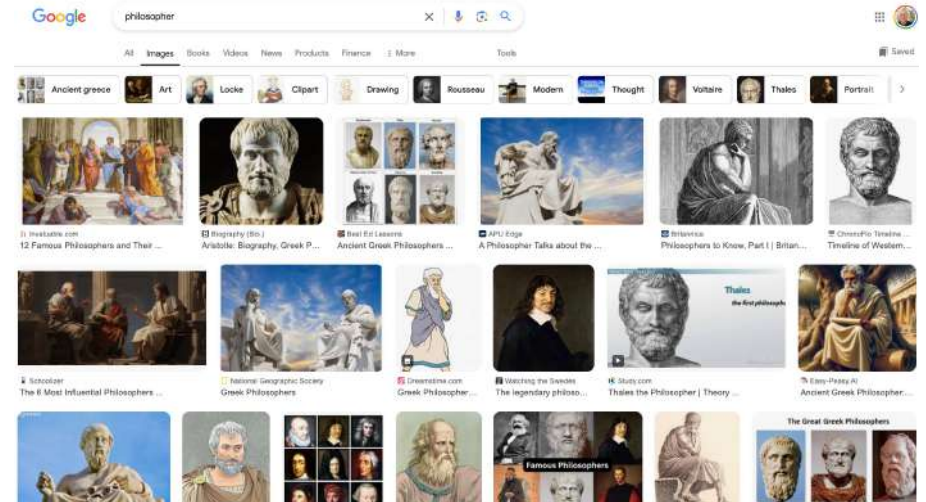
Prompt: En gammal kvinnlig filosof studerar intelligens i levande organismer - celler, djur, växter

Exempel på web snedvridning/partiskhet/bias

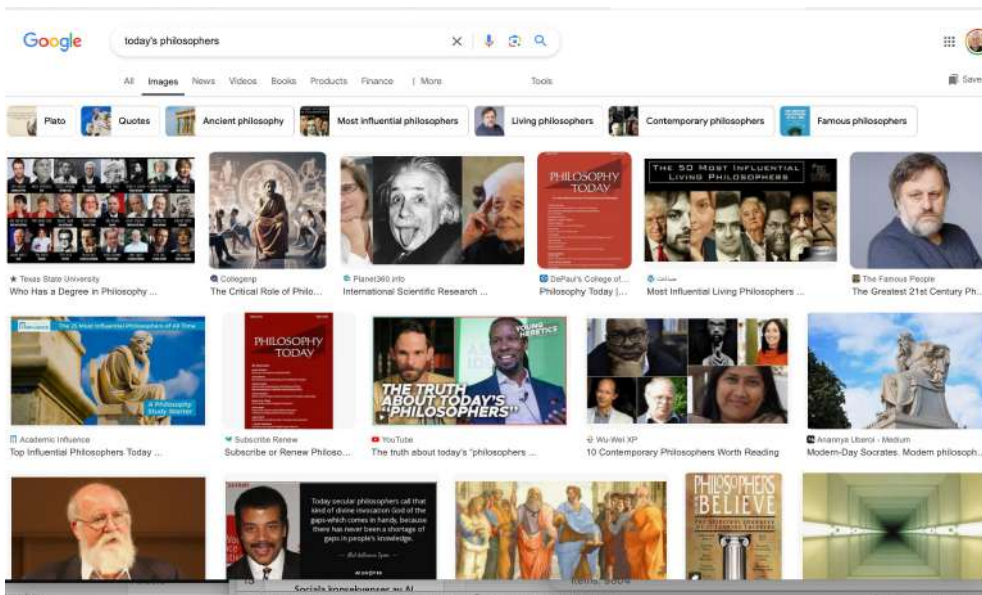
Web bias exempel



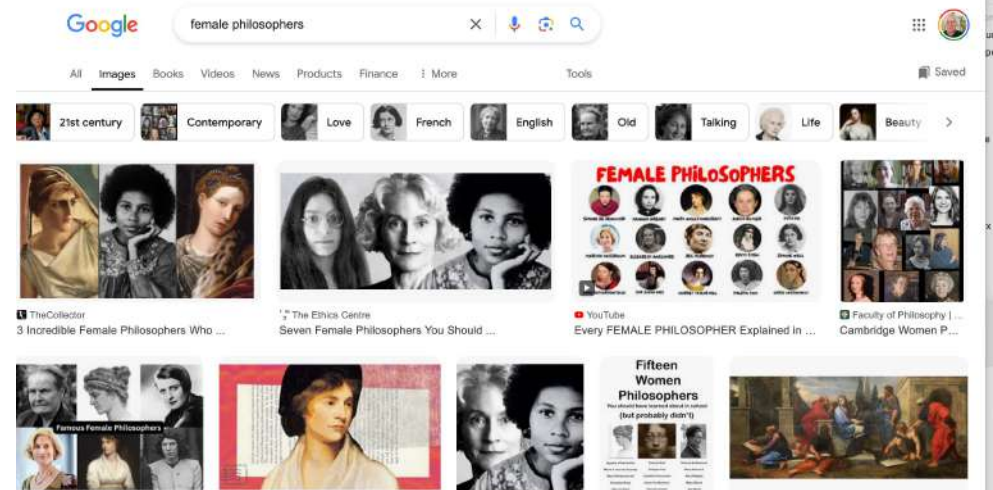
Google sök på "ingenjör": både manliga och kvinnliga bilder



Google sök på "filosof": enbart manliga bilder



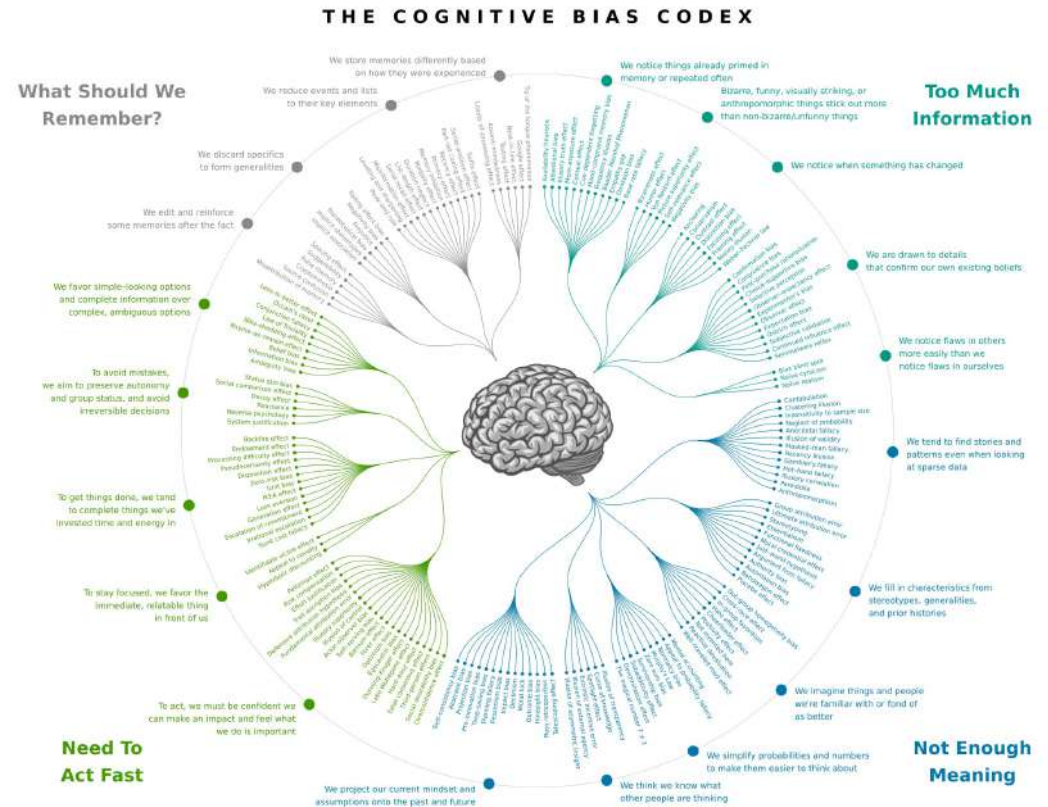
Google sök på "today's philosophers": även kvinnor förekommer



Google sök på "female philosophers" - de finns!

188 Mänskliga kognitiva bias 188 Human cognitive biases

https://en.wikipedia.org/wiki/List_of_cognitive_biases List of cognitive biases



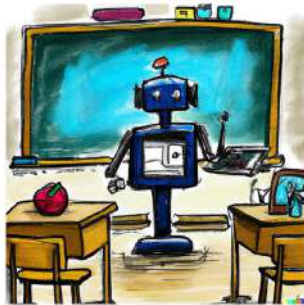
https://en.wikipedia.org/wiki/Cognitive_bias#/media/File:Cognitive_bias_codex_en.svg

Sociala konsekvenser av AI

FUNDACION TATIANA

Proyecto ELAI:
Lecciones éticas de la
inteligencia artificial
Ethical Lessons of Artificial Intelligence

uc3m Universidad Carlos III de Madrid



Alexandra Koch, Pixabay

Gordana Dodig-Crnković
Mälardalen University &
Chalmers University of Technology,
Sweden

**Navigating the White-Water World
with Digital Humanism**

April 12th, 2024



<https://www.youtube.com/watch?v=Mccpg8fpUI8>

Arbetsmarknadens förändring
och automatisering

AI:s påverkan på utbildning
och lärande

Sociala medier och
informationsspridning

AI i vården och etiska
dilemman

AI och samhället



<https://www.dagenssamhalle.se/samhalle-och-valfard/digitalisering/experterna-se-upp-for-farorna-med-ai-stora-olyckor/> Fredrik Heintz, Daniel Gillblad och Magnus Mähring, Experterna: Se upp för farorna med AI

- Farorna med nya tekniken.
- Så kan riskerna förebyggas.
- Då kan AI-modellerna börja hallucinera.

Demokrati och AI: påverkan på beslutsfattande och opinionsbildning

AI och ojämlikhet: risker för ökade klyftor i samhället

Kulturell påverkan och kreativitet

Mina reflektioner

Etiken är ett stort område.
Vad den här föreläsningen kan göra är att öppna fönstret med utsikt



Frågan om värderingar

Alltför ofta utvecklas ny teknik med liten hänsyn till dess inverkan på mänskliga värderingar



VÄRDERINGAR OCH ETIK I KUNSKAPSPRODUKTION

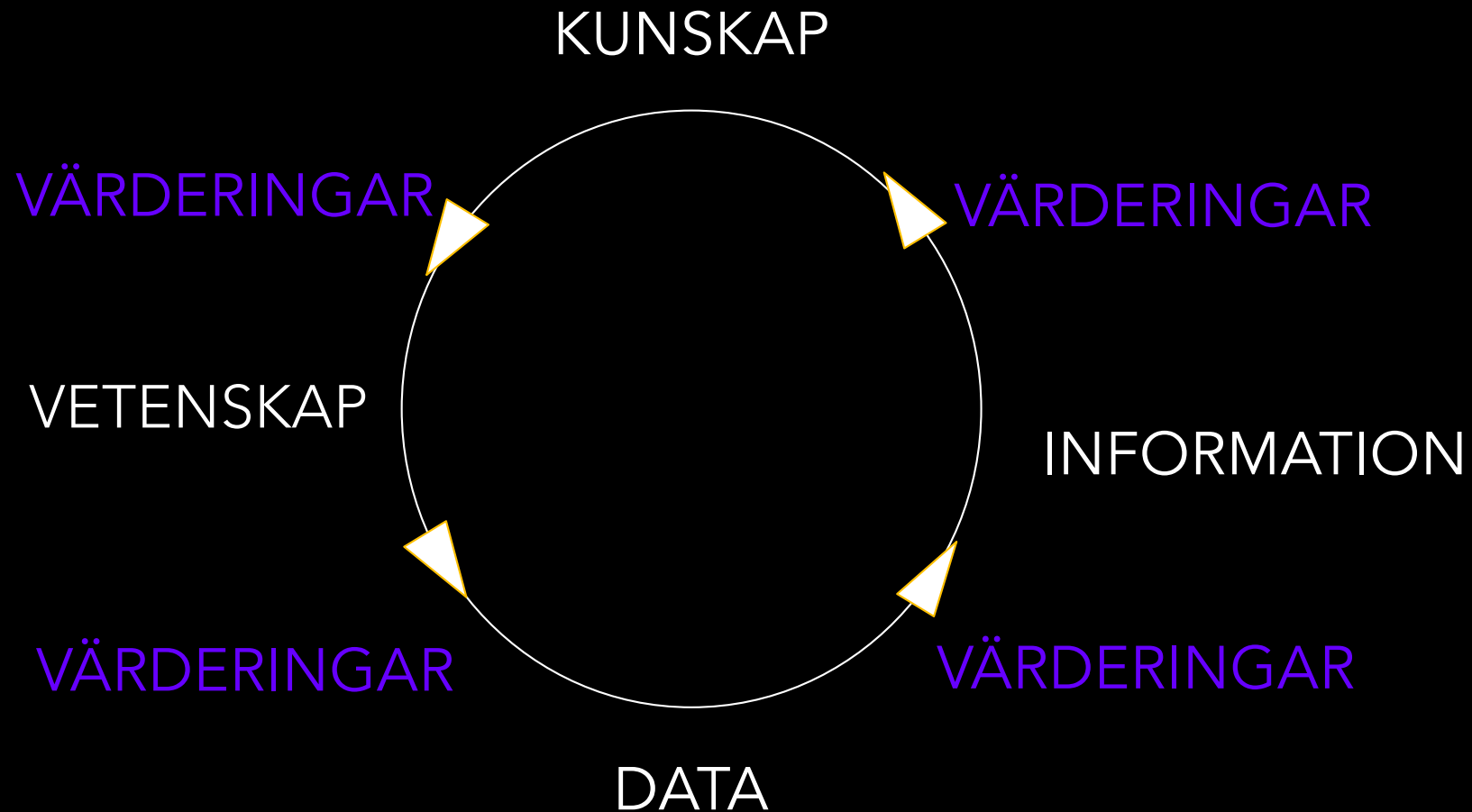


Baserad på:

Nancy Tuana (2015)
Coupled Ethical-Epistemic Analysis in Teaching
Ethics. Critical reflection on value choices.
CACM VOL. 500 NO. 12. Pages 27-29

<http://cacm.acm.org/magazines/2015/12/194630-coupled-ethical-epistemic-analysis-in-teaching-ethics/abstract>

Värderingar i kunskapsproduktion



Se: <https://link.springer.com/article/10.1007/s11023-024-09697-7> The Inherent Normativity of Concepts

VÄRDERINGAR

Värderingar fungerar som vägledning till handling och kunskap. De är relevanta för alla aspekter av vetenskaplig och teknisk praxis, inklusive upptäckt, analys och tillämpning.

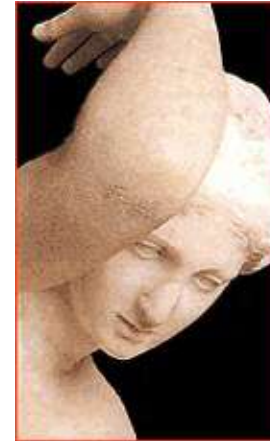
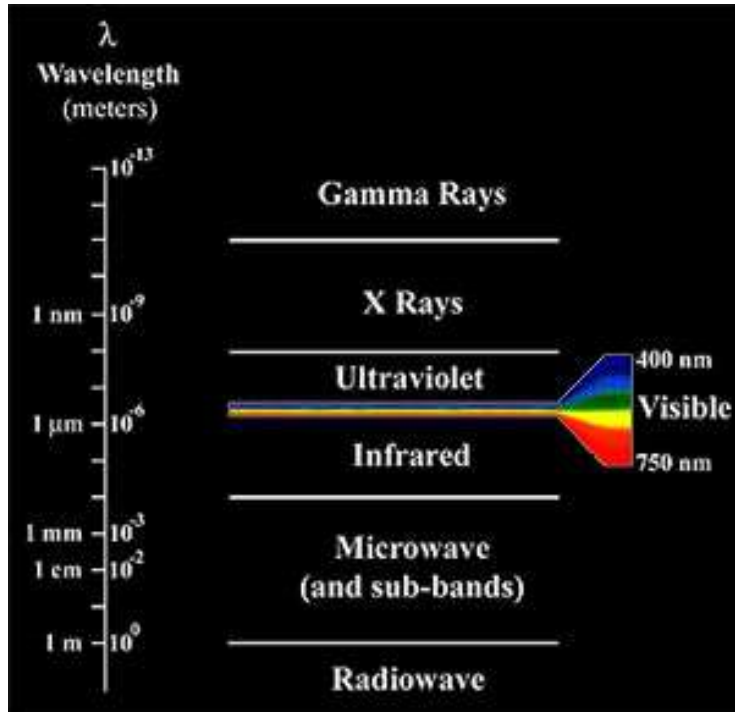
Kognitionsforskare har funnit att värden är en integrerad del av STEM-forskningen (Science, Technology, Engineering, and Mathematics).

TYPER AV VÄRDEN

Olika typer av värderingar kan vara involverade i beslutsfattande och resonemang, några exempel:

- etiska värderingar (samhällets bästa, rättvisa, hållbarhet)
- estetiska värden (enkelhet, elegans, komplexitet),
- epistemiska värden (prediktiv kraft, tillförlitlighet, koherens, omfattning)
- ekonomiska värden m.m.

Världen kan ses i olika ljus



Tänk om vi kunde se i vilken våglängd som helst av det elektromagnetiska spektrumet, från gammastrålar till radiovågor? Hur skulle världen se ut för oss?

Etiska aspekter i multikriterie-beslutsanalys

Ethical Aspects of Technology in the Multi-Criteria Decision Analysis

Gordana Dodig Crnkovic, Chalmers University of Technology and University of Gothenburg, Sweden
gordana.dodig-crnkovic@chalmers.se

Gaetana Sapienza, ABB Corporate Research and Mälardalen University, Sweden
gaetana.sapienza@se.abb.com

Abstract. In technological systems, decisions are often governed by multi criteria decision analysis (MCDA) techniques that take into account mutually opposing criteria for the system, and it results in ranking of alternatives. MCDA is based on value systems of decision-makers, and ethical deliberation in the process is implicit. We argue that it is necessary to make decision-making in technological systems transparent such that value basis and ethical considerations become explicit and subject for scrutiny of involved stakeholders. As different priorities, value systems and ethical choices result in different technical solutions, such solutions when put in use will promote those intrinsic and implicit values. In a society with ubiquitous technology, value aspects of technology are essential. At present there is no explicit mechanism to expose ethical aspects in these analyses, so they can easily be forgotten. As a support to encourage introduction of transparent value-based deliberation we propose an extended MCDA scheme that explicitly takes into account ethical analysis.

<https://tinyurl.com/mre9knw5>

Inclusion of Ethical Aspects in Multi-criteria Decision Analysis

Publisher: IEEE

[Cite This](#)

[PDF](#)

Gaetana Sapienza ; Gordana Dodig-Crnkovic ; Ivica Crnkovic [All Authors](#)

2

Cites in
Papers

216

Full
Text Views



Abstract

Document Sections

- I. Introduction
- II. Background
- III. The Importance of Ethical Aspects
- IV. Explicating Ethical Aspects in MCDA
- V. Case Study:
HW/SW
Partitioning for a
Wind Turbine

Abstract:

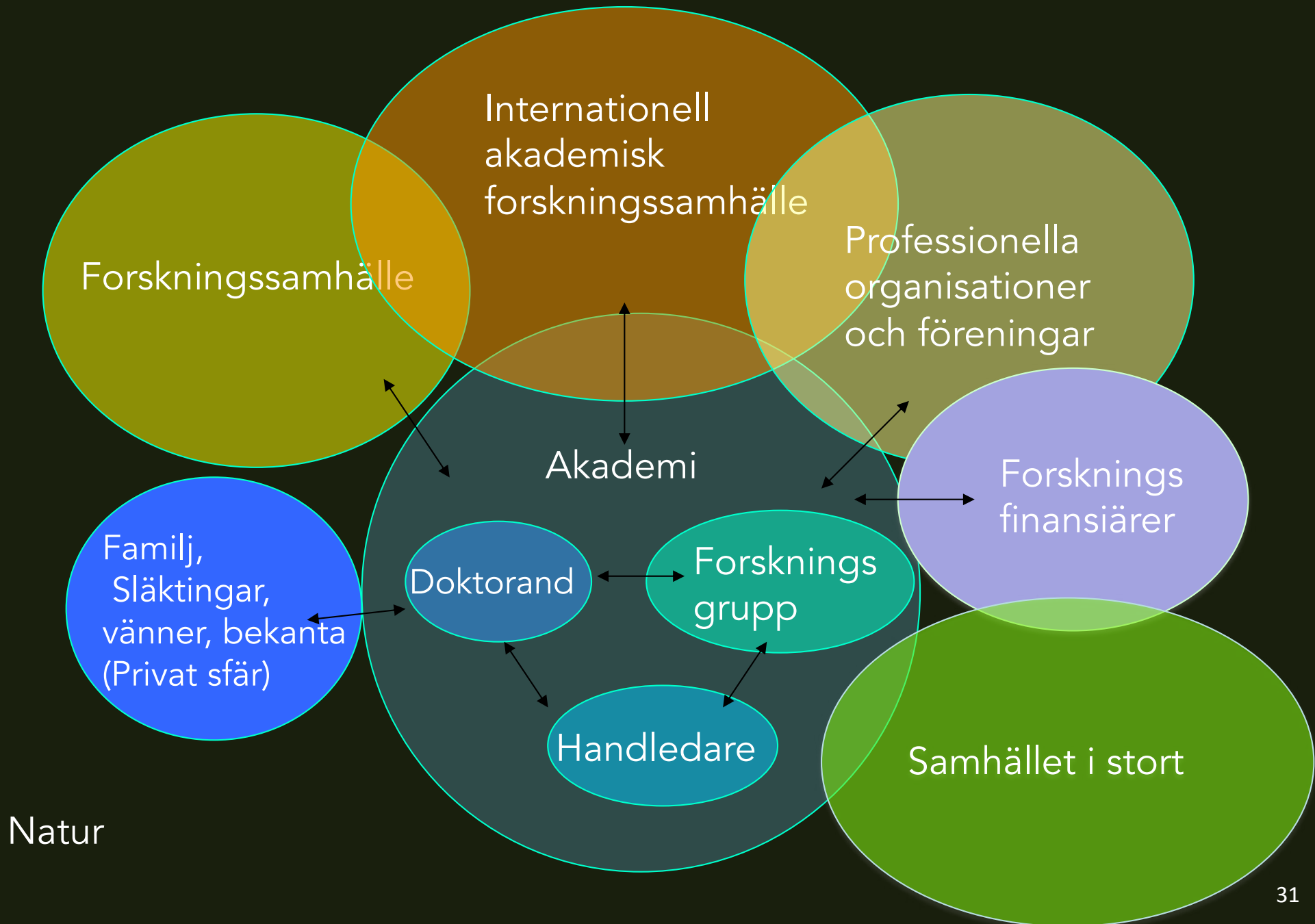
Decision process is often based on multi-faceted and mutually opposing criteria. In order to provide rigorous techniques for problem structuring and criteria aggregation used for classification and ranking of alternatives, Multiple Criteria Decision Analysis (MCDA) has been used as a method to achieve architectural decisions. Even though it has already been argued in literature that MCDA essentially depends on value systems of decision-makers, it is a question how the decision result reflects a particular criterion, requirement or a particular decision. This is especially true if a criterion is not precisely specified. In this paper we analyse the ethical aspects of MCDA. In our analysis we argue that it is in the long run necessary to make value basis of decision-making and ethical considerations explicit and subject for scrutiny. As a support to encourage introduction of transparent value-based deliberation we propose an extended MCDA scheme that would explicitly take into account ethical analysis. As an illustration, we present an industrial case study for the Software (SW)/Hardware (HW) partitioning of a wind turbine application in which different decisions can be taken, depending on the ethical aspects.

Published in: 2016 1st International Workshop on Decision Making in Software ARCHitecture

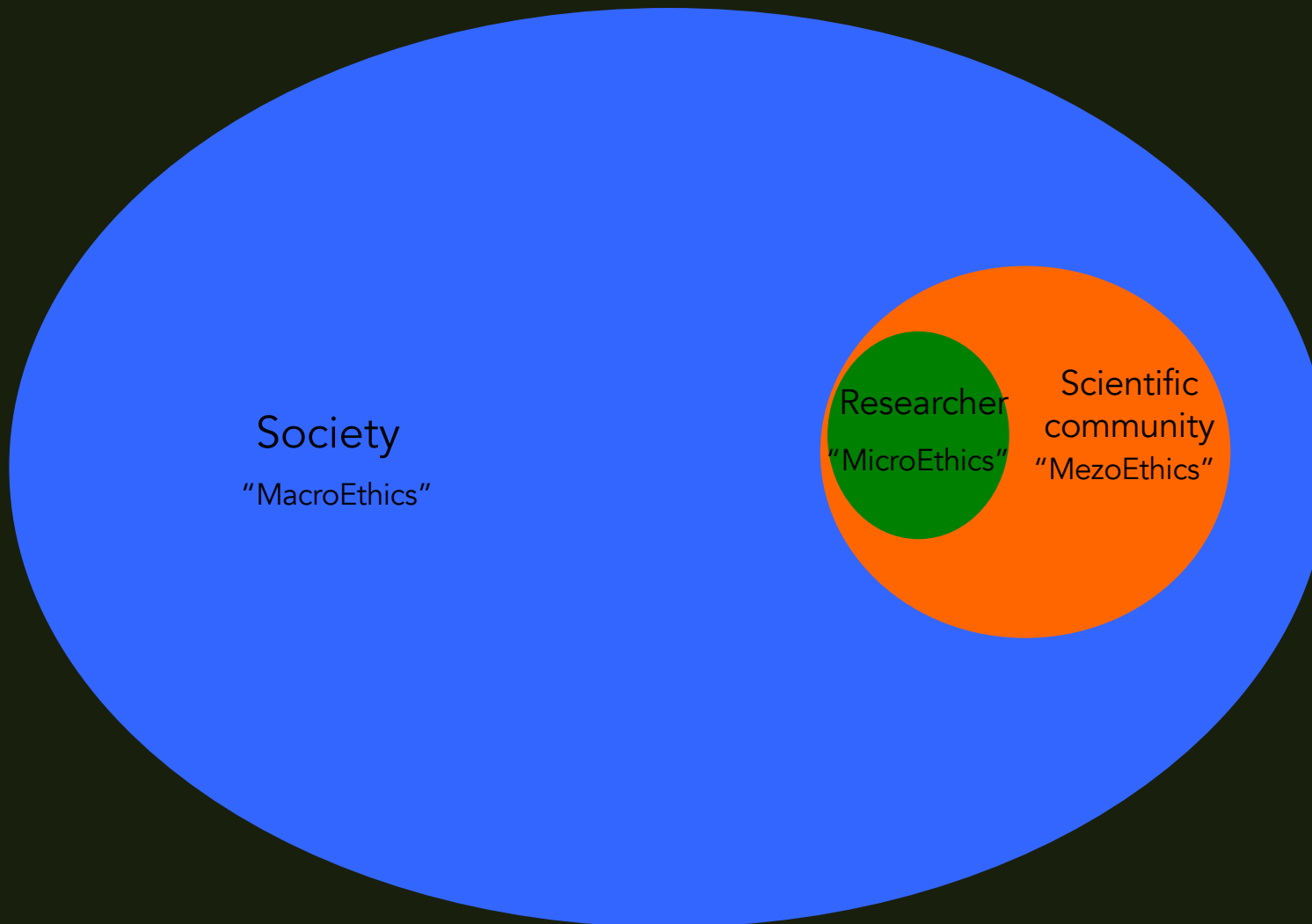


<https://ieeexplore.ieee.org/document/7496439>

AKTÖRER I ETT AKADEMISKT FORSKNINGSPROJEKT



Forskningsetiska domäner



Is it true?

1. Scientific integrity

Is it fair?

2. Collegiality

3. Protection of human subjects

4. Animal welfare

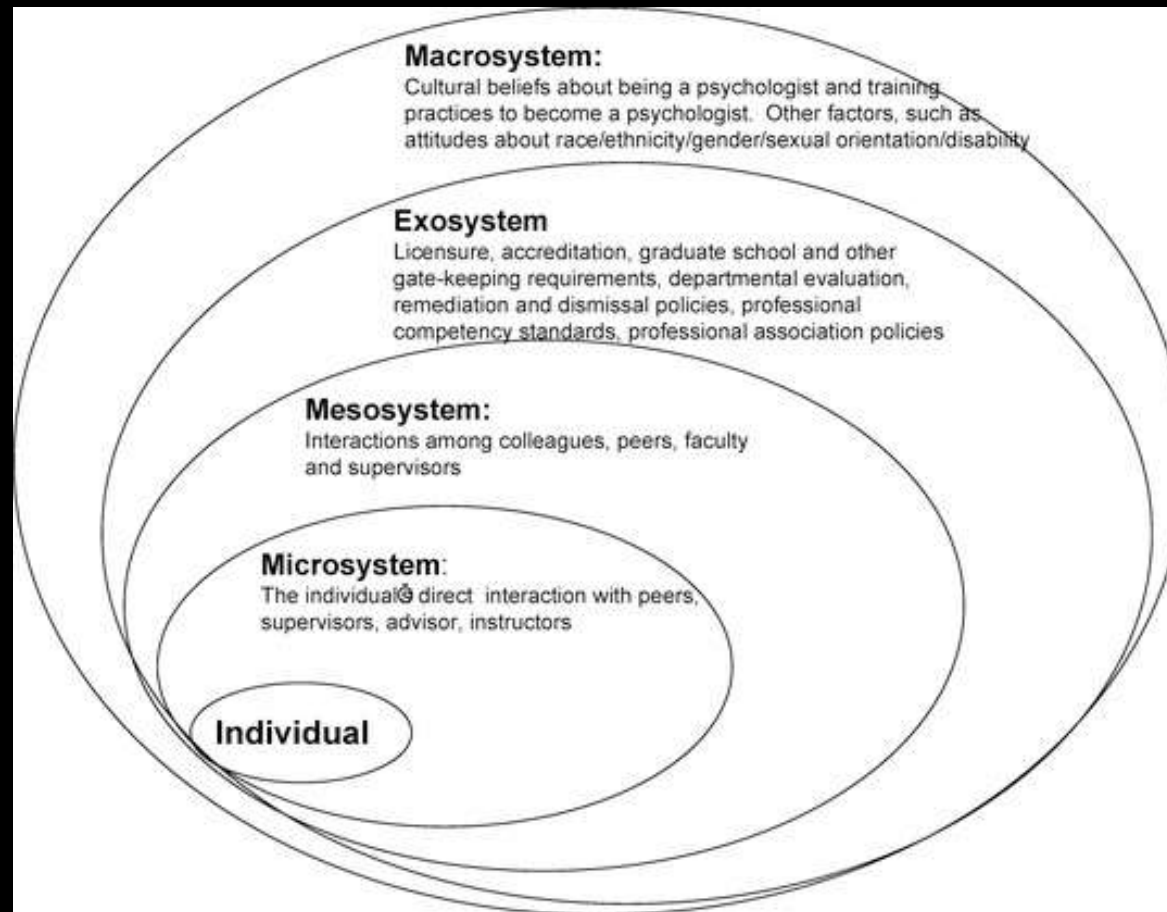
5. Institutional integrity

Is it wise?

6. Social responsibility

Kenneth D. Pimple (2002) "Six Domains of Research Ethics. A Heuristic Framework for the Responsible Conduct of Research". *Science and Engineering Ethics* 8 , 191-205

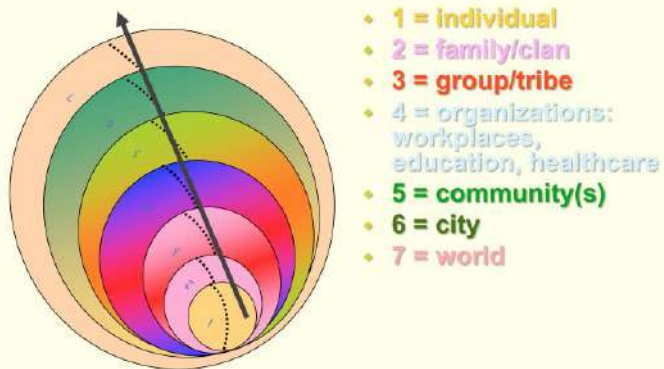
Micro – Meso – Exo – Macro Domains



you will recognize this **domain-based** view in the analysis of many different types of problems – organization of society, sustainability of cities, ecology, economics, ethical aspects etc.

Complexity Aspects Relating Micro – Meso – Exo – Macro Levels of Analysis – Example of City

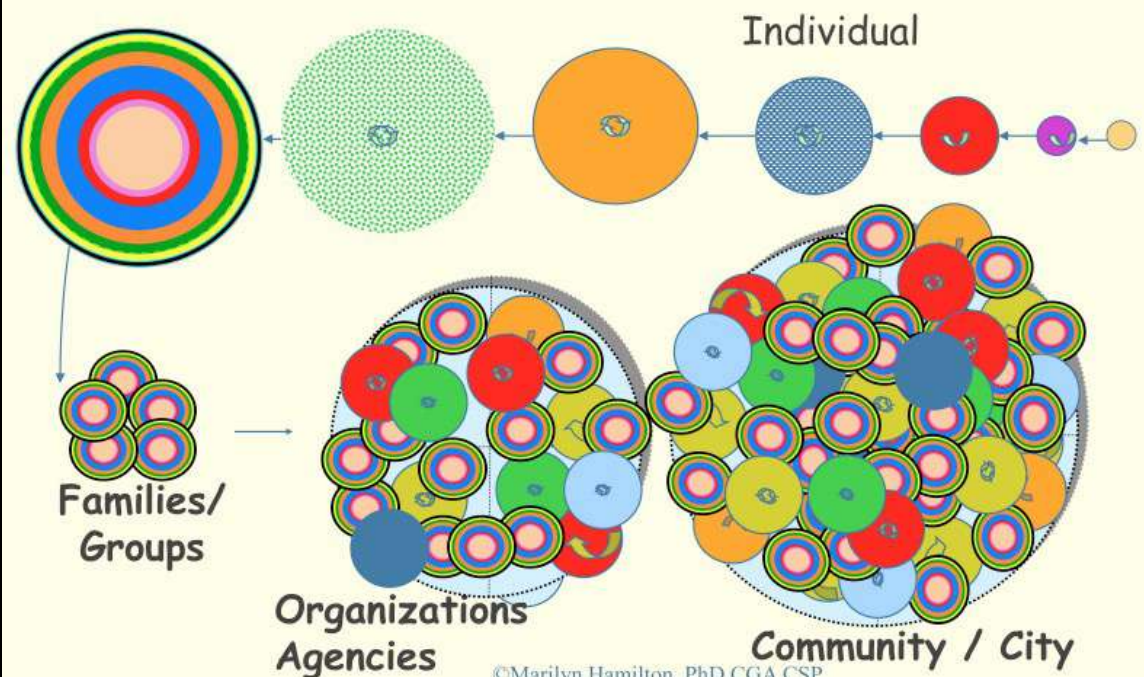
Map 2: The Nested Holarchy of City Systems



©Marilyn Hamilton PhD CGA CSP

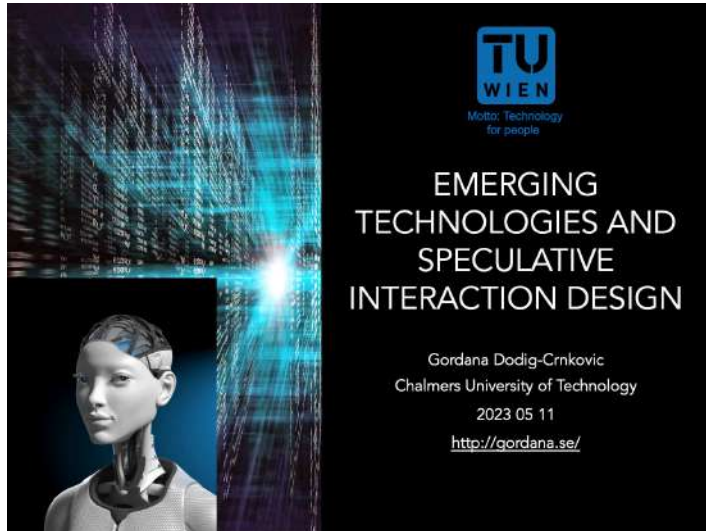
A holarchy, in the terminology of Arthur Koestler, is a connection between holons, where a holon is both a part and a whole. The term was coined in Koestler's 1967 book *The Ghost in the Machine*.

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



©Marilyn Hamilton PhD CGA CSP

Framtidsperspektiv



<https://tinyurl.com/3s3784dc>



<https://tinyurl.com/3hbyfts4>

Potentiella framtida etiska utmaningar med AI

Vikten av tvärvetenskapligt samarbete inom AI-etik

Elevernas roll i att forma en etisk AI-framtid

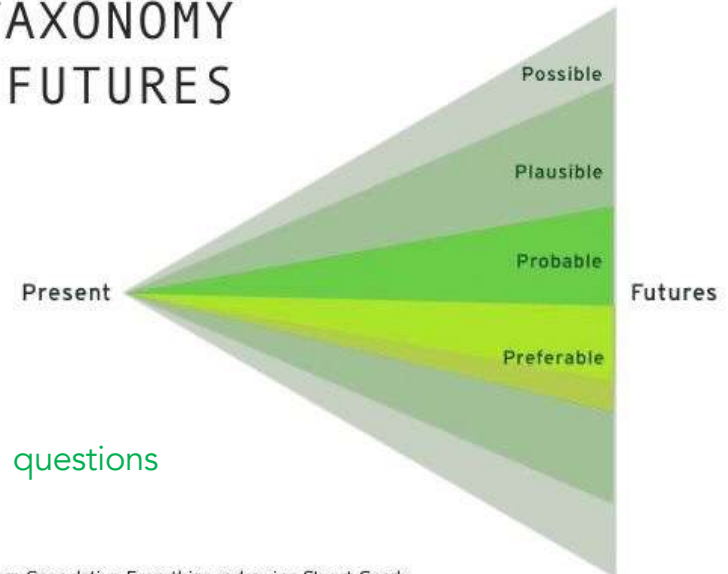
<https://www.youtube.com/watch?v=Ctuhh8Vqtfl>

Fei Fei Li, Professor, Stanford University on the History and Future of AI at Data + AI Summit 2024

SPEKILATIVT ALLT / SPECULATIVE EVERYTHING



A TAXONOMY OF FUTURES



"what if" questions

Redrawn from *Speculative Everything* redrawing Stuart Candy

Table of Contents:

- Beyond radical design?
- A map of unreality
- Design as critique
- Consuming monsters: big, perfect, infectious
- A methodological playground: fictional worlds and thought experiments
- Physical fictions: invitations to make believe
- Aesthetics of unreality
- Between reality and the impossible
- Speculative everything.

Design utan gränser. Att designa för emergens i en värld av
"forsande vatten".

Design Unbound. Designing for emergence in a 'white water world'.
(1) Designing for Emergence & (2) Ecologies of Change

Design Unbound. Designing for
Emergence in a White Water World.

Ann Pendleton-Jullian and
John Seely Brown, MIT Press 2018

<https://www.desunbound.com/>
<https://www.youtube.com/watch?v=-U8h4wNBfCO>
<https://www.youtube.com/watch?v=tFPvK1mO6Sg>
<https://www.youtube.com/watch?v=Lto8szGvPfM>
https://www.desunbound.com/assets/DesUnbound_chapter_8.pdf



Richard Buchanan (1992) Wicked Problems in Design Thinking. Design
Issues, Vol. 8, No. 2, pp. 5-21. The MIT Press
<http://www.jstor.org/stable/1511637>.

En värld som forsande vatten – komplex och dynamisk

A 'White Water World' – complex & dynamic

"We are forcing the past as a solution set. But the past as a solution set is not a viable option. We need a new toolset."

Design Unbound presents a new tool set for having agency in the world today, which we characterize as a 'white water world' – one that is rapidly changing, hyperconnected and radically contingent.

Imagination is a 'muscle that must be exercised' (John Seely Brown)

Hyperconnectivity causes a transition from equilibrium to constant **non-equilibrium**. The need for adaptivity, anticipation, and **resilience**.

Complexity science gives us a new lens through which to view **the world as one that is entangled and emerging**.



'**Wicked problems**': As soon as you start to solve them, they morph.
"Computational irreducibility": You must run the model to see the outcome. Computation takes the same time as the process itself.

MÅNGVETENSKAPLIGT/ TVÄRVETENSKAPLIGT SAMARBETE MULTIDISCIPLINARY/ TRANSDISCIPLINARY COLLABORATION

nature methods

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature methods](#) > [comment](#) > [article](#)

Comment | Published: 20 June 2024

Collaborative ethics: innovating collaboration between ethicists and life scientists

[Jeantine E. Lunshof](#)  & [Julia Rijssenbeek](#)

[Nature Methods](#) 21, 1571–1574 (2024) | [Cite this article](#)

673 Accesses | 1 Citations | 52 Altmetric | [Metrics](#)

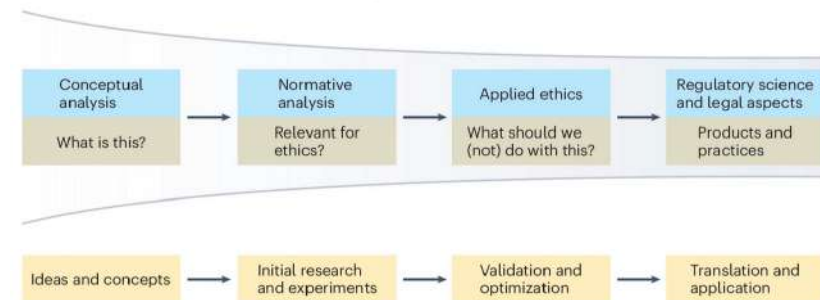
Scientific breakthroughs can change how we understand and live in the world, disrupting long-held assumptions and concepts and raising new questions for philosophy and science. To address these challenges, we describe a model for collaboration of scientists with philosophers and ethicists, and its benefits to the research process and outcomes.

Is there a place for ethics *in* scientific research, not *about* science or *after* scientific breakthroughs? We are convinced that there is, and we describe here our model for

Fig. 1: Collaborative ethics innovation funnel.

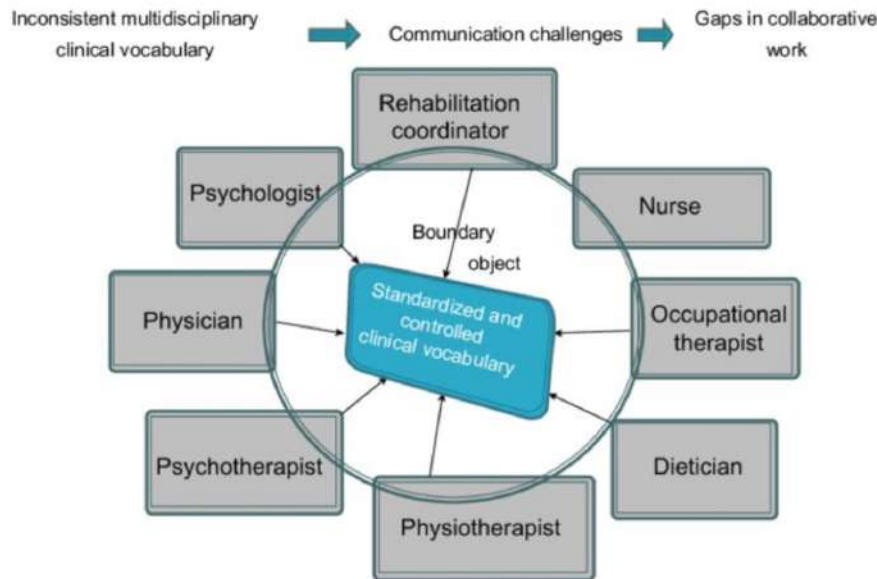
From: [Collaborative ethics: innovating collaboration between ethicists and life scientists](#)

Collaborative ethics innovation in the research process

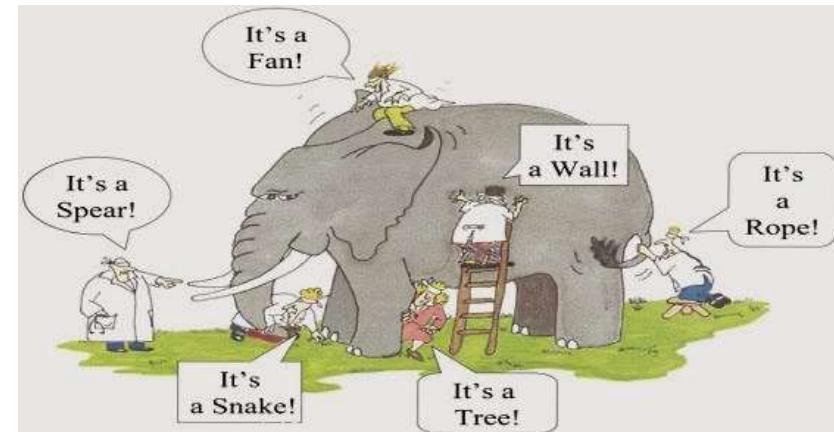


Modeled after the Wyss Institute Technology Innovation Funnel (Supplementary Fig. 1). The light blue boxes show steps 1–4 of the collaborative ethics process. In gray the specific questions for each step. The yellow boxes show the corresponding steps in the research and technology innovation process. The funnel shape illustrates the fact that out of a multitude of initial ideas (step 1), only a few reach the stage of translation (step 4).

Ett gränsojekt/ Boundary object



<https://tinyurl.com/4ky2tszb>



<https://medium.com/betterism/the-blind-men-and-the-elephant-596ec8a72a7d>

Boundary Object

Origin: The term "boundary object" comes from the sociology of science and was introduced by Susan Leigh Star and James R. Griesemer in their 1989 work.

Definition: Boundary objects are objects that are flexible enough to be interpreted differently by different social groups but robust enough to maintain a common identity across these groups. They help facilitate communication and collaboration across different communities of practice or disciplines.

Function: Boundary objects allow people with different perspectives or expertise to work together without necessarily needing full agreement on the object's meaning. For example, a map might be used differently by a hiker, a biologist, and a city planner, but the map itself facilitates collaboration because all parties find it useful in their specific ways.

Ett gränsobjekt/ Boundary object

Ett gränsobjekt (boundary object) är ett begrepp som beskriver föremål eller fenomen vilka fungerar som gemensamma referenspunkter mellan olika grupper eller kunskapsområden. Gränsobjekt gör det möjligt för personer med olika bakgrund, intressen och perspektiv att kommunicera, samarbeta och skapa gemensam förståelse utan att behöva ha identiska uppfattningar eller mål.

Gränsobjekt kännetecknas vanligtvis av följande egenskaper:

- Flexibilitet: De kan tolkas olika av olika aktörer.
- Gemensamt fokus: De hjälper till att samordna aktiviteter och diskussioner.
- Sammanlänkning: De överbryggar olika grupper, discipliner eller kunskapsfält.

Exempel på gränsobjekt kan vara ritningar, prototyper, kartor, modeller eller begrepp som används inom tvärvetenskapliga projekt eller samarbete mellan forskare och praktiker.

Ett gränsobjekt/ Boundary object

Begreppet "*information*" passar utmärkt som ett gränsobjekt, eftersom:

Inom **fysik och datavetenskap** (Landauers princip) information ses som en fysisk entitet, något som kräver fysisk implementation.

Inom **filosofi eller semiotik** definieras information ofta utifrån *mening* och tolkning snarare än strikt fysikaliska aspekter.

Inom **psykologi och neurovetenskap** kopplas information ofta till *hjärnans bearbetning* av sinnesintryck och stimuli.

Inom **sociologi och kommunikationsvetenskap** fokuseras på informationsöverföring mellan individer, grupper eller organisationer, ofta med betoning på sociala relationer och kultur.

Just därför fungerar "*information*" som ett gränsobjekt. Begreppet är flexibelt nog för att anpassas efter olika discipliners specifika behov och perspektiv.

Samtidigt utgör det en gemensam grund för tvärvetenskaplig kommunikation och samarbeten.

Flytande betecknare/Floating signifier

Floating Signifier

Origin: This concept comes from semiotics and structuralism, particularly the work of Claude Lévi-Strauss and later poststructuralist thinkers like Jacques Lacan and Roland Barthes.

Definition: A floating signifier is a symbol or term that can take on a wide range of meanings depending on the context, yet it doesn't have a fixed or intrinsic meaning of its own. It "floats" in the sense that it can be attached to different interpretations, which makes it a flexible but also ambiguous tool for communication.

Function: Floating signifiers are often used in ideological, political, or cultural discourse, where different groups or individuals ascribe their meanings to terms like "freedom," "justice," or "democracy." These terms have no single, fixed meaning but carry significant weight in discourse precisely because they can be filled with different meanings.

Relationship Between Boundary Object and Floating Signifier

Flexibility and Multiplicity of Meaning: Both boundary objects and floating signifiers share the characteristic of being interpretable in different ways by different groups or individuals. They are flexible and serve as tools for communication across social or discursive boundaries.

- A **boundary object** is generally more focused on practical collaboration across different social worlds, where each group applies its own interpretation to the object while still working together. It maintains a stable identity across contexts, even as its meaning shifts.
- A **floating signifier**, on the other hand, is more fluid and context-dependent, used in discourses where the meaning can vary widely, often without any single, stable identity. The lack of a fixed meaning is what gives it power in political, cultural, or ideological debates.

Contextual Use: Boundary objects are often used in **interdisciplinary or inter-organizational settings**, helping people collaborate despite different understandings. Floating signifiers, meanwhile, appear in **discourse analysis or cultural studies**, where the focus is on how meaning is contested and shaped by different ideologies or social forces.

Flytande betäcknare

En flytande betecknare är ett begrepp från semiotik och diskursteori som syftar på ett ord eller en term vars betydelse inte är fast eller entydig, utan är öppen och förändras beroende på sammanhang och vem som använder det.

Flytande betecknare kännetecknas av följande:

- Instabil betydelse - de saknar en definitiv, permanent innebörd och kan tolkas på många sätt.
- Kulturellt och ideologiskt laddade- de används ofta för att uttrycka ideologiska eller kulturella värderingar som är föremål för konflikt eller debatt.
- Sociala och politiska effecter - de används ofta strategiskt inom politik, media och kommunikation eftersom deras vaga innebörd möjliggör olika typer av identifikation och mobilisering.

Exempel på flytande betecknare kan vara begrepp som:

"frihet", "rättvisa", "demokrati", "hållbarhet", ...

The Digital Humanism Initiative

Initiativet för digital humanism

The Digital Humanism Initiative is an international collaboration seeking to build a community of scholars, policy makers, and industrial players who are focused on ensuring that technology development remains centered on human interests.

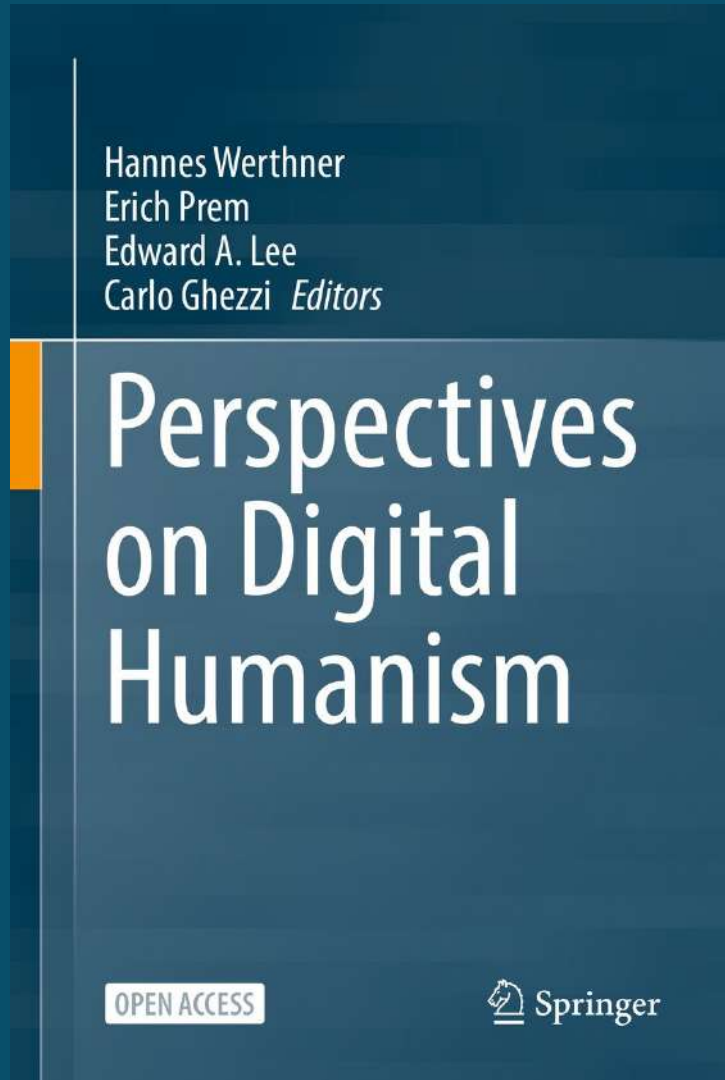
- Digital humanism is a global, international issue.
- The approach: scientific, transdisciplinary, interdisciplinary, multidisciplinary, in the tradition of the Enlightenment.
- **People are the central focus**, as individuals and societies. [this is not in conflict with environment and animal rights!]*
- Technology is for people and not the other way around.
- Humankind is at the center.
- Building a just and democratic society with humans at the center of technological progress.

<https://dighum.ec.tuwien.ac.at/> Digital Humanism movement web page @ TUW – Technical University in Vienna

E. Prem, L. Hardman, H. Werthner, P. Timmers (eds.). Research, innovation, and education roadmap for digital humanism. The Digital Humanism Initiative. Vienna, 2022. <https://dighum.ec.tuwien.ac.at/>

*Mark Coeckelbergh (2024) What is Digital Humanism? Journal of Responsible Technology 17 100073

Perspectives on Digital Humanism – Open Access

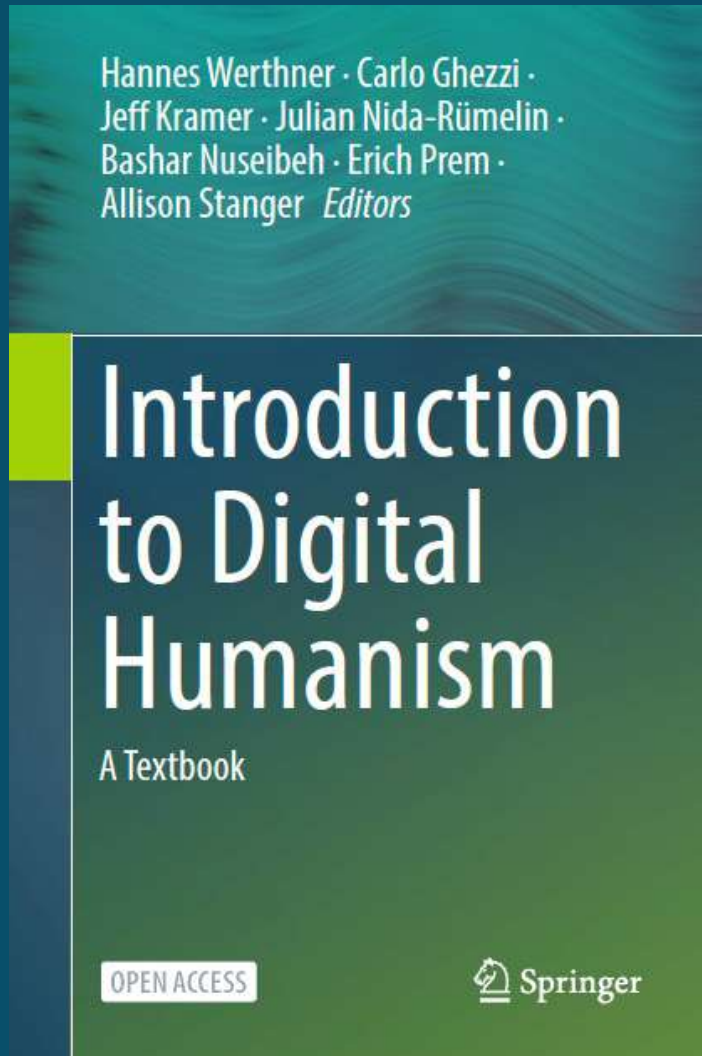


Hannes Werthner, Erich Prem, Edward A. Lee, and Carlo Ghezzi (eds): *Perspectives on Digital Humanism*, Springer, 2022.

<https://link.springer.com/book/10.1007/978-3-030-86144-5>

Introduction to Digital Humanism – A Textbook

Open Access

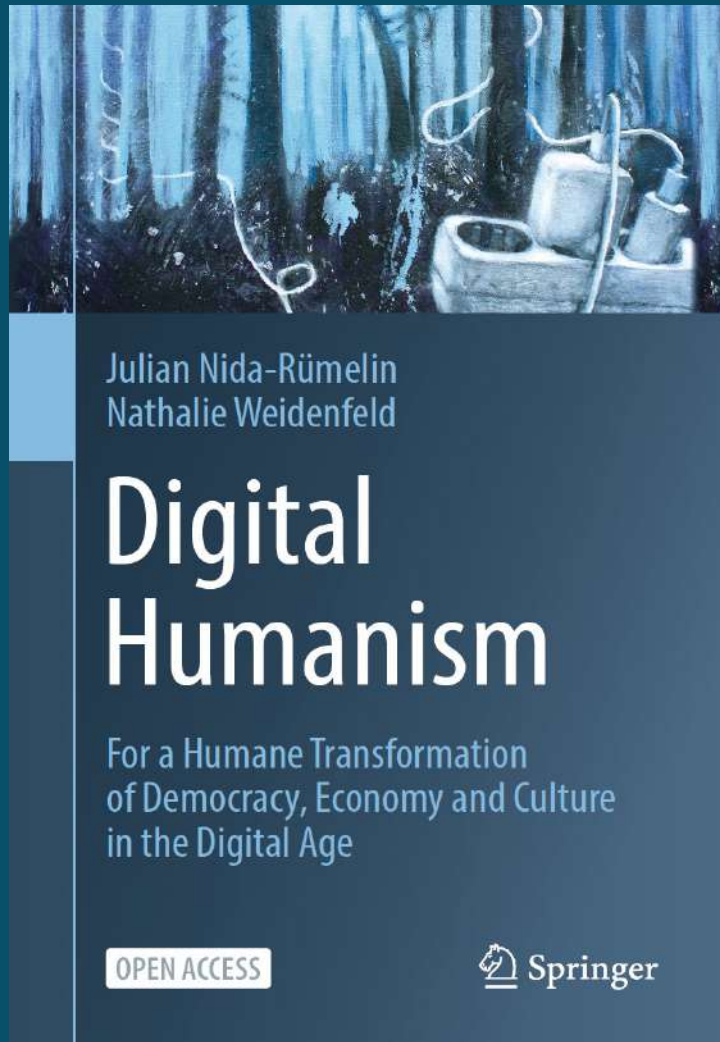


Hannes Werthner, Carlo Ghezzi, Jeff Kramer, Julian Nida-Rümelin, Bashar Nuseibeh, Erich Prem, and Allison Stanger (eds): **Introduction to Digital Humanism**, Springer, 2024.

<https://link.springer.com/book/10.1007/978-3-030-86144-5>

Digital Humanism – For a Humane Transformation Of Democracy, Economy, and Culture in the Digital Age

Open Access



Julian Nida-Rümelin, Nathalie Weidenfeld (eds):
Digital Humanism. For a Humane Transformation of
Democracy, Economy and Culture in the Digital Age,
Springer, 2022.

<https://link.springer.com/book/10.1007/978-3-031-12482-2>

Digital Humanism Lecture Series

<https://dighum.ec.tuwien.ac.at/news-events/>

<https://www.youtube.com/@DigitalHumanism> Youtube channel
(Stuart Russel, Gary Marcus, Edward Lee, Deborah G. Johnson, Julian Nida-Rümelin,...)

Digital Humanism Manifesto

“Today, we experience the co-evolution of technology and humankind. The flood of data, algorithms, and computational power is disrupting the very fabric of society by changing human interactions, societal institutions, economies, and political structures. Science and the humanities are not exempt. This disruption simultaneously creates and threatens jobs, produces and destroys wealth, and improves and damages our ecology. It shifts power structures, thereby blurring the human and the machine.”

<https://dighum.ec.tuwien.ac.at/dighum-manifesto/>

DIGITAL HUMANISM CONFERENCE



<https://dighum.wien>

Avslutning och diskussion



Tankar om AI risker och om vikten att reglera AI:
<https://www.youtube.com/watch?v=QEGjCcU0FLs>
Will AI Be Humanity's Last Act? with Stuart Russell
(författaren till "AI boken")

<https://www.apogeonline.com/articoli/regole-di-composizione-fotografica-usare-la-prospettiva-a-piu-punti-harold-davis/>

Resurser för vidare läsning och fördjupning

Lund University

AI Lund is an interdisciplinary network for research, education, and innovation in the area of Artificial Intelligence coordinated by Lund University. Read more at <https://ai.lu.se>

Umeå University

Centre for Transdisciplinary AI (TAIGA)
<https://www.umu.se/en/centre-for-transdisciplinary-ai>

#frAlday

<https://www.umu.se/en/research/our-research/features-and-news/artificial-intelligence/friday/>

Chalmers University of Technology

<https://www.youtube.com/@chalmersairesearchcentre>
Chalmers AI Research Centre

https://www.youtube.com/watch?v=S_c33fnSM9o

AI i klassrummet hur rustar vi för framtiden, Karl de Fine Licht

<https://tinyurl.com/bdh8nxpx> AI Ethics – Neither tyrant nor slave- AI Ethics with Aksel Sterri

