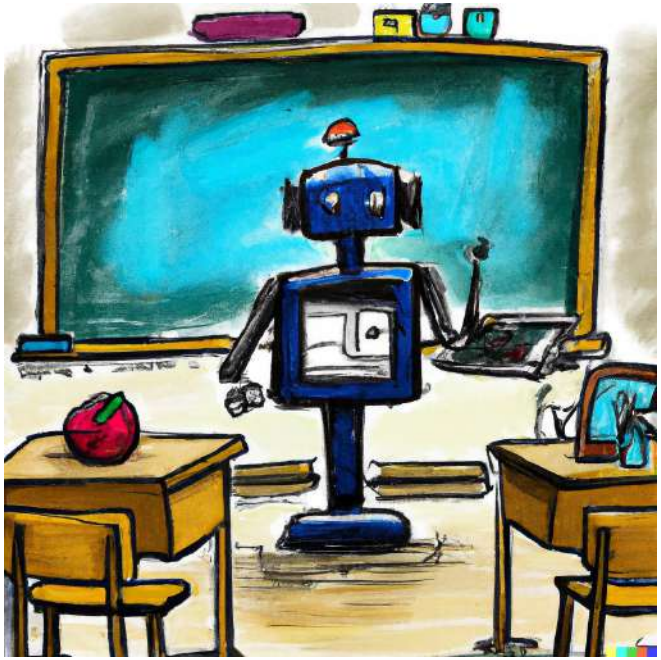
FUNDACION TATIANA

Proyecto ELAI: Lecciones éticas de la inteligencia artificial

Ethical Lessons of Artificial Intelligence

uc3m | Universidad Carlos III de Madrid

homo hominis sacra res



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://demaquinaseintenciones.wordpress.com/elai/
Salón de Grados, Edificio Padre Soler, campus de Leganés.

https://www.usi.ch/en/feeds/27126 12 April 2024

Navigating the White-Water World with Digital Humanism

Emergent Intelligent Technologies between Utopia and Dystopia

Gordana Dodig Crnkovic

Senior Professor of Computer Science at Mälardalen University and

Professor of Interaction Design, Chalmers University of Technology, Sweden, http://gordana.se/

My affiliations

GÖTEBORGS
UNIVERSITET

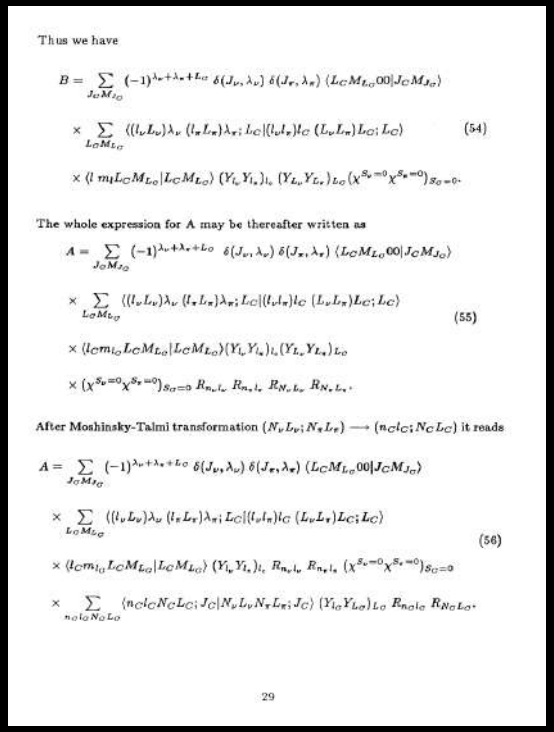
Division:
Computer Science and Software Engineering

Research groups:
Interaction Design and Software Engineering
Critical Robotics

Division:
Division of Computer Science and Software Engineering

Research groups:
Artificial Intelligence and Intelligent Systems
Ubiquitous Computing

My background - from formal to natural languages



Thus we have

$$B = \sum_{J_C M_{J_C}} (-1)^{\lambda_\nu+\lambda_\pi+L_C} \delta(J_\nu,\lambda_\nu)\,\delta(J_\pi,\lambda_\pi)\,\langle L_C M_{L_C} 00|J_C M_{J_C}\rangle$$

$$\times \sum_{L_C M_{L_C}} \langle (l_\nu L_\nu)\lambda_\nu\,(l_\pi L_\pi)\lambda_\pi; l_C|(l_\nu l_\pi)l_C\,(L_\nu L_\pi)L_C; L_C\rangle \qquad (54)$$

$$\times \langle l\ m_l L_C M_{L_C}|L_C M_{L_C}\rangle\,(Y_{l_\nu} Y_{l_\pi})_{l_\nu}\,(Y_{L_\nu} Y_{L_\pi})_{L_C} (\chi^{S_\nu=0}\chi^{S_\pi=0})_{S_C=0}.$$

The whole expression for A may be thereafter written as

$$A = \sum_{J_C M_{J_C}} (-1)^{\lambda_\nu+\lambda_\pi+L_C}\ \delta(J_\nu,\lambda_\nu)\,\delta(J_\pi,\lambda_\pi)\,\langle L_C M_{L_C} 00|J_C M_{J_C}\rangle$$

$$\times \sum_{L_C M_{L_C}} \langle (l_\nu L_\nu)\lambda_\nu\,(l_\pi L_\pi)\lambda_\pi; L_C|(l_\nu l_\pi)l_C\,(L_\nu L_\pi)L_C; L_C\rangle \qquad (55)$$

$$\times \langle l_C m_{l_C} L_C M_{L_C}|L_C M_{L_C}\rangle (Y_{l_\nu} Y_{l_\pi})_{l_\nu} (Y_{L_\nu} Y_{L_\pi})_{L_C}$$

$$\times (\chi^{S_\nu=0}\chi^{S_\pi=0})_{S_C=0}\ R_{n_\nu l_\nu}\ R_{n_\pi l_\pi}\ R_{N_\nu L_\nu}\ R_{N_\pi L_\pi}.$$

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

$$A = \sum_{J_C M_{J_C}} (-1)^{\lambda_\nu+\lambda_\pi+L_C}\ \delta(J_\nu,\lambda_\nu)\,\delta(J_\pi,\lambda_\pi)\,\langle L_C M_{L_C} 00|J_C M_{J_C}\rangle$$

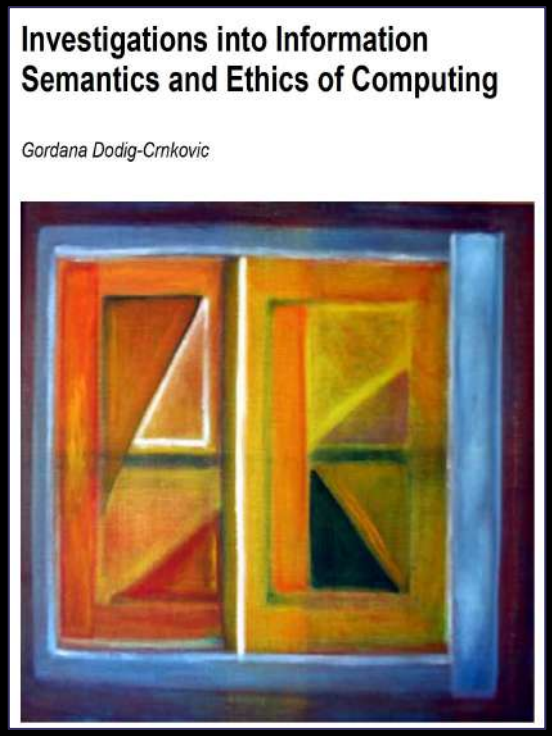
$$\times \sum_{L_C M_{L_C}} \langle (l_\nu L_\nu)\lambda_\nu\,(l_\pi L_\pi)\lambda_\pi; L_C|(l_\nu l_\pi)l_C\,(L_\nu L_\pi)L_C; L_C\rangle \qquad (56)$$

$$\times \langle l_C m_{l_C} L_C M_{L_C}|L_C M_{L_C}\rangle\,(Y_{l_\nu} Y_{l_\pi})_{l_\nu}\ R_{n_\nu l_\nu}\ R_{n_\pi l_\pi}\ (\chi^{S_\nu=0}\chi^{S_\pi=0})_{S_C=0}$$

$$\times \sum_{n_C l_C N_C L_C} \langle n_C l_C N_C L_C; J_C|N_\nu L_\nu N_\pi L_\pi; J_C\rangle\,(Y_{l_C} Y_{L_C})_{L_C}\ R_{n_C l_C}\ R_{N_C L_C}.$$

29

PhD in Physics, 1988
On Alpha-decay, Department of Physics, University of Zagreb



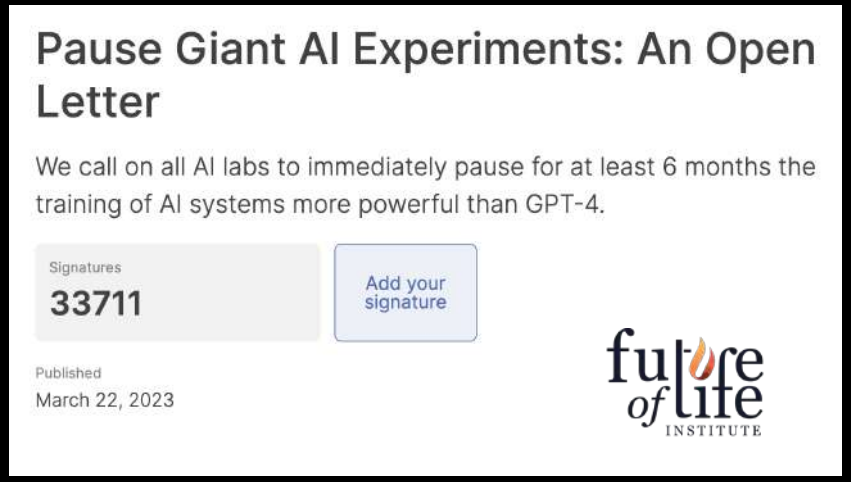
PhD in Computing, 2006
Computer Science,
Mälardalen University

Current: Morphological Computing and Cognition
AI Ethics, Digital Ethics, Digital Humanism

Transformative emerging intelligent technologies

- We live in an era of transformative AI technologies that profoundly alter our civilization, reshape existing software and hardware, and challenge our understanding of fundamental concepts such as intelligence, consciousness, language, education, research, ethics, sustainability, government, democracy, being human, and more.
- The pace of technological advancement is accelerating.
- Today's technology isn't an isolated domain managed solely by specialists and industries. Instead, it's an integral component of a broader techno-social system.
- As stakeholders in this development—both professionals and citizens—we must maintain a long-term perspective and actively participate in decision-making about future technologies. We can't assume that a few years from now technology will remain as it is today.
- The most dramatic development we are experiencing is in AI

Responses to the dramatic development of AI
Examples of collective action



Signatories include: Yoshua Bengio, Stuart Russell, Gary Marcus, Emad Mostaque, Elon Musk, Tristan Harris, Steve Wozniak and Yuval Noah Harari.

Geoffrey Hinton and Yoshua Bengio warned in May 2023:

"Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war," The letter published by nonprofit organization Center for AI Safety.

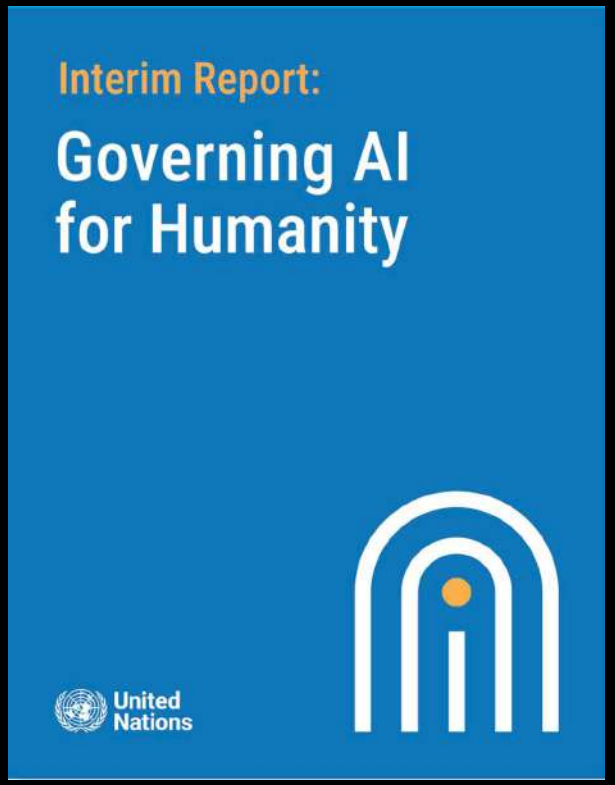
Other signatories include researchers from the Vector Institute and Mila, as well as professors from universities across Canada. Open AI CEO Sam Altman, Microsoft CTO Kevin Scott, etc.

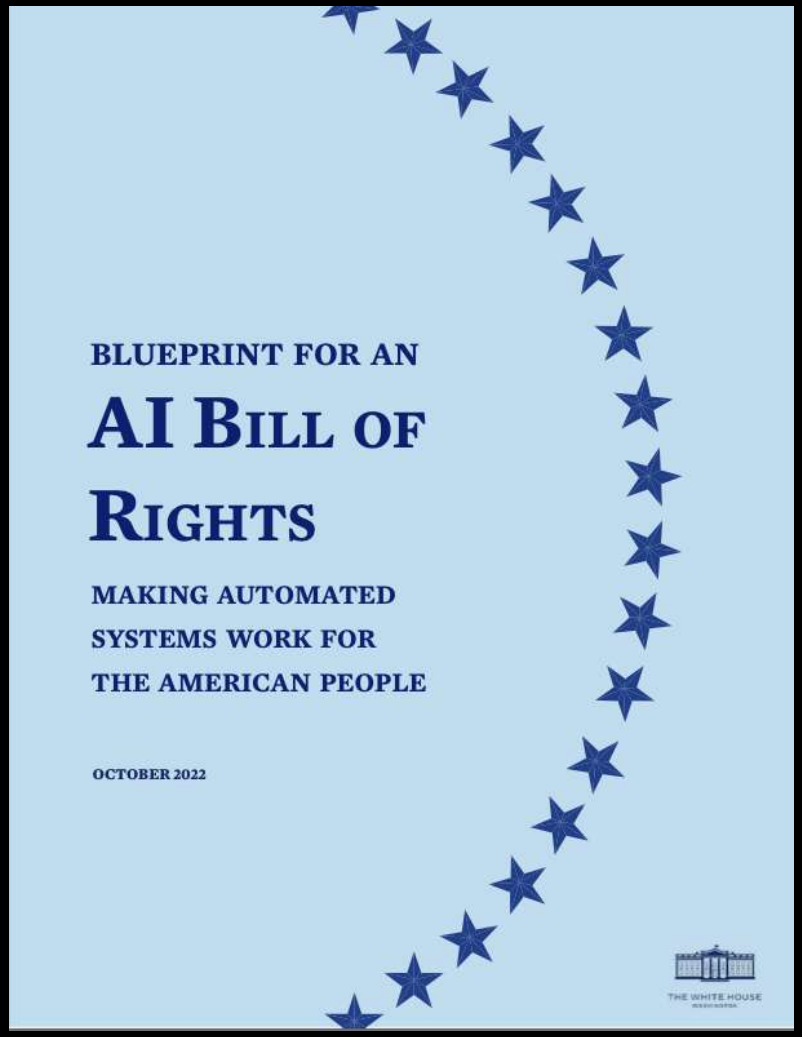
Academics, CEOs sign on in support of AI regulation and Bill C-27 as Canadian companies race to adopt the technology

Since Last year, work on AI regulation

United Nations report (2023)
“Governing AI for Humanity”

https://w.un.org.techenvoy/files/ai_advisory_body_interim_report.pdf





https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf

The US AI Bill of Rights outlines principles, including that people have a **right to control how their data is used** and **to not be discriminated against by unfair algorithms.**

It is a white paper, which does not have the force of law. It's primarily aimed at **the federal government** and could influence **which technologies government agencies acquire**, or help parents, workers, policymakers, and designers **ask tough questions about artificial intelligence systems**.

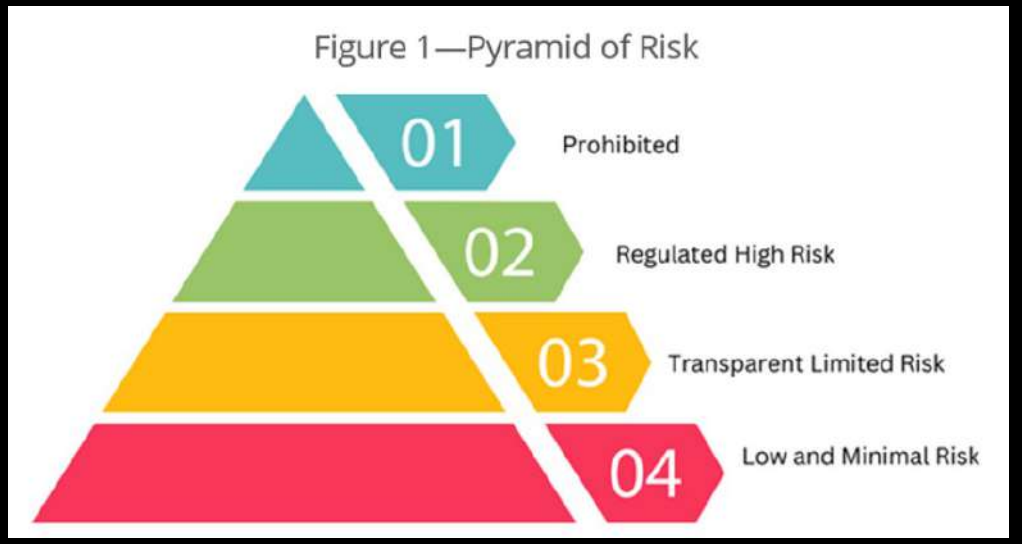
However, **it can't constrain large tech companies**, which arguably play a bigger role in shaping future applications of AI.

EU's "AI Act" (2024)

The world's first AI legislation

AI Act, European Commission. Shaping Europe's digital future

https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai



ISACA

The European Parliament granted final approval of the EU Artificial Intelligence Act on March 13, 2024, by a vote of 523 for passage, 46 against, and 49 abstaining. The Act faces a final step – approval by EU member states – as its provisions gradually take effect.

Timeline for the adoption of the European AI Act

EU AI Act timeline (as of March 2024)

date	Milestone
21 April 2021	EU Commission proposes the AI Act
6 December 2022	EU Council unanimously adopts the general approach of the law
9 December 2023	European Parliament negotiators and the Council Presidency agree on the final version
2 February 2024	EU Council of Ministers unanimously approves the draft law on the EU AI Act
13 February 2024	parliamentary committees approve the draft law
13 March 2024	EU Parliament approves the draft law
20 days after its publication in the Journal	Entry into force of the law
6 months after entry into force	Ban on AI systems with unacceptable risk
9 months after entry into force	Codes of conduct are applied
12 months after entry into force	Governance rules and obligations for General Purpose AI (GPAI) become applicable
24 months after entry into force, with specific exceptions	Start of application of the EU AI Act for AI systems (including Annex III)
36 months after entry into force, with specific exceptions	Application of the entire EU AI Act for all risk categories (including Annex II)

https://www.alexanderthamm.com/en/blog/eu-ai-act-timeline/

THINKING ABOUT THE RESPONSIBILITIES FOR NEW TECHNOLOGY

ASSIGNMENT OF RESPONSIBILITY: WHO DECIDES?

Time perspective

- Short-term perspective
  We, humans, decide
- Middle-term perspective
  AGI & We co-decide
- Long-term perspective
  Superintelligence? Who decides?

Levels of AI

- ANI (Narrow AI)
- AGI (Artificial General Intelligence)
- ASI (Artificial Super Intelligence)

Stakeholders

- Politicians
- Legislators
- Businesses
- Requirements engineers
- Designers, Developers
- Programmers
- Deployment engineers, testers
- Maintenance engineers

Learning from experience. Feedback on development & design

https://tinyurl.com/pjbdyn95 Global AI governance: barriers and pathways forward- Huw Roberts, Emmie Hine, Mariarosaria Taddeo, Luciano Floridi

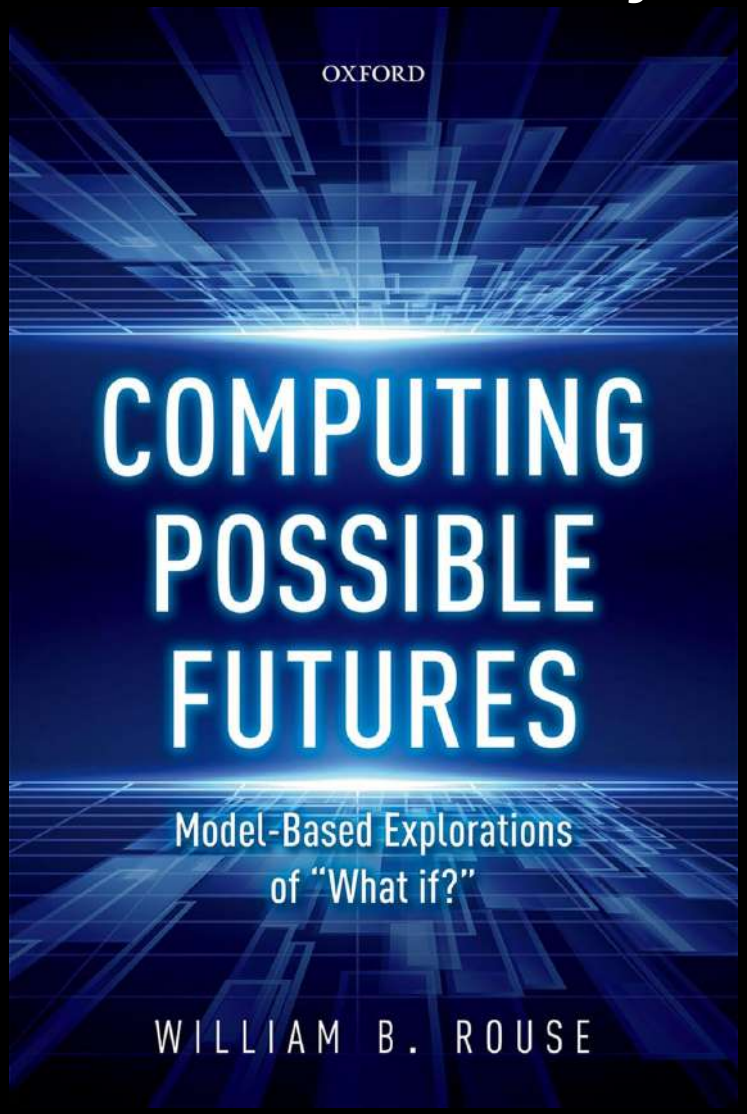
Questions

- In the turbulent currents of today's world, filled with disruptive intelligent technologies, how can we navigate to evade dystopic scenarios? (AI controlling humans, taking over, and eventually destroying humans. Humans with the help of AI enslaving other humans)
- How can we envision the broader landscape of a future human-centered digital society? What would human flourishing mean?
- What does a desirable future look like for both humans and our planet, steering towards common preferred futures/utopias?

Plan of the talk

- Navigating Possible Futures: Speculative Design
- Complexity & Systemic Thinking
- A White Water World & Emergence in Ecologies of Change
- Value-based Human-centric Design
- Digital Humanism
- A Case Study: Ethics Of Autonomous Cars
- Wrap-up

We are discussing possible futures with socially disruptive technologies



OF COURSE, PRESENT-DAY TECHNOLOGY CAN NOT BE NEGLECTED, LIKE FEMINIST APPROACHES AND CRITICAL DESIGN, BUT WE DO NOT FOCUS ON THAT.

Design for possible & preferable futures –
SPECULATIVE DESIGN

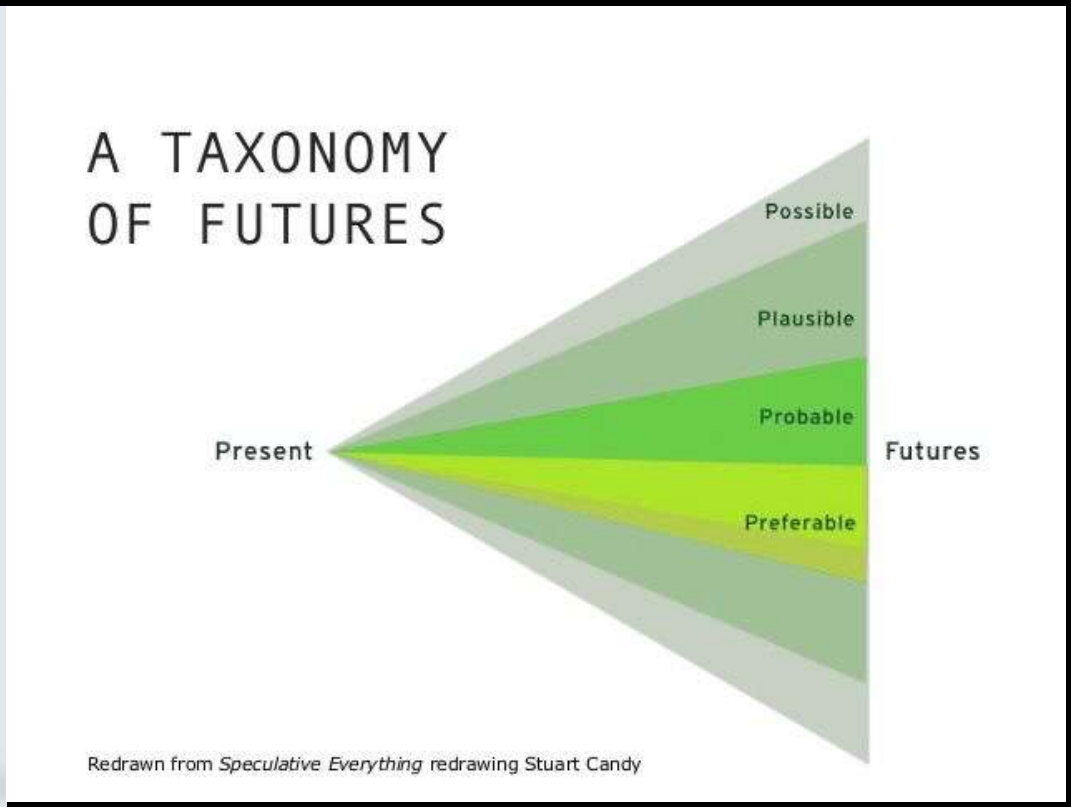
Speculative design combines informed, hypothetical extrapolations of an emerging technology's development with a deep consideration of the cultural landscape into which it might be deployed, to speculate on future products, systems and services.

These speculations are then used to examine and encourage dialogue on the impact a specific technology may have on our everyday lives.

Auger Loizeau

https://elviavasconcelosblog.wordpress.com/2017/01/15/what-is-speculative-critical-fiction-design-part-1/

Speculative Everything – Antony Dunne and Fiona Raby

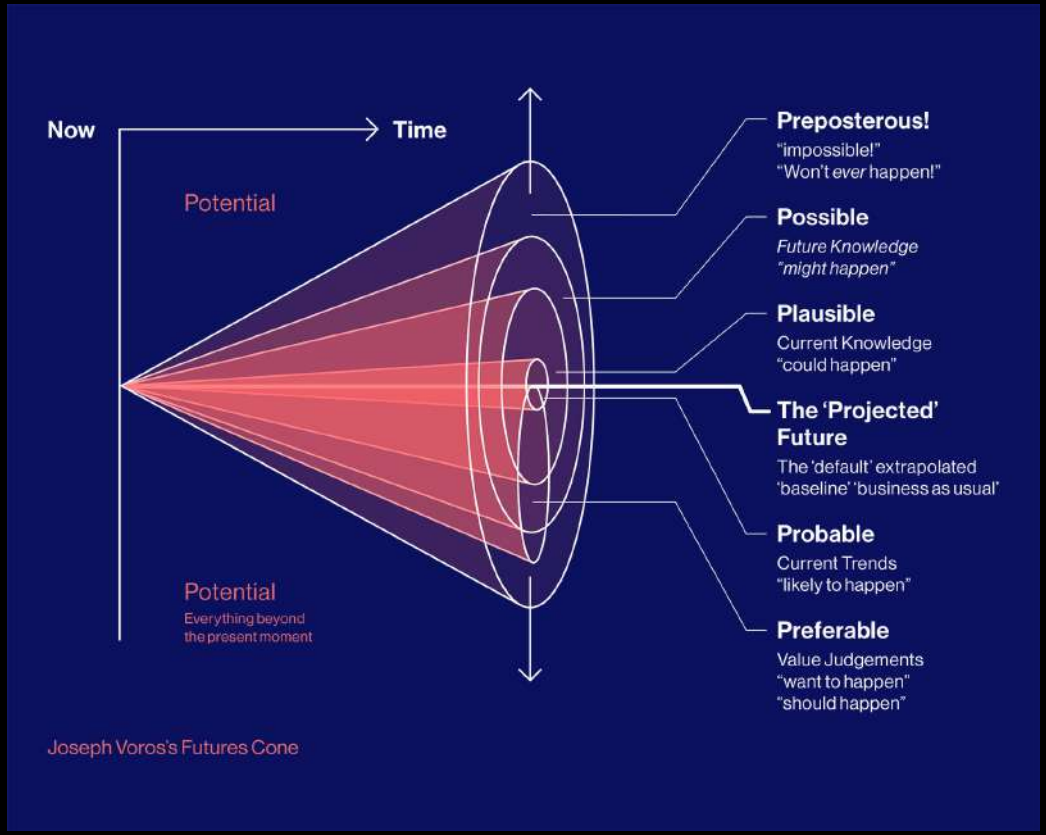


"what if" questions

https://www.youtube.com/watch?v=kmibm20UsoA

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.

Speculative Design creates space to…



Arrange emerging (not yet available) technological 'elements' to hypothesize future, products and artifacts.

Apply alternative plans, motivations, or ideas to those currently driving technological development, in order to facilitate new arrangements of existing elements.

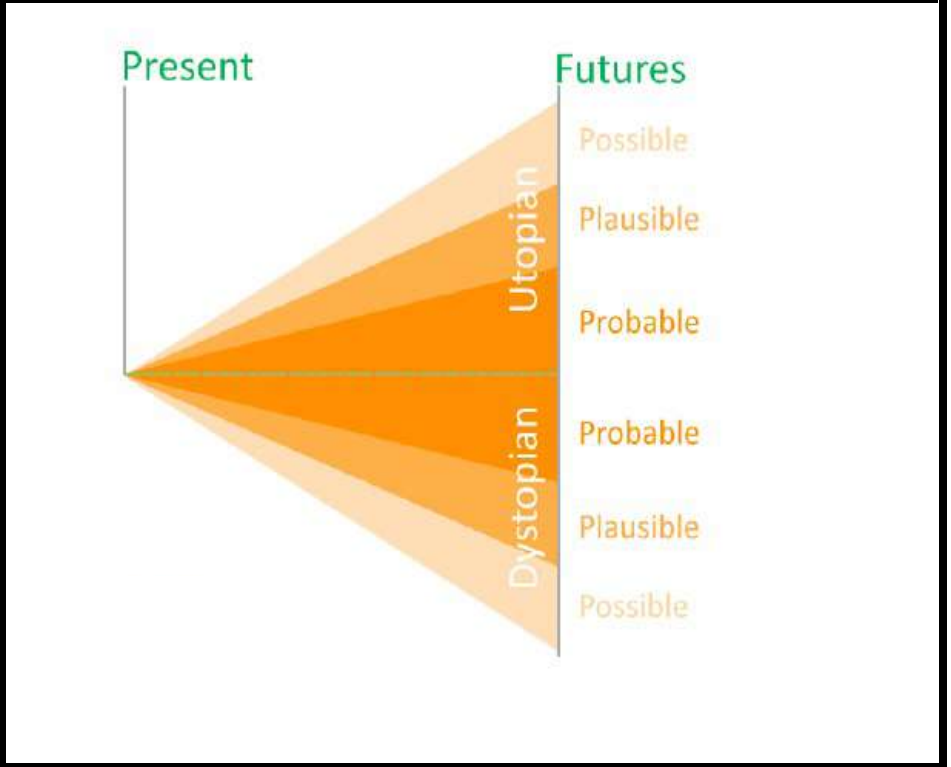
Develop new perspectives on big systems.

Speculative Design Facilitates…

Exploration of 'What is a better future (with respect to the present)?'

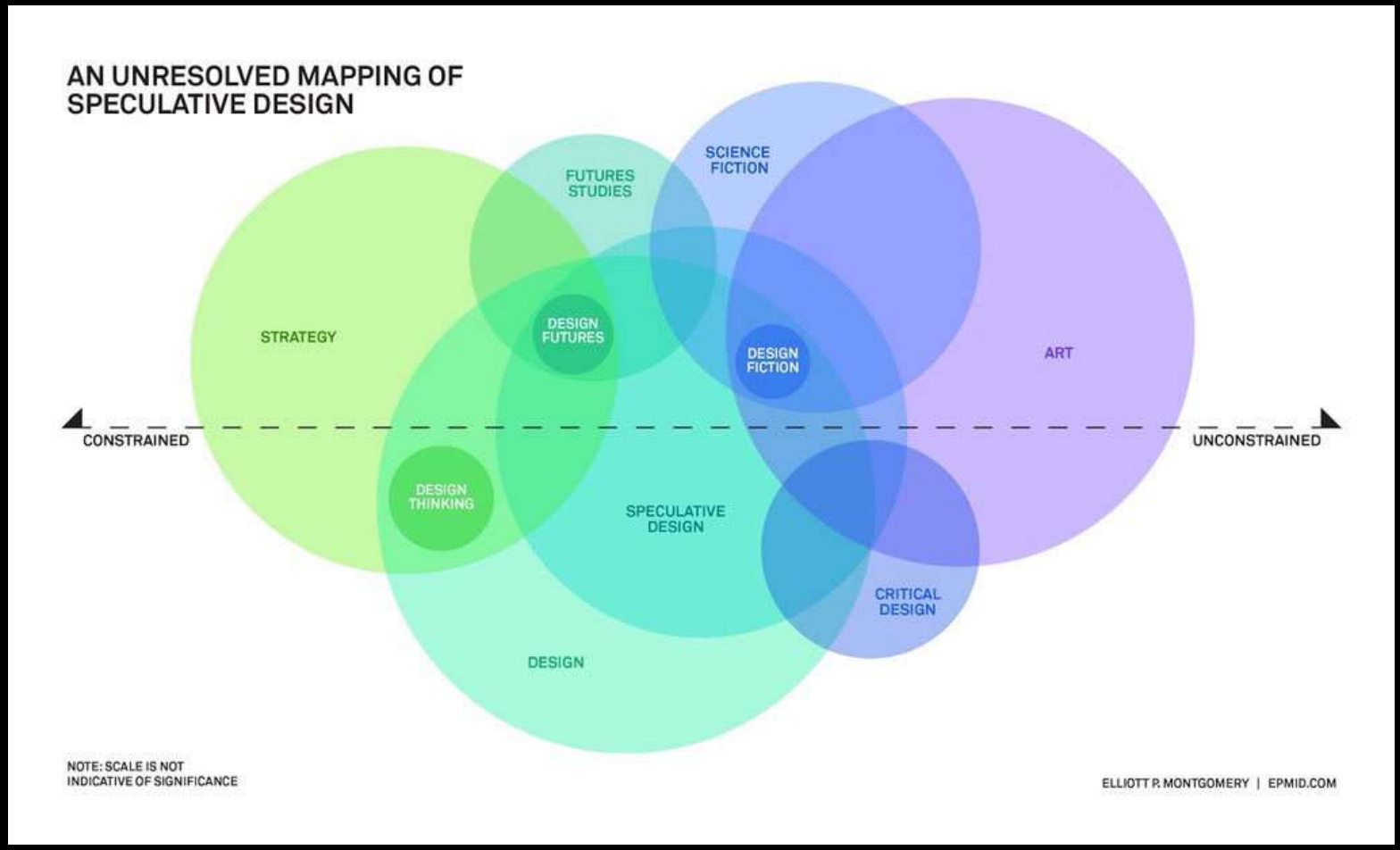
Generating a better understanding of the potential implications of a specific (disruptive) technology in various contexts and on multiple scales – with a particular focus on everyday life.

Moving design 'upstream' – to not simply package technology at the end of the technological journey but to impact and influence that journey from its genesis.



Giovanni M Troiano, Matthew Wood, Mustafa Feyyaz Sonbudak, Riddhi Chandan Padte, and Casper Harteveld. 2021. "Are We Now Post-COVID?": Exploring Post-COVID Futures Through a Gamified Story Completion Method. In Proceedings of the 2021 ACM Designing Interactive Systems Conference (DIS '21). ACM, New York, NY, USA, 48–63. https://doi.org/10.1145/3461778.3462069

Speculative Design and its context



https://speculativeedu.eu/new-reflections-on-speculativity/

Complexity & systemic thinking in hyper-connected society

https://humanparts.medium.com/making-sense-of-complexity-ee78755d56b9

ADD TO THIS PICTURE (INTELLIGENT) INTERNET OF THINGS!

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=-U8h4wNBfCQ
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.

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."

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

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

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

'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.

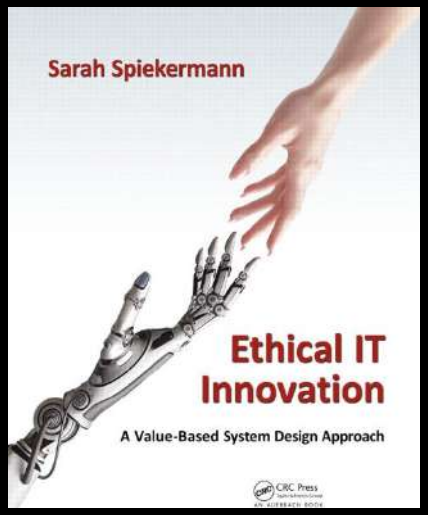
VALUE-BASED HUMAN-CENTRIC DESIGN

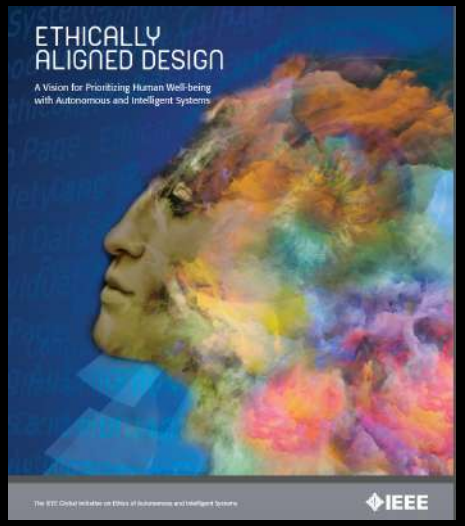
Values

Values serve as a guide to action and knowledge.

They are relevant to all aspects of scientific and engineering practice, including discovery, analysis, and application.

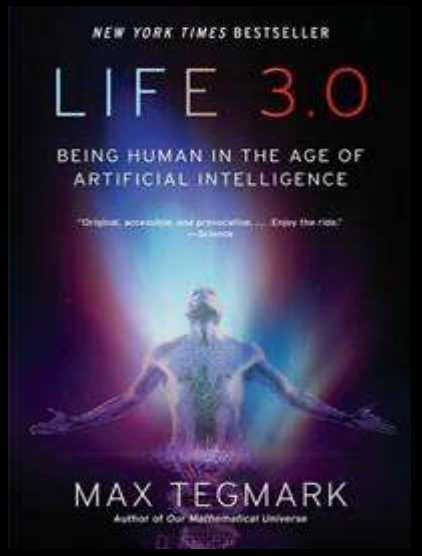
A VALUE-BASED DESIGN APPROACH

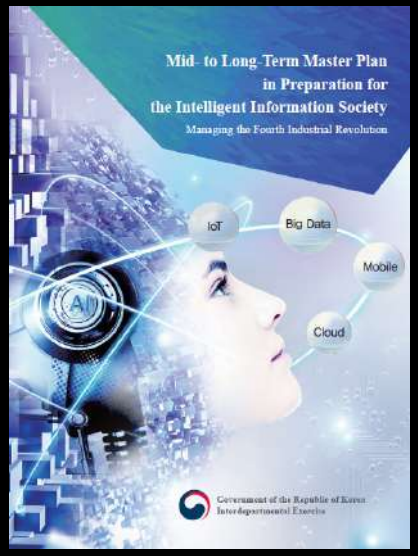


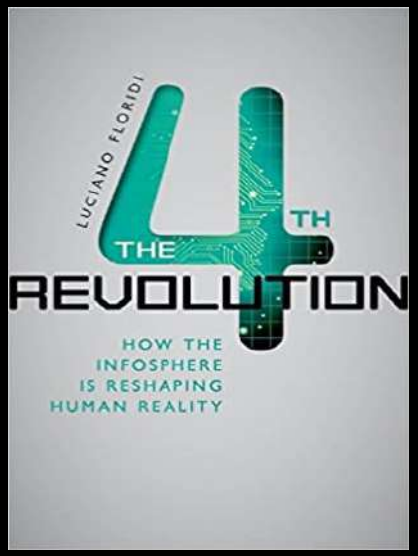


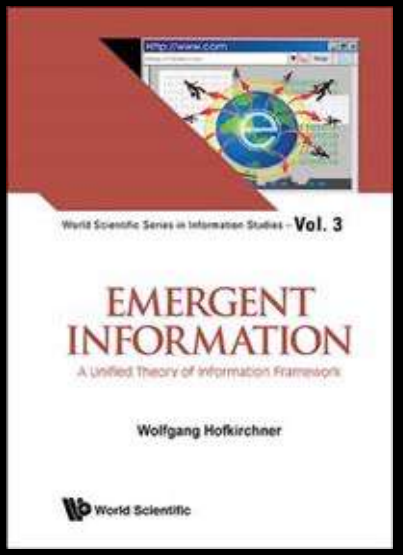
One question we can ask is: How much time can we afford to spend on the "ideation phase" before starting to actually build technology?
Andrew Ng points out for a startup it is more profitable to identify which technology can be built, and then go and build it, instead of spending a lot of time thinking about all possible alternatives: https://www.youtube.com/watch?v=5p248yoa3oE (29:08)

Human-centered future intelligent society









"In the Fourth Industrial Revolution, the convergence of artificial intelligence, robot technology, big data and software disrupts fields such as labor, welfare, employment, education and defense. This has sparked revolutionary change across society."

Wikipedia, https://en.wikipedia.org/wiki/Intelligent_information_society

The Digital Humanism Initiative

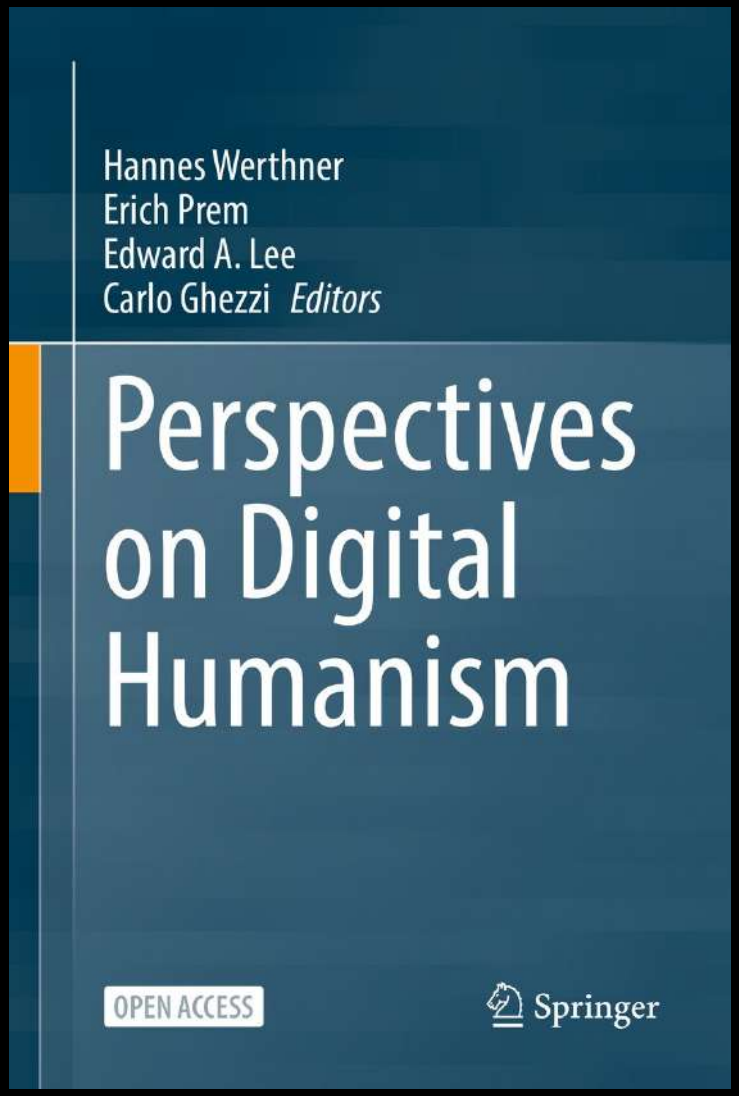
The Digital Humanism Initiative is an international collaboration seeking to build a community of scholars, policy makers, and industrial players who are focused on ensuring that technology development remains centered on human interests.

- 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.
- 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/

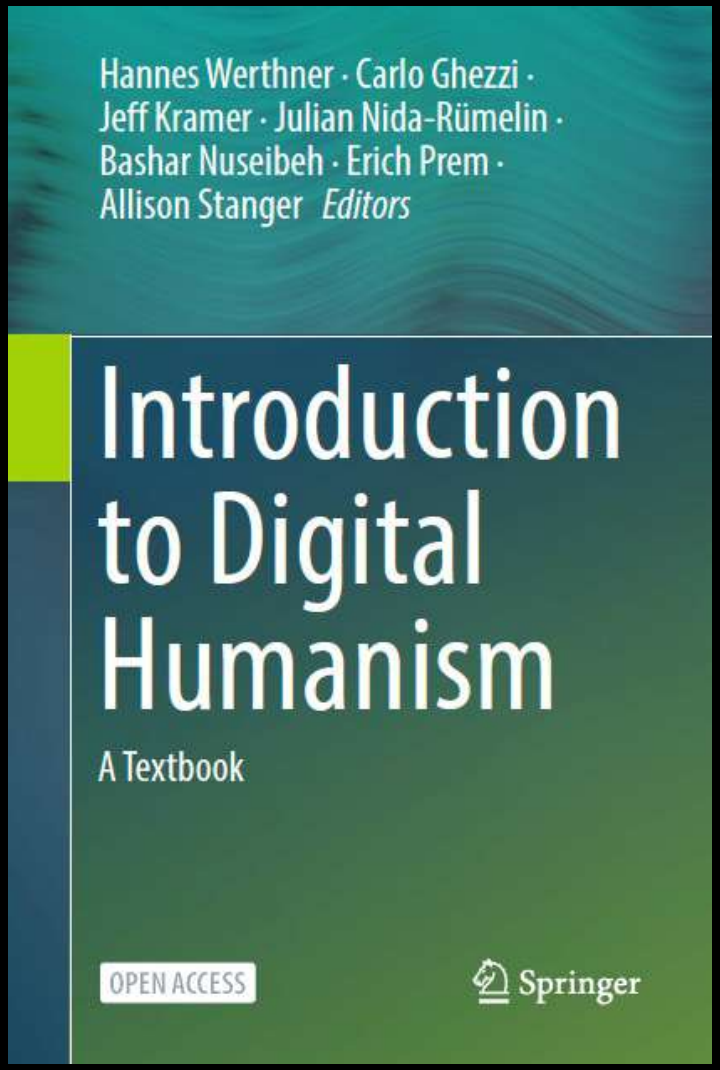
Perspectives on Digital Humanism – Open Access



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

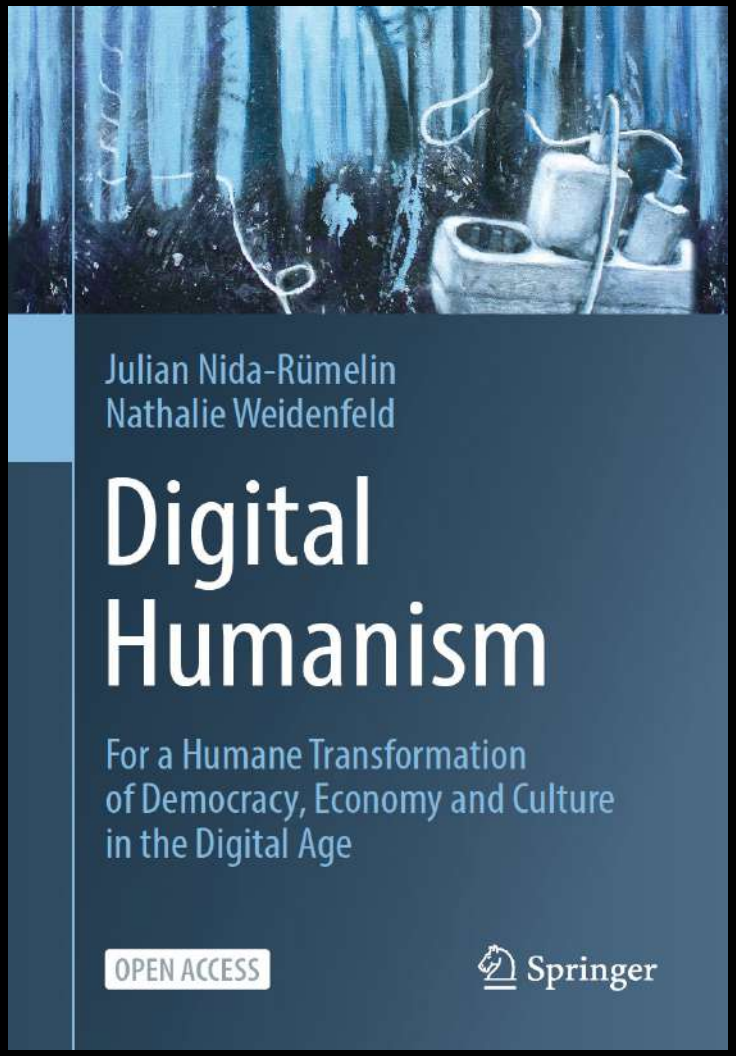
Open Access



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.”

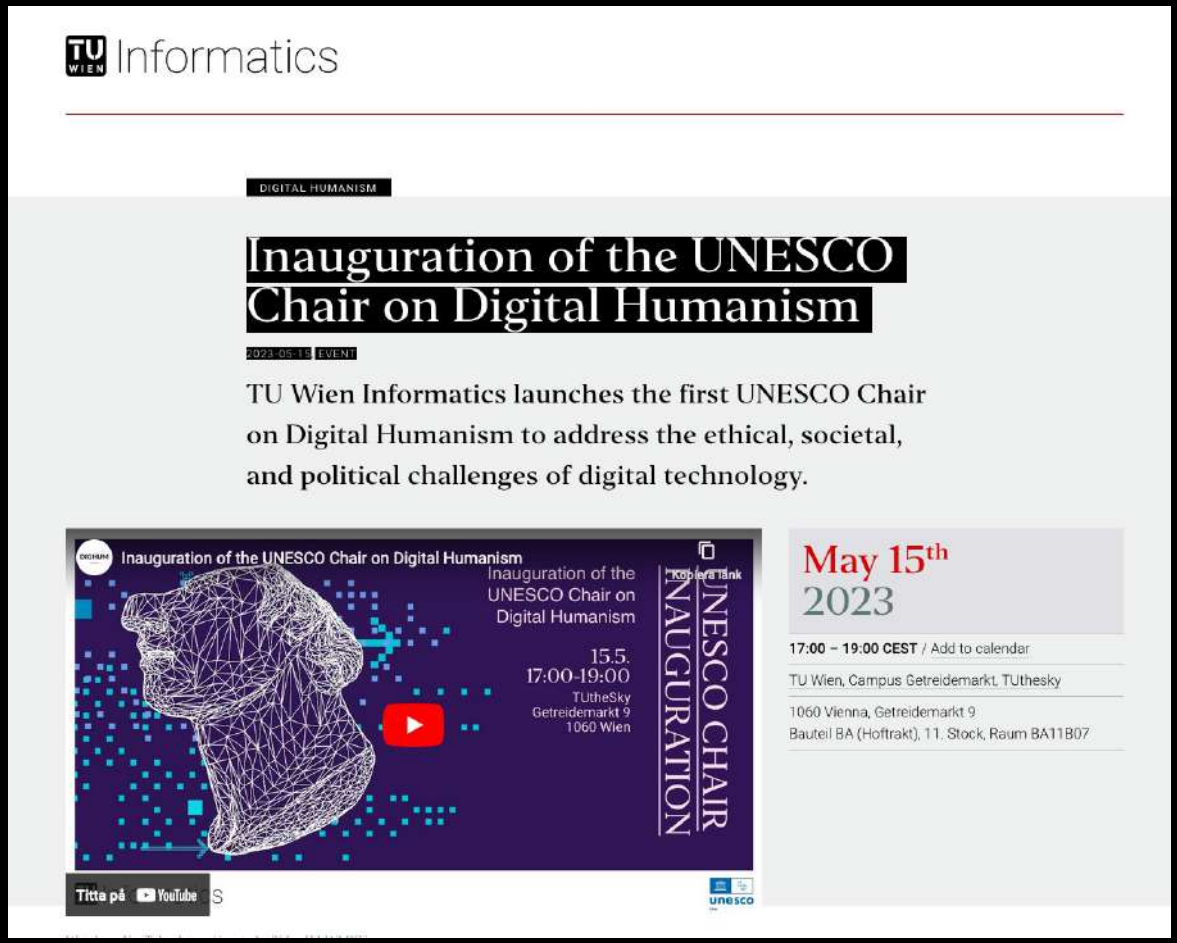
https://dighum.ec.tuwien.ac.at/dighum-manifesto/

Viable Initiatives in a Hyperconnected, Dynamic, Emergent World

https://medium.com/@bill.bannear/the-new-zeitgeist-relationships-and-emergence-e8359b934e0

UNESCO Chair on Digital Humanism

Peter Knees Chair and Julia Neidhardt Co-Chair

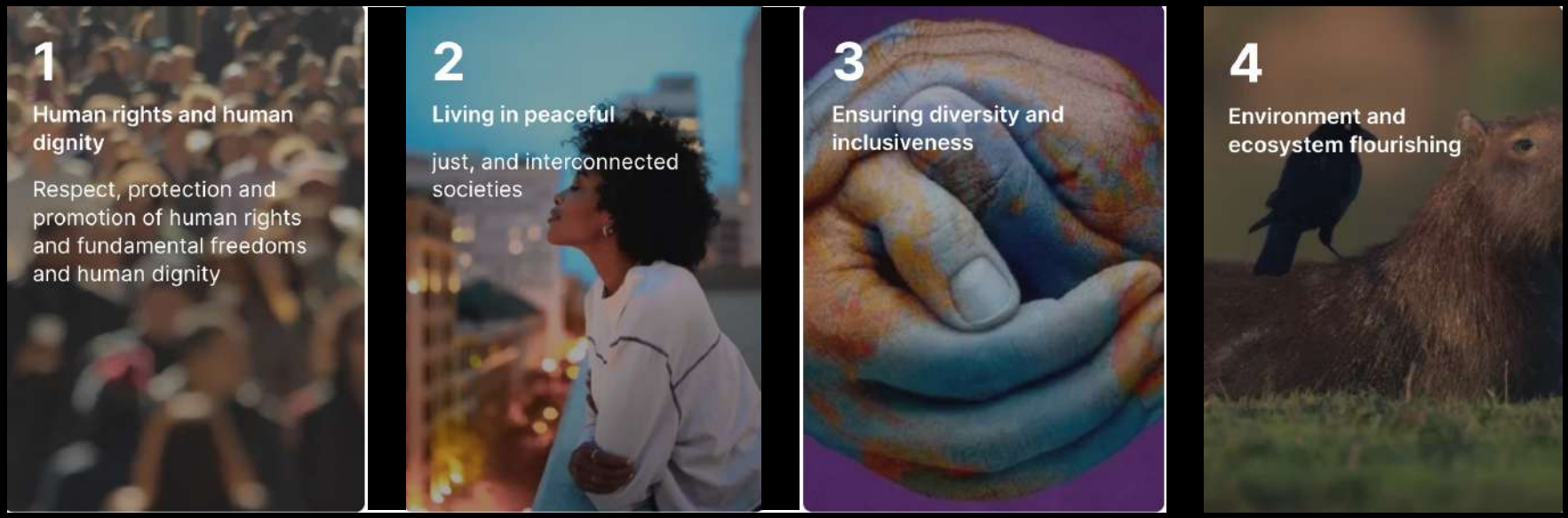


"UNESCO uses education, science, culture, communication and information to foster mutual understanding and respect for our planet."

CAIML - Center for Artificial Intelligence and Machine Learning. https://www.tuwien.at/caiml/

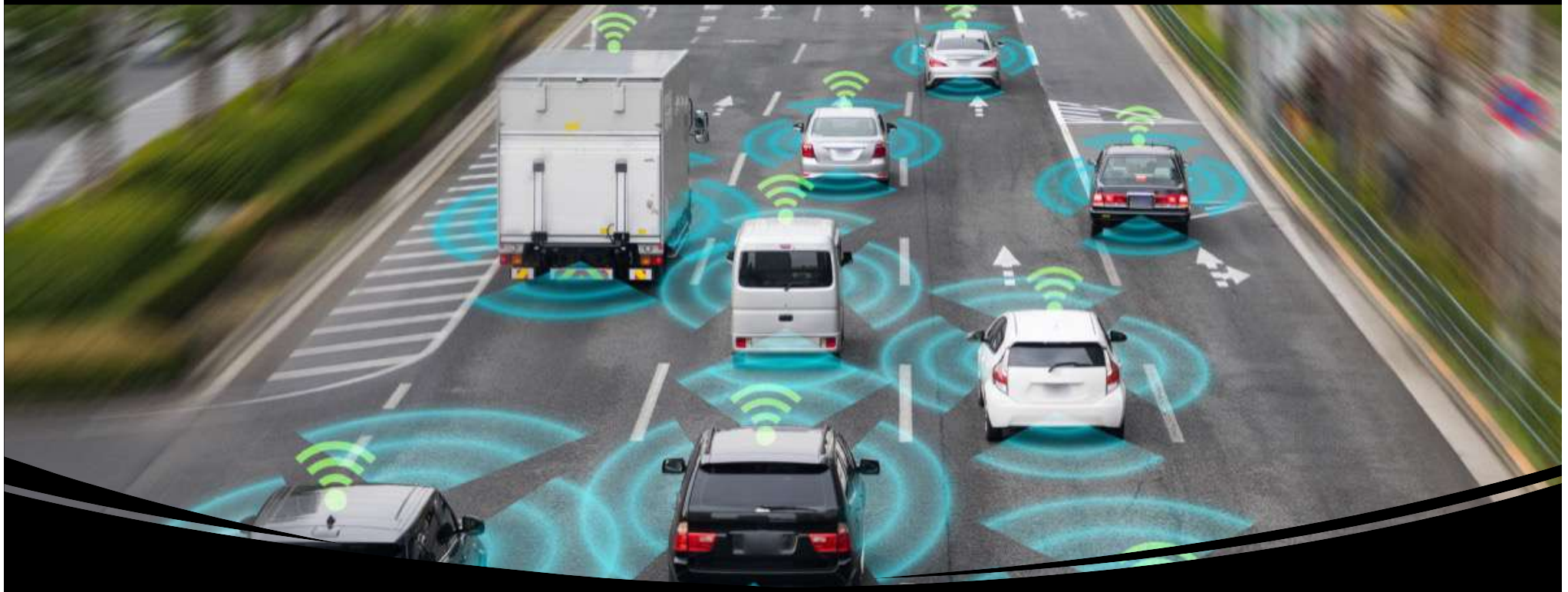
https://informatics.tuwien.ac.at/stories/2383

UNESCO ‘Recommendation on the Ethics of Artificial Intelligence’



https://www.unesco.org/en/artificial-intelligence/recommendation-ethics

Case study - Autonomous Cars Ethics



Autonomous cars

As a special case of intelligent emerging technology

Book chapter:
"Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem".

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

Picture: https://www.aarete.com/insights/what-is-the-business-case-for-autonomous-vehicles-in-the-supply-chain/

Safety

Challenges

- Hardware and software adequacy
- Vulnerabilities of machine-learning algorithms
- Control of trade-offs between safety and other factors (like economic) in the design, manufacturing and operation
- Possibility of intervention in case of major failure of the system and graceful degradation
- Systemic solutions to guarantee safety in organizations (regulations, authorities, safety culture)

Approaches

- Setting safety as the first priority
- Learning from the history of automation
- Learning from experience of current use
- Specification of how a system will behave in cases when autonomous operation is disabled (safe mode)
- Preparedness for handling “loss of control” situations- autonomous systems running amok
- Regulations, guidelines, standards being developed as the technology develops

Security

Challenges

- Minimal necessary security requirements for deployment of the system
- Security in the context and connections
- Deployment of software updates
- Storing and using received and generated data in a secure way

Approaches

- Technical solutions to guarantee minimum security under all foreseeable circumstances
- Anticipation and prevention of the worst-case scenarios
- Accessibility of data, even in the case of accidents, learning from experience

Non-maleficence

Challenges

- Risk of technology causing harm, physical, cognitive, psychological, social, etc.
- Disruptive changes in the labor market
- Transformation of related businesses, markets, and business models (manufacturers, insurance, etc.)
- Loss of human skills
- Loss of autonomy

Approaches

- Partly covered by technical solutions, but interdisciplinary approaches are needed
- Preparation of strategic solutions for people losing jobs
- Learning from historic parallels to industrialization and automatization

Responsibility and Accountability

Challenges

- Assignment and distribution of responsibility and accountability as some of central regulative mechanisms for the development of new technology

Approaches

- The Accountability, Responsibility, and Transparency (ART) principle (Virginia Dignum) based on a Design for Values approach that includes human values and ethical principles in the design processes

Stakeholders Interests

Humans in the loop

Freedom of choice

To what extent will the user be in control?

Will the AI do, what I want it to do?

Implementation of restrictions

Loss of jobs compensation

Impacts on society as a whole

Social Trust

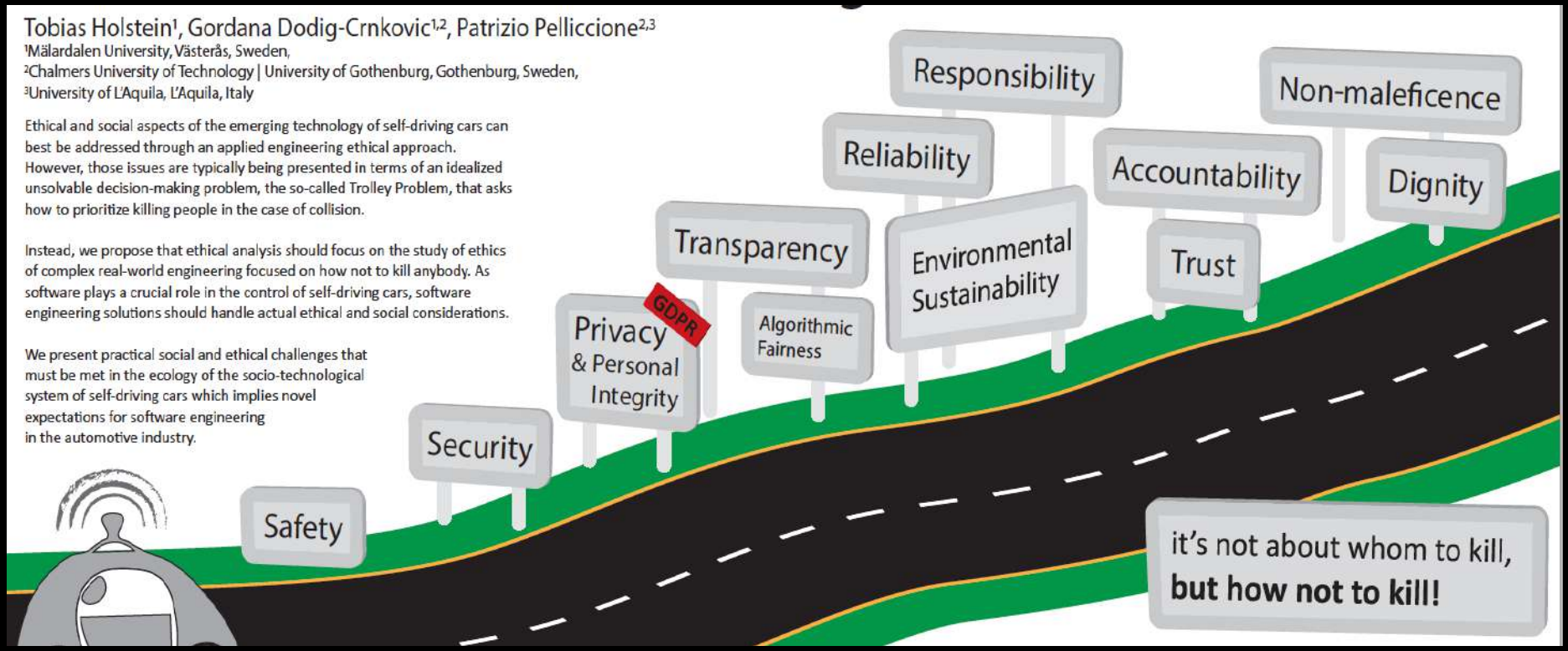
Challenges

- Establishing trust between humans and robots as well as within the social system involving robots

Approaches

- Further research on how to implement trust across multiple systems
- Provision of trusted connections between components as well as external services

Value-based Ethical Guidelines for Self-Driving Cars



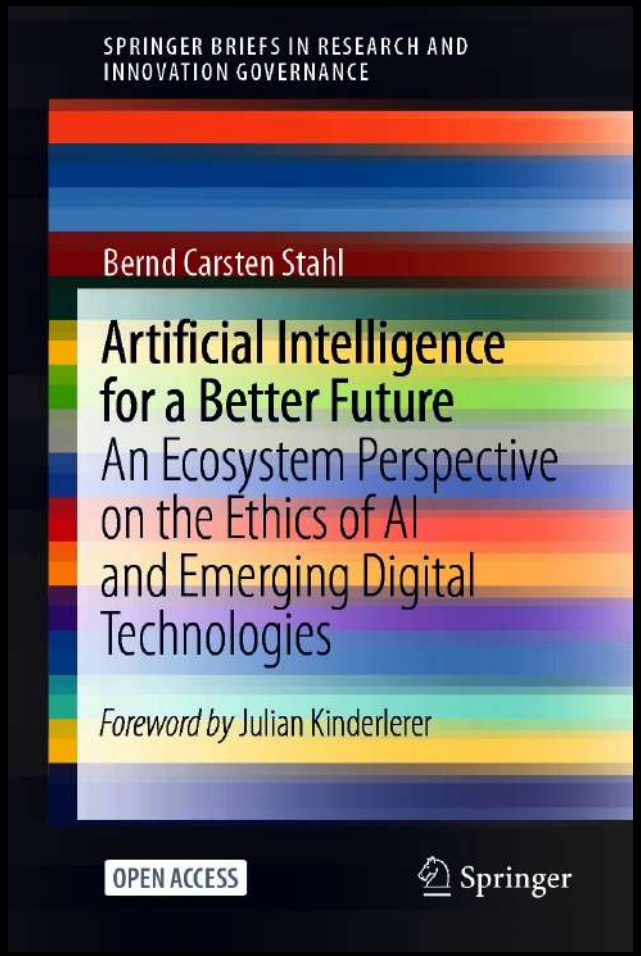
Ehics of Self-Driving Cars

Presented at major SE conference ICSE2020 as poster

Extended version in a book chapter:

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

Our Future with AI



AI FOR A BETTER FUTURE

An Ecosystem Perspective
on the Ethics of AI
and Emerging Digital Technologies

Bernd Carsten Stahl

https://link.springer.com/book/10.1007/978-3-030-69978-9 OPEN ACCESS

Organizational Ethical Issues of AI

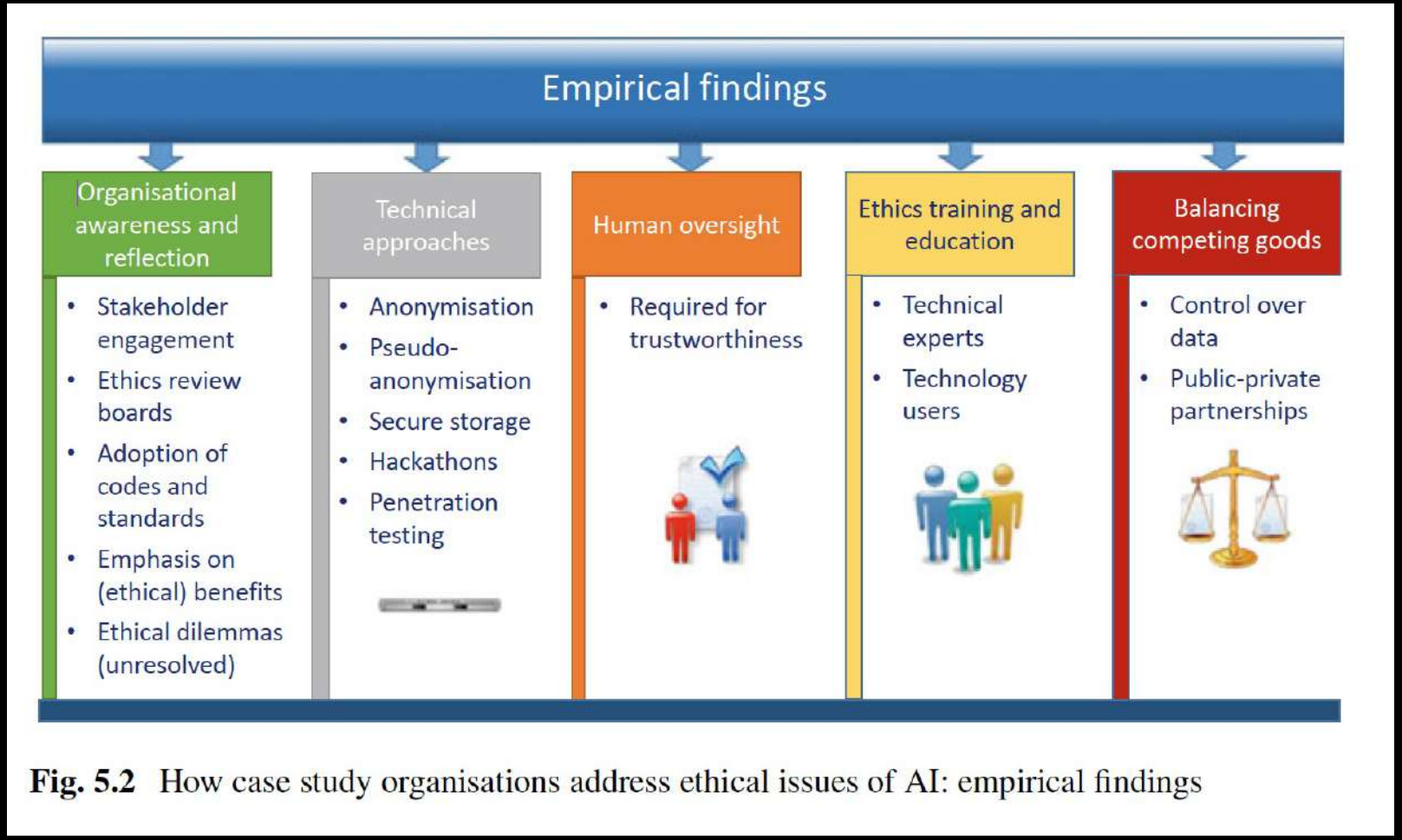
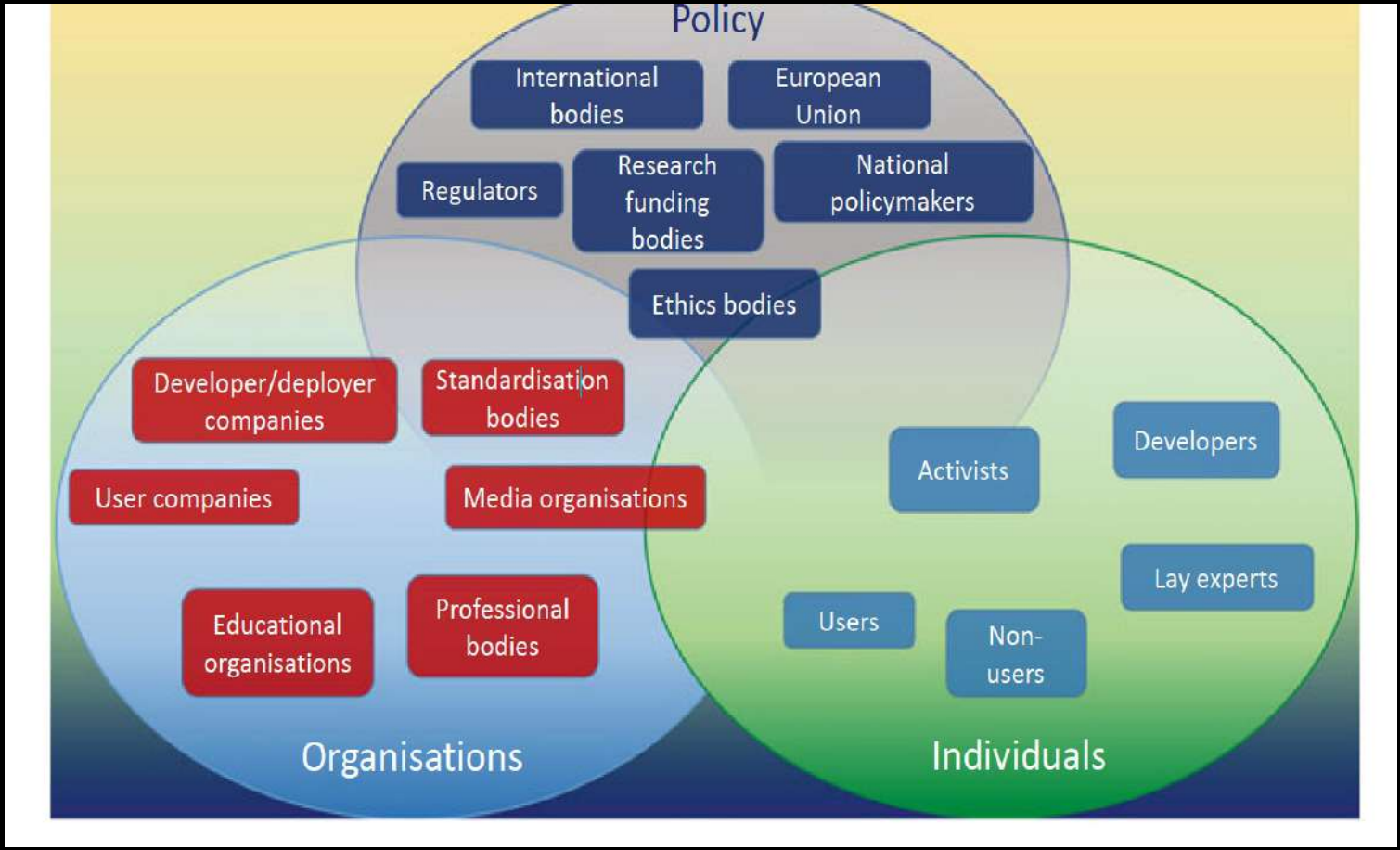


Fig. 5.2 How case study organisations address ethical issues of AI: empirical findings

Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies https://link.springer.com/book/10.1007%2F978-3-030-69978-9

Overview of AI stakeholders



Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9

Key Challenges of Ethical Governance of AI

Fig. 7.1 Key challenges of ethical governance of AI ecosystems

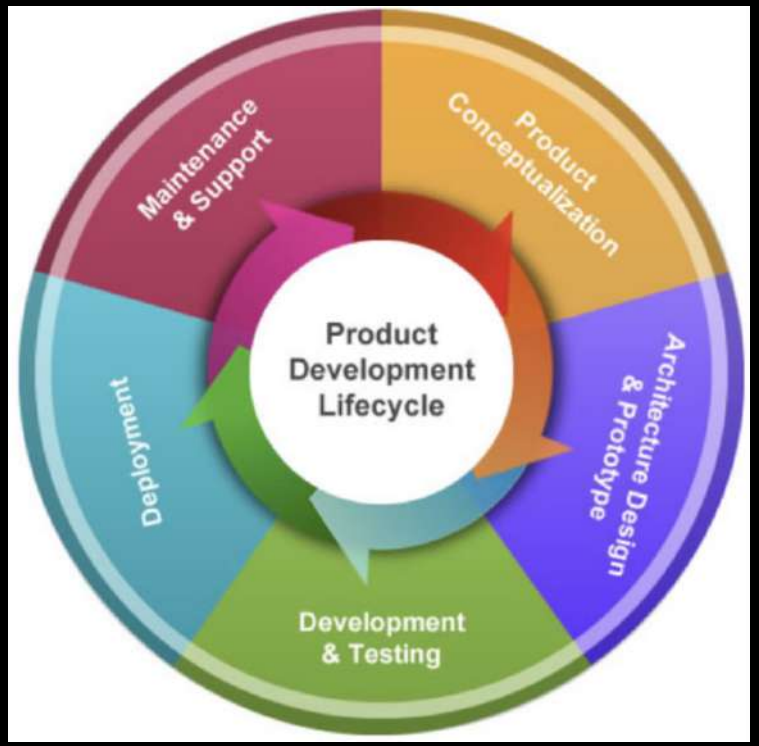
Bernd Carsten Stahl (2021) Artificial Intelligence for a Better Future, https://link.springer.com/book/10.1007%2F978-3-030-69978-9

Practical Use of the Proposed Ethical Program for Intelligent Emergent Technologies -

Importance of Transdisciplinarity and Transversal Knowledge

Ethical requirements must be fulfilled in all phases in the life-cycle of technology, in the context of:

- Conceptualization/Design/Prototyping/ Construction/Development/Testing/Production
- Deployment/Application/
- Maintenance/Support
- Oversight/Regulation



Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

https://www.indiamart.com/kaynes-tecnology-india-pvt-ltd/product-life-cycle-management.html

Challenges for Emergent Technologies

Legislation	Global framework	Guidelines	Implementation
Keeping legislation up-to-date with current level of automated driving, and emergence of self-driving cars	Creating and defining global legislation frameworks for the implementation of interoperable and development of increasingly automated vehicles	Defining the guidelines that will be adopted by society for building self-driving cars	Including ethical guidelines in design and development processes

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

Building Ethical Technology in an Ethical Way

Work on the shared vision of emergent technologies.

Anticipation and consideration of uncertainties/Speculative design

A system-level approach involving the entire software-hardware system as well as human stakeholders, with organizational, and social factors.

Multi-criteria decisions. Multidisciplinary approach.

Learning from experience from the whole life cycle of technology.

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global

Ethical Lessons of Artificial Intelligence

Responsibility in AI Development: recognizing the responsibility of developers and engineers to create AI systems that are not only effective but also fair, transparent, and non-discriminatory.

Impact on Society: There are lessons to be learned regarding the societal impact of AI, such as the potential for job displacement, privacy concerns, and changes in social dynamics.

Bias and Fairness: AI can inadvertently perpetuate or amplify existing biases if not carefully designed and monitored. Understanding and addressing these issues is a crucial ethical lesson.

Responsibility in AI Development: recognizing the responsibility of developers and engineers to create AI systems that are not only effective but also fair, transparent, and non-discriminatory.

Transparency and Explainability: As AI systems become more complex, ensuring that they are transparent, and their decisions can be explained and understood by humans is an important ethical consideration.

Accountability: Establishing clear lines of accountability for AI's decisions and actions, particularly when they lead to harm or injustice, is an ethical challenge that must be addressed.

Safety and Security: Ensuring that AI systems are safe from malicious uses and are secure against potential breaches is an ongoing ethical concern.

Regulation and Governance: Determining the appropriate level of regulation and the governance structures needed to oversee AI development and implementation is an essential ethical lesson.

Beneficence and Nonmaleficence: AI should be designed and used in ways that benefit people and society at large while avoiding harm, reflecting these core ethical principles.

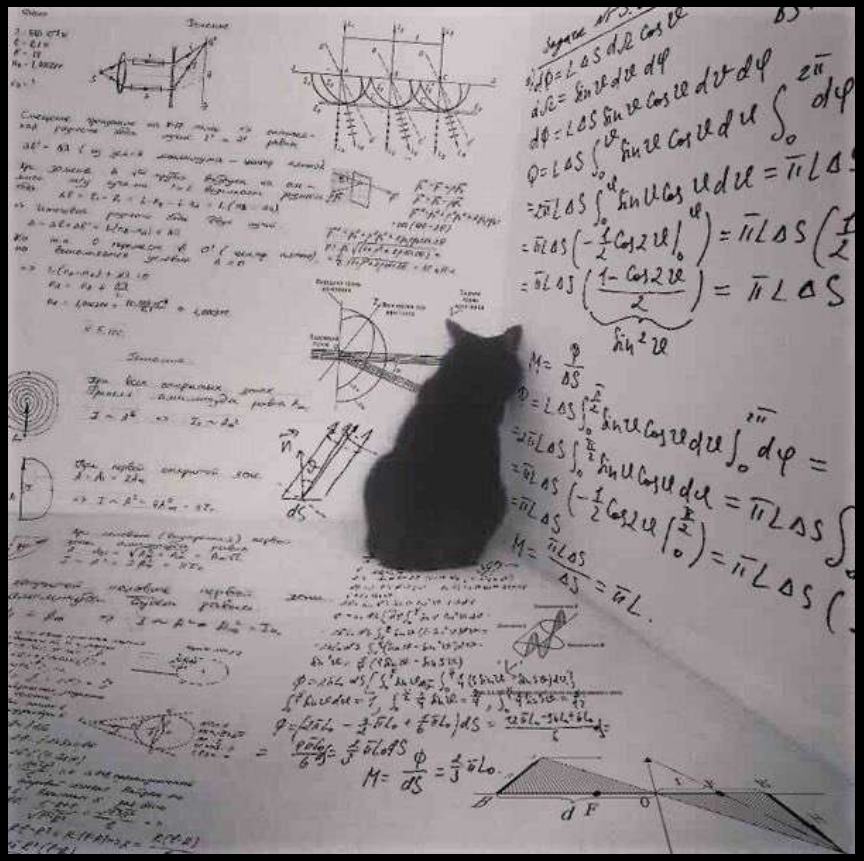
Wrap-up

The main topics we visited during this talk

- Navigating Possible Futures: Speculative Design
- A White Water World & Emergence in Ecologies of Change
- Value-based Human-centric Design
- Digital Humanism
- Case Study: Ethics of Autonomous Cars

As AI technology becomes more and more powerful, the age-old wisdom applies: "With great power comes great responsibility."

The perspective of Digital Humanism was presented as a way of approaching the contemporary white-water world, driven by the prospect of a more humane and inclusive future.



Q & A TIME!

References

- G. Dodig-Crnkovic, T. Holstein, P. Pelliccione and, Jathoosh Thavarasa (2023) "Future Intelligent Autonomous Robots, Ethical by Design. Lessons Learned from Autonomous Cars Ethics." Proc. ICSIT 2023 conference. ISSN: 2771-6368 (Print) ISBN: 978-1-950492-70-1 (Print) DOI: 10.54808/ICSIT2023.01 https://www.iiis.org/CDs2023/CD2023Spring//
- Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. In Steven John Thompson (Ed.), Machine Law, Ethics, and Morality in the Age of Artificial Intelligence. IGI Global
- 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. https://ethics.se
- Holstein, T. Dodig-Crnkovic G. Avoiding the Intrinsic Unfairness of the Trolley Problem. Accepted for the Proceedings of FairWare workshop at ICSE2018, to be published by ACM.
- Holstein, T. Dodig-Crnkovic G. and Pelliccione P. Ethical and Social Aspects of Self-Driving Cars, http://arxiv.org/abs/1802.04103
- Dodig Crnkovic, G. and B. Çürüklü. Robots: ethical by design. Ethics and Information Technology, 14(1):61–71, Mar 2012.
- Dodig Crnkovic, G. and B. Çürüklü. Robots: ethical by design. Ethics and Information Technology, 14(1):61–71, Mar 2012.
- Dodig-Crnkovic, G. and D. Persson. Sharing moral responsibility with robots: A pragmatic approach. In Proceedings of the 2008 Conference on Tenth Scandinavian Conference on Artificial Intelligence: SCAI 2008, pages 165–168, Amsterdam, The Netherlands, IOS Press. 2008.
- Dodig-Crnkovic, G. and D. Persson. Sharing moral responsibility with robots: A pragmatic approach. In Proceedings of the 2008 Conference on Tenth Scandinavian Conference on Artificial Intelligence: SCAI 2008, pages 165–168, Amsterdam, The Netherlands, IOS Press. 2008.
- Johnsen A., G. Dodig- Crnkovic, K. Lundqvist, K. Hänninen, and P. Pettersson. Risk- based decision-making fallacies: Why present functional safety standards are not enough. In 2017 IEEE International Conference on Software Architecture Workshops (ICSAW), pages 153–160, April 2017.
- Sapienza, G., Dodig-Crnkovic, G. and I. Crnkovic. Inclusion of ethical aspects in multi-criteria decision analysis. In 2016 1st International Workshop on Decision Making in Software ARCHitecture (MARCH), pages 1–8, April 2016.
- Thekkilakattil A. and G. Dodig-Crnkovic. Ethics aspects of embedded and cyber-physical systems. In 2015 IEEE 39th Annual Computer Software and Applications Conference, volume 2, pages 39–44, July 2015.
- Margarita Georgieva (student) and Gordana Dodig-Crnkovic (2011) Who Will Have Irresponsible, Untrustworthy, Immoral Intelligent Robot? Proceedings of IACAP 2011. The Computational Turn: Past, Presents, Futures?, p 129, Mv-Wissenschaft, Münster, Århus University, Danmark, Ed. Ess and Hagengruber, July 201

- Regulation (E.U.) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016. https://eur-lex.europa.eu/eli/reg/2016/679/oj
- euRobotics Topics Group 'Ethical Legal and Socio-Economic Issues.' Policy Documents & Institutions - ethical, legal, and socio-economic issues of Robotics and artificial intelligence, 2022. https://www.pt-ai.org/TG-ELS/
- F. Operto, "Ethics in Advanced Robotics," IEEE Robot. Autom. Mag., vol. 18, no. 1, pp. 72–78, Mar. 2011.
- N. Leveson, "Are You Sure Your Software Will Not Kill Anyone?," Commun. ACM, vol. 63, no. 2, pp. 25–28, Jan. 2020. https://dl.acm.org/doi/10.1145/3376127
- P. Lin, K. Abney, and G. A. Bekey, Robot Ethics: The Ethical and Social Implications of Robotics. MIT Press, 2011. http://kryten.mm.rpi.edu/Divine-Command_Roboethics_Bringsjord_Taylor.pdf
- P. M. Asaro, "What should we want from a robot ethic?," in Machine Ethics and Robot Ethics, 2017. https://peterasaro.org/writing/Asaro%20IRIE.pdf
- P. M. Asaro, Autonomous Weapons and the Ethics of Artificial Intelligence," in S. Matthew Liao (ed.) Ethics of Artificial Intelligence, Oxford University Press, pp. 212-236. https://global.oup.com/academic/search?q=ethics+of+artificial+intelligence&cc=us&lang=en
- S. G. Tzafestas, Roboethics - A Navigating Overview, vol. 79. Springer International Publishing, 2016. https://link.springer.com/book/10.1007/978-3-319-21714-7
- V. C. Müller, "Ethics of Artificial Intelligence and Robotics," in The Stanford Encyclopedia of Philosophy (Summer 2021 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2021/entries/ethics-ai/
- W. Wallach and C. Allen, Moral Machines: Teaching Robots Right from Wrong. New York: Oxford University Press, 2009. https://academic.oup.com/book/10768
- https://www.ethics.se ETHICS & SELF-DRIVING CARS
- Baran Çürüklü, Gordana Dodig-Crnkovic, Batu Akan (2010) Towards Industrial Robots with Human Like Moral Responsibilities, 5th ACM/IEEE International Conference on Human-Robot Interaction, Osaka, Japan, March 2010
- Gordana Dodig-Crnkovic (2010) Information Ethics for Robotic Agents European Computing and Philosophy Conference ECAP 2010 @The Technische Universität München, 4-6 October, 2010
- Gordana Dodig-Crnkovic (2009) Delegating Responsibilities to Intelligent Robots. ICRA2009 IEEE International Conference on Robotics and Automation. Workshop on Roboethics Kobe, Japan, May 17, 2009.
- Gordana Dodig-Crnkovic and Daniel Persson (student) (2008) Sharing Moral Responsibility with Robots: A Pragmatic Approach. Tenth Scandinavian Conference on Artificial Intelligence, SCAI 2008.
Volume 173, Frontiers in Artificial Intelligence and Applications. Eds. A. Holst, P. Kreuger and P. Funk
- Gordana Dodig-Crnkovic and Daniel Persson (student) (2008) Towards Trustworthy Intelligent Robots, NA-CAP@IU 2008, North American Computing and Philosophy Conference, Indiana University, Bloomington, July 10-12, 200

Digital Humanism References

https://www.youtube.com/watch?v=V-XvfMEZgPc The Challenge of Being Humanely Digital - UCAI '22 Keynote by Erich Prem

https://informatics.tuwien.ac.at/digital-humanism/

https://dighum.ec.tuwien.ac.at

https://link.springer.com/book/10.1007/978-3-030-86144-5 Perspectives on Digital Humanism – book freely available for download

https://dighum.ec.tuwien.ac.at/dighum-manifesto/ Vienna Manifesto on Digital Humanism

https://nextconf.eu/2017/11/what-is-digital-humanism/#gref

https://www.erichprem.at/publications-press-videos/ Erich Prem videos