Workshop on human-centric cybersecurity - HCC2023
Mälardalen University, Västerås, Sweden
August 24th, 2023 https://hcc2023-mdu.github.io/

Robots Ethical by Design in the Perspective of Digital Humanism

Gordana Dodig Crnkovic

Professor of Computer Science

Chalmers Technical University | University of Gothenburg & Mälardalen University, Sweden

http://gordana.se/

http://www.gordana.se/work/presentations

Based on the work of
Gordana Dodig Crnkovic,
Baran Çürüklü,
Tobias Holstein
JathooshThavarasa

https://techdissected.com/editorials-and-discussions/advantages-of-having-a-humanoid-robot

The article we revisited

Ethics Inf Technol (2012) 14:61–71
DOI 10.1007/s10676-011-9278-2

ORIGINAL PAPER

Robots: ethical by design

Gordana Dodig Crnkovic · Baran Çürüklü

Published online: 24 August 2011
© Springer Science+Business Media B.V. 2011

Abstract Among ethicists and engineers within robotics there is an ongoing discussion as to whether ethical robots are possible or even desirable. We answer both of these questions in the positive, based on an extensive literature study of existing arguments. Our contribution consists in bringing together and reinterpreting pieces of information from a variety of sources. One of the conclusions drawn is that artifactual morality must come in degrees and depend on the level of agency, autonomy and intelligence of the machine. Moral concerns for agents such as intelligent search machines are relatively simple, while highly intelligent and autonomous artifacts with significant impact and complex modes of agency must be equipped with more advanced ethical capabilities. Systems like cognitive robots are being developed that are expected to become part of our everyday lives in future decades. Thus, it is necessary to ensure that their behaviour is adequate. In an analogy with artificial intelligence, which is the ability of a machine to perform activities that would require intelligence in humans, artificial morality is considered to be the ability of a machine to perform activities that would require morality in humans. The capacity for artificial (artifactual) morality, such as artifactual agency, artifactual responsibility, artificial intentions, artificial (synthetic) emotions, etc., come in varying degrees and depend on the type of agent. As an illustration, we address the assurance of safety in modern High Reliability Organizations through responsibility distribution. In the same way that the concept of *agency* is generalized in the case of *artificial agents*, the concept of *moral agency*, including *responsibility*, is generalized too. We propose to look at artificial moral agents as having functional responsibilities within a network of distributed responsibilities in a socio-technological system. This does not take away the responsibilities of the other stakeholders in the system, but facilitates an understanding and regulation of such networks. It should be pointed out that the process of development must assume an evolutionary form with a number of iterations because the emergent properties of artifacts must be tested in real world situations with agents of increasing intelligence and moral competence. We see this paper as a contribution to the macro-level Requirement Engineering through discussion and analysis of general requirements for design of ethical robots.

Keywords Artificial morality · Machine ethics · Machine morality · Roboethics · Autonomous agents · Artifactual responsibility · Functional responsibility

G. Dodig Crnkovic (✉)
Computer Science Laboratory, School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden
e-mail: gordana.dodig-crnkovic@mdh.se

B. Çürüklü
Computational Perception Laboratory, School of Innovation, Design and Engineering, Mälardalen University, Västerås, Sweden
e-mail: baran.curuklu@mdh.se

Introduction

Robots as intelligent agents are one of the most promising future emerging technologies (Gates 2007; Warwick 2009). The more intelligent they become the more useful and effective they are. However, historical experience shows that highly intelligent agents without ethical qualities may easily turn out to be unscrupulous and destructive. The purpose of this article is to show why and how ethics should enter the field of intelligent robots/softbots and contribute to the promotion of the idea that *intelligence*

In the meantime: Research in real-world ethics for specific robots: self-driving cars

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.), Handbook of Research on Machine Ethics and Morality. 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.

Holstein, T., and Dodig-Crnkovic, G. (2018). Avoiding the Intrinsic Unfairness of the Trolley Problem. In Proceedings of the International Workshop on Software Fairness (FairWare '18). Association for Computing Machinery, New York, NY, USA, 32–37.

Holstein, T., Dodig-Crnkovic, G., & Pelliccione, P. (2018). Ethical and Social Aspects of Self-Driving Cars. ArXiv, abs/1802.04103.

Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem

Downloaded from: https://research.chalmers.se, 2023-02-20 09:50 UTC

Citation for the original published paper (version of record):
Holstein, T., Dodig Crnkovic, G., Pelliccione, P. (2021). Steps Towards Real-world Ethics for Self-driving Cars: Beyond the Trolley Problem. Machine Law, Ethics, and Morality in the Age of Artificial Intelligence (Steven John Thompson (Ed.)) : 85-107.
http://dx.doi.org/10.4018/978-1-7998-4894-3.ch006

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology.
It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004.
research.chalmers.se is administrated and maintained by Chalmers Library

Based on experiences from autonomous cars studies

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//

Proceedings of the 14th International Conference on Society and Information Technologies (ICSIT 2023)

Future Intelligent Autonomous Robots, Ethical by Design. Lessons Learned from Autonomous Cars Ethics

Gordana DODIG-CRNKOVIC
Division of Computer Science and Software Engineering, Mälardalen University
Västerås, Sweden
Department of Computer Science and Engineering, Chalmers University of Technology
Gothenburg, Sweden

Tobias HOLSTEIN
Division of Computer Science and Software Engineering, Mälardalen University
Västerås, Sweden

Patrizio PELLICCIONE
Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila
L'Aquila, Italy

Jathoosh THAVARASA
Software Engineering, Efrei Paris
Villejuif, France
Division of Computer Science and Software Engineering, Mälardalen University
Västerås, Sweden

ABSTRACT

The "ethical by design" approach involves examining all stages of a lifecycle of technology to ensure that they are ethically justifiable and socially sustainable. Building on our work on the ethics of autonomous intelligent robocars, and studies of the literature on the ethics of robotics, we propose for robot applications a set of values and ethical principles including safety, security, privacy, transparency, and explainability, accountability, fairness, human control, well-being, autonomy and freedom, and sustainability. This may help stakeholders in the field of intelligent autonomous robotics to connect ethical principles with their applications. Most ethical considerations we identified in our work on autonomous cars are relevant to all AI-powered robots, but robots require additional examination depending on their application domain, such as social robots (care robots, personal companions, robots used in education, health care, elderly care, education, entertainment, chat-bots), industrial robots, exoskeletons. Thus, existing ethical frameworks need to be applied in a context-sensitive way, by assessments in interdisciplinary, multi-competent teams through multi-criteria analysis. Furthermore, we argue for the need for continuous development of ethical principles, guidelines, and regulations, informed by the progress of technologies and involving relevant stakeholders. This implies designing the socio-technical system as an intelligent learning ecology.

Keywords: Ethics, Artificial Intelligence, Autonomous Robots, Intelligent Robots, Roboethics, Autonomous cars, Emerging Technologies, ELSA

I. INTRODUCTION

This article builds on the findings of our book chapter [1] on ethical and social aspects of self-driving cars, which are robots classified as "mobile service robots" [2]. Those are vehicles capable of perceiving their environment and driving without (or with little) human intervention. They combine advanced sensing, controlling, and artificial intelligence with autonomous safety-critical decision-making. Ethical aspects of autonomous cars (also called self-driving cars, autonomous vehicles, driverless cars, automated cars, or robocars) have lately got a lot of attention from the general public, ethicists, researchers, industry, and decision-makers [3].

Our studies on autonomous car ethics led us to the insight that the same approach may be applied to intelligent autonomous robots in general, having in mind that autonomous cars are a special type of intelligent autonomous robots. The research question was: how much of our recommendations for the ethical analysis of autonomous cars can apply to the ethical analysis of other types of robots that present an important emerging technology?

"As a game-changing technology, robotics naturally will create ripple effects through society.", according to Liu, Almay, and Bolay, [3]. The impact of robotics technology on society is significant and far-reaching, potentially leading to major changes in every-day life, business, and culture. Therefore, it is crucial to examine the effects on ethics, law, and policy within ELSA (Ethical, Legal, and Social Aspects) studies, through the fields known as robot ethics, roboethics, and ethics of robotics, which are closely related to AI ethics, machine ethics, technology ethics, and ethical technology.

Peter Asaro posed the fundamental question: "What should we want from a robot ethic?" [4]. His answer is that we need to develop robots that progressively acquire stronger ethical abilities. The primary focus of robot ethics should be avoiding the harm caused by robots. The assignment of responsibility in complex socio-technical systems should be governed by legal theory, [4].

ISBN: 978-1-950492-70-1
ISSN: 2771-6368

https://doi.org/10.54808/ICSIT2023.01.92

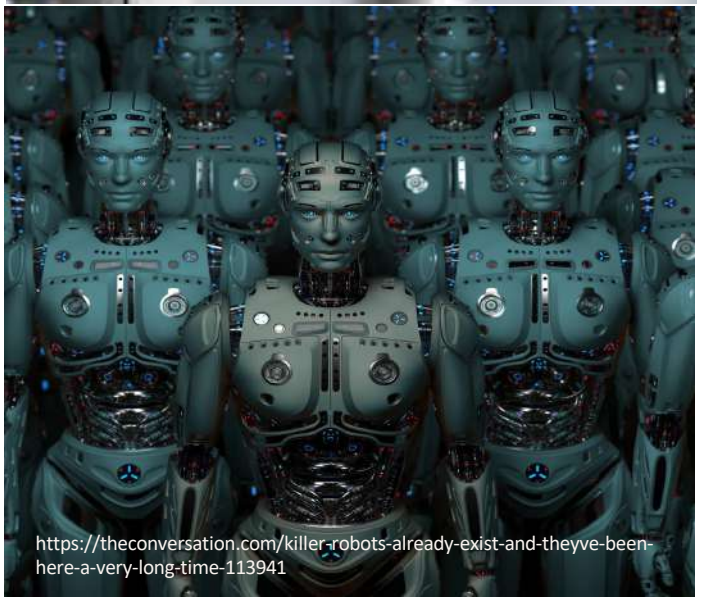
How can we trust intelligent robots?

Trusting the intelligent autonomous robot technology with our future presupposes their anticipated beneficial influence on the societies and individuals, globally. Question of good and bad, right and wrong, and values, in general, are studied within the field of ethics. The emerging fields of Artificial Intelligence (AI) ethics and specifically ethics of intelligent autonomous robotic cars are good examples of ethics research with actionable practical value.

In those ethical fields, a variety of stakeholders, including the legal system with other societal and governmental actors, companies and businesses, collaborate bringing about shared view of ethics.

Drawing from the existing literature on ethics of AI and robotics, and our work on autonomous intelligent robocars, our contribution consists in lessons learned for ethics of autonomous intelligent robots in general, that can help us overview the field with the common set of values and ethical principles, which may help stakeholders in the broader field of intelligent autonomous robotics to connect ethical principles with their applications.

https://robohub.org/from-disembodied-bytes-to-robots-that-think-and-act-like-humans/



https://theconversation.com/killer-robots-already-exist-and-theyve-been-here-a-very-long-time-113941

The Perspective

The aim of this presentation is to offer new views of the Ethics of Robotics as the topic of Design Ethics.
AI ethics and even Robotic ethics are huge.
This lecture wants to open the window with a view, giving you a glimpse of a huge unexplored territory in front of us.

”I invite readers not on a visit to an archaeological museum, but rather on an
adventure in science in making”

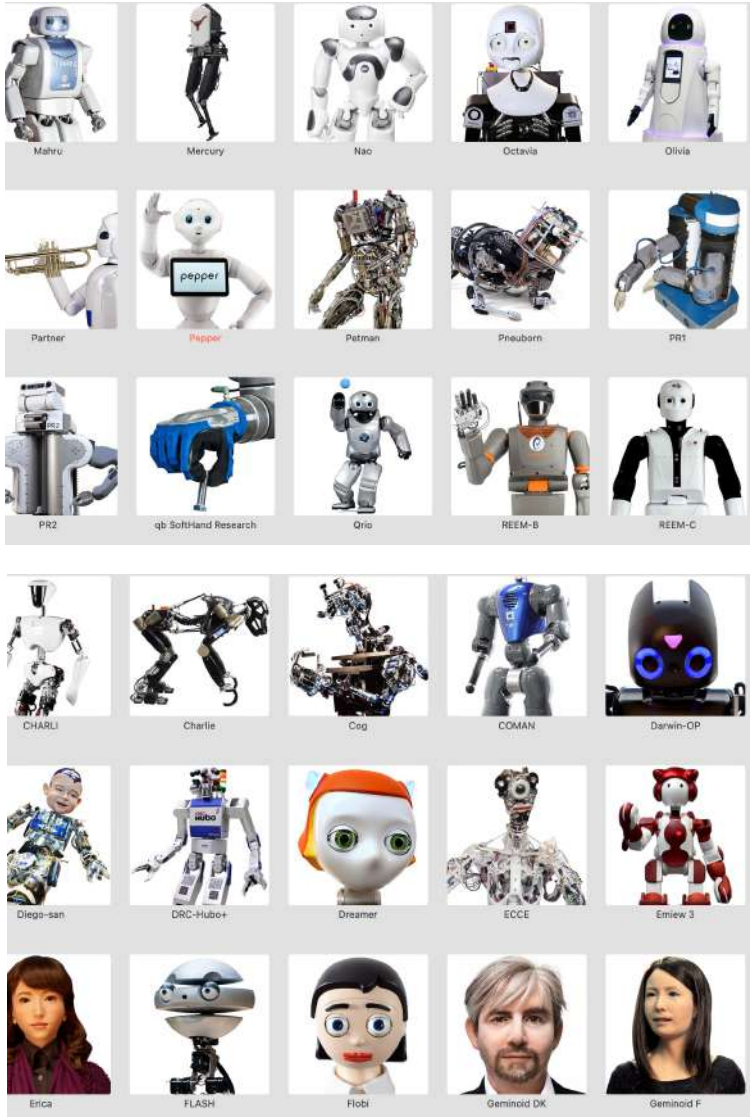
Ilya Prigogine. The End of Certainty: Time, Chaos and New Laws of Nature, 1997

https://www.onventanas.com/historia-vidrio/ventana-japonesa/#iLightbox[postimages]/0

ROBOTS

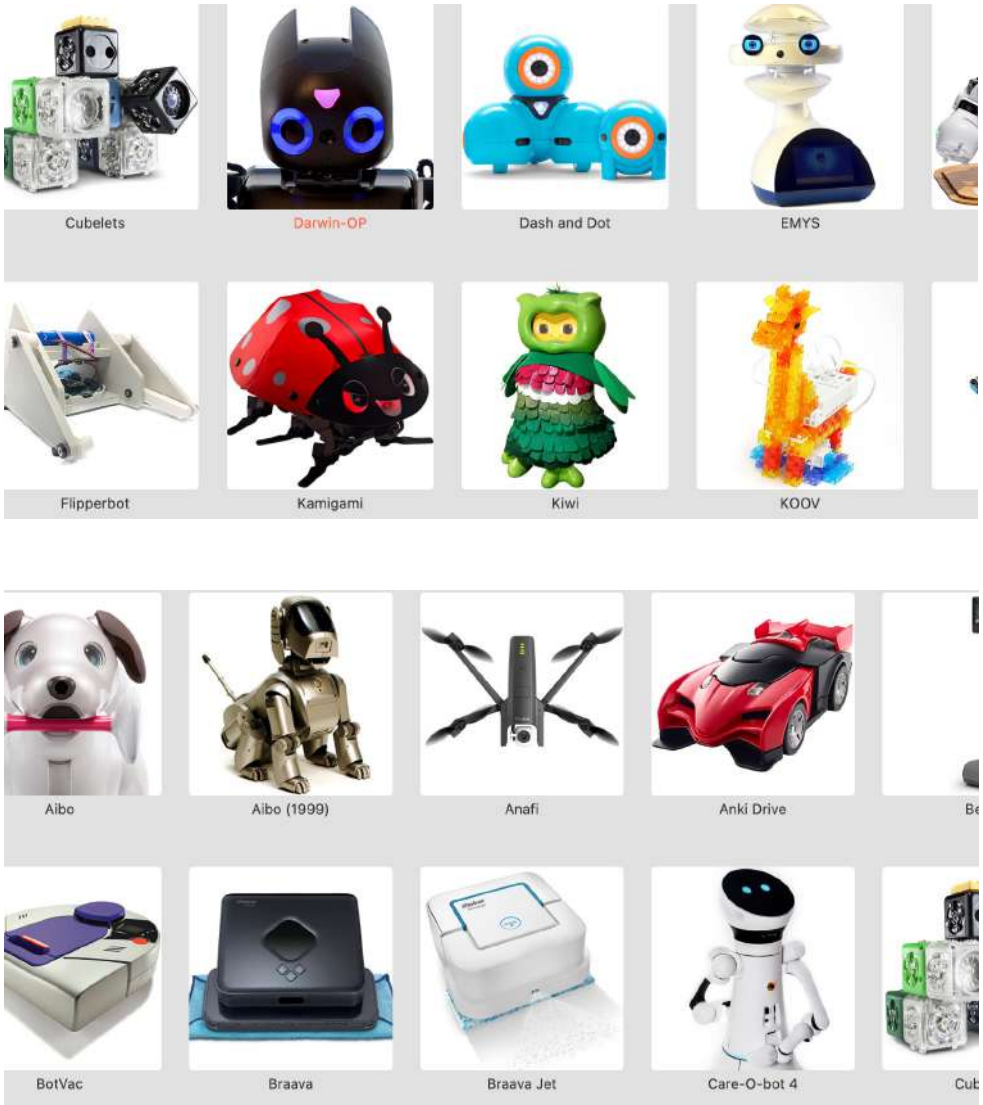
Humanoid robots
Education robots
Consumer robots
Research robots
Medical robots
Nano robots
Disaster response robots
Industrial robots
Aerospace robots
Underwater robots
Aerospace robots
Military and Security robots
Telepresence robots
Drones
Autonomous cars

https://robots.ieee.org/



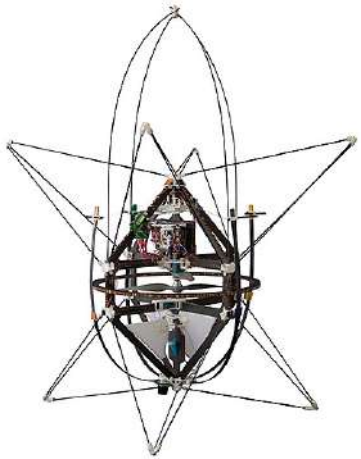
Humanoid Robots

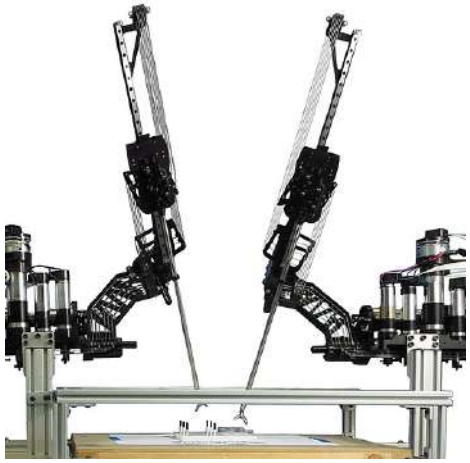
https://robots.ieee.org/



Education & Consumer Robots

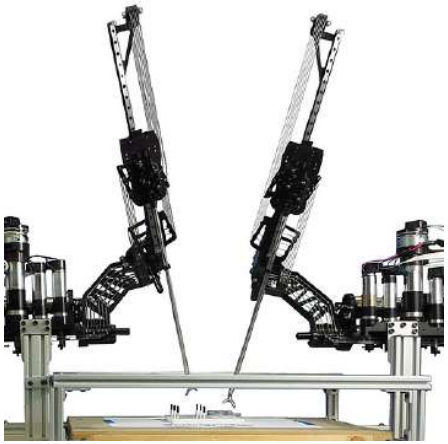
https://robots.ieee.org/





Research Robots

https://robots.ieee.org/



Medical Robots

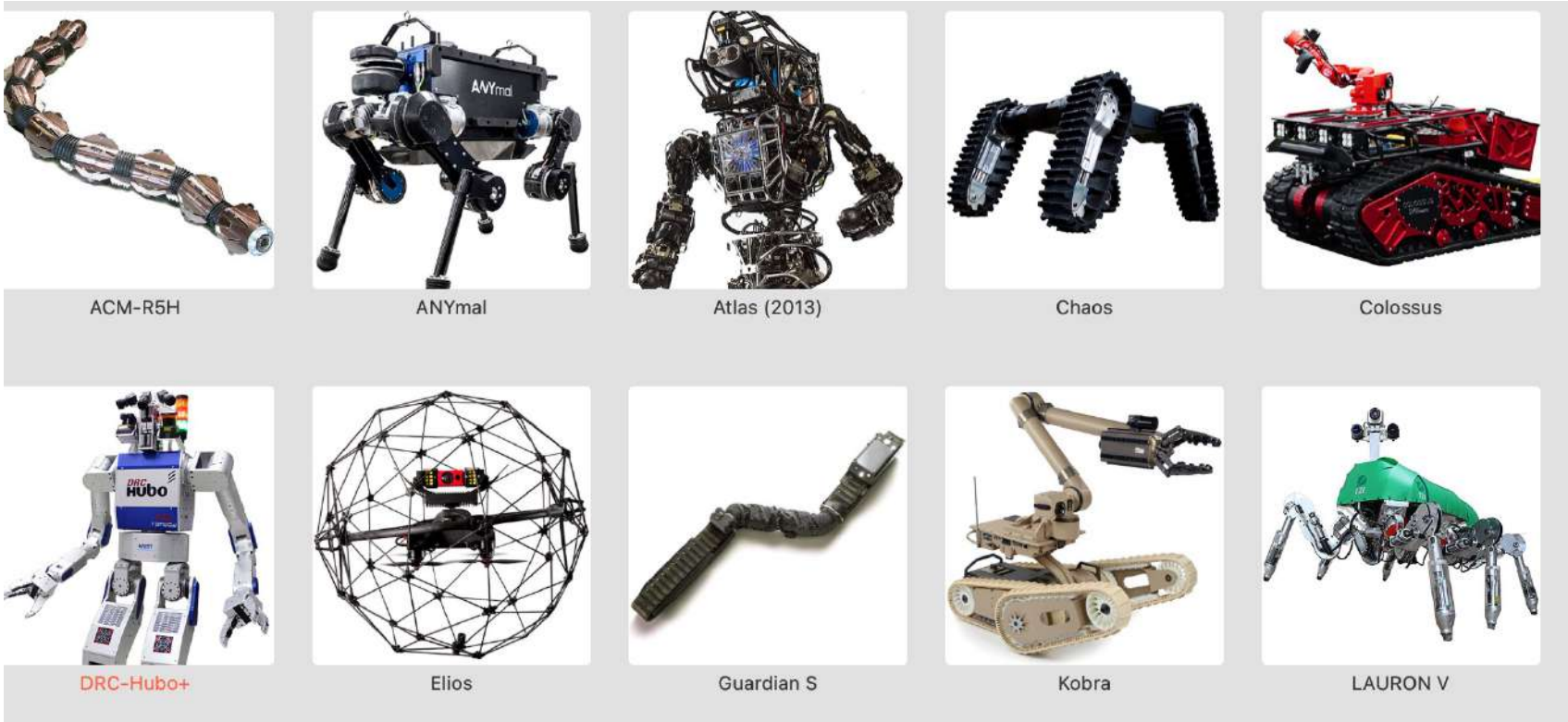
https://robots.ieee.org/

Nano Robots

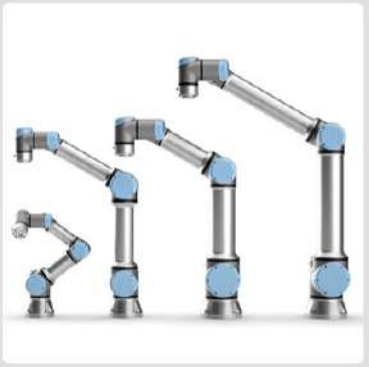
https://www.europeanpharmaceuticalreview.com/news/

Disaster response

https://robots.ieee.org/



Unimate


UR

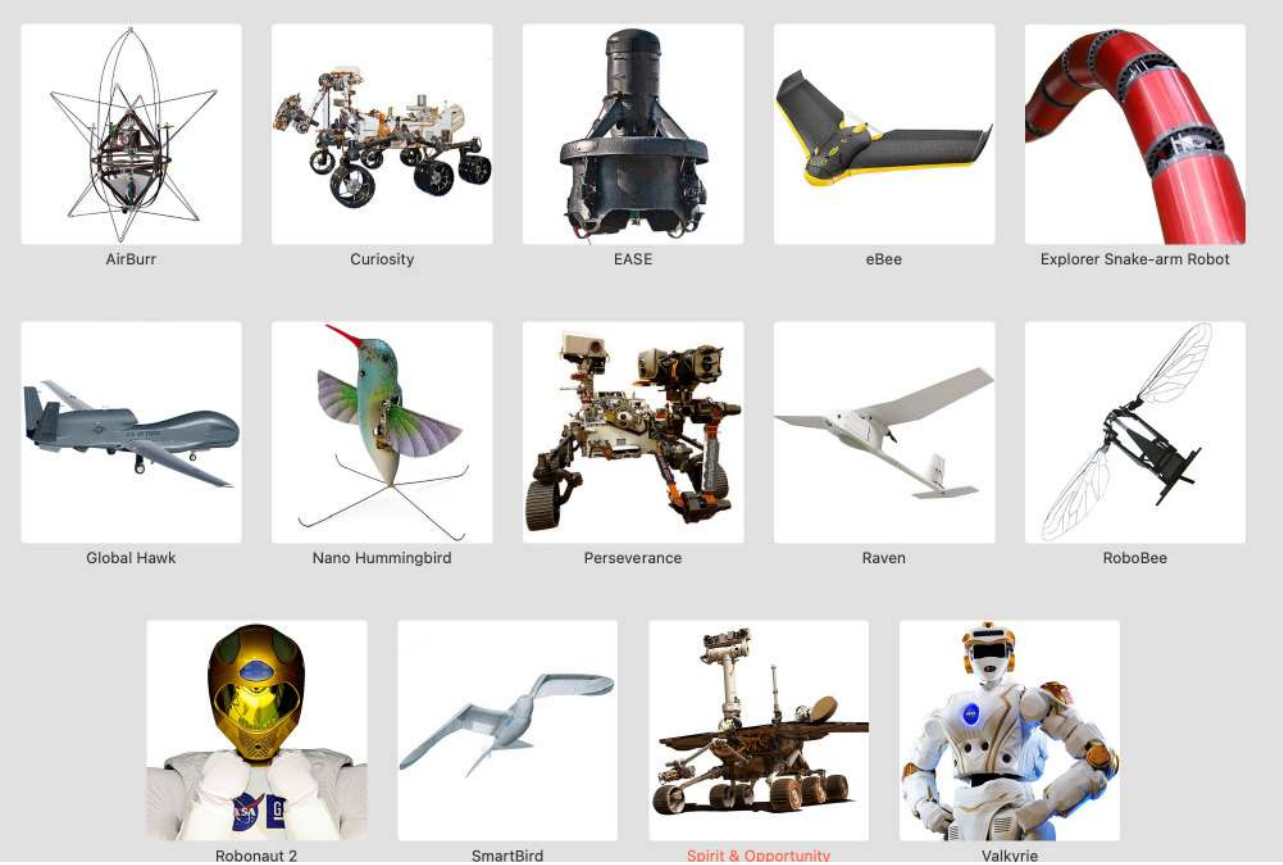
Versatrax

WAM

YuMi

Industrial Robots

https://robots.ieee.org/



Aerospace Robots

https://robots.ieee.org/

ACM-R5H

Aqua2

Aquanaut

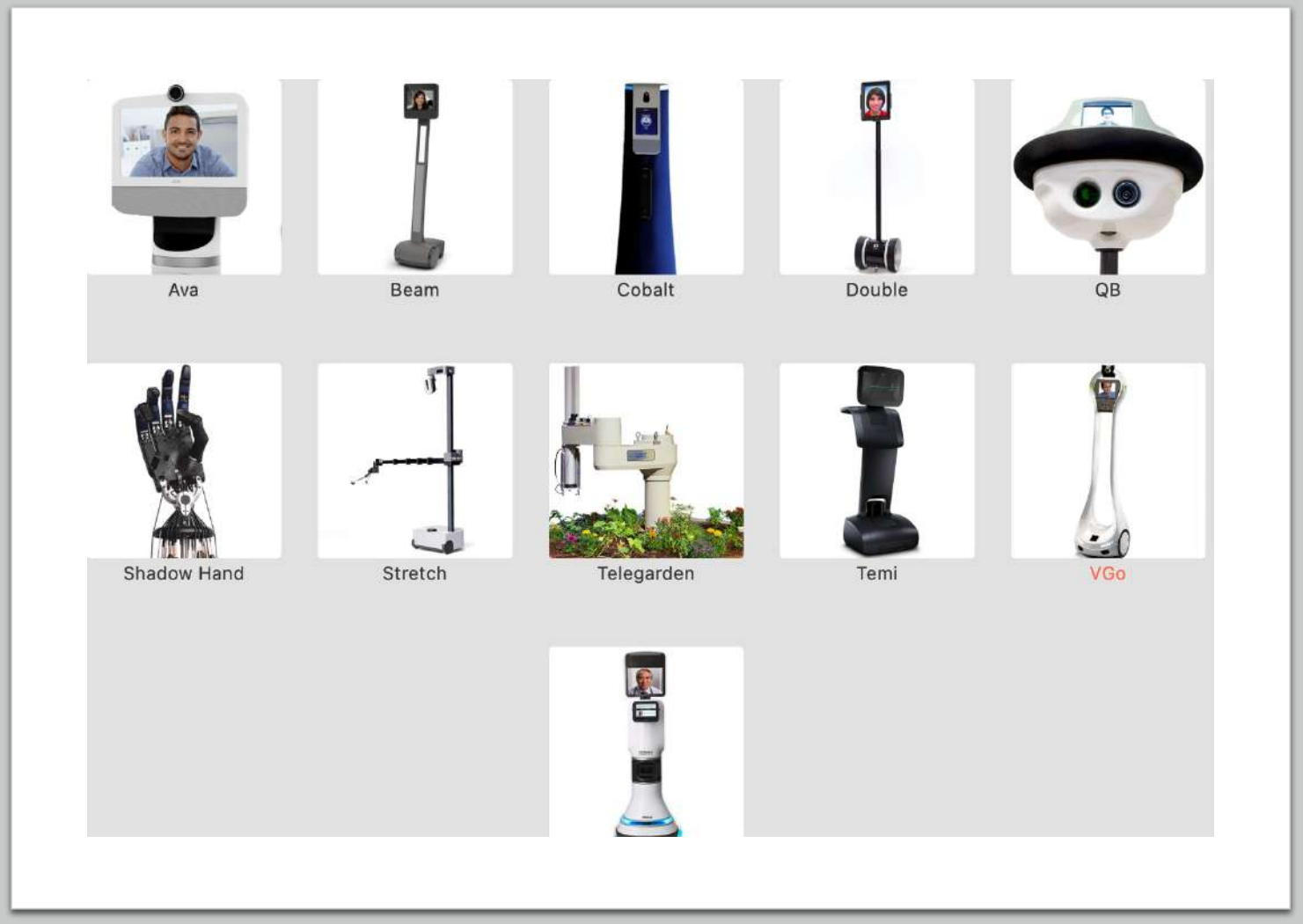
Wave Glider

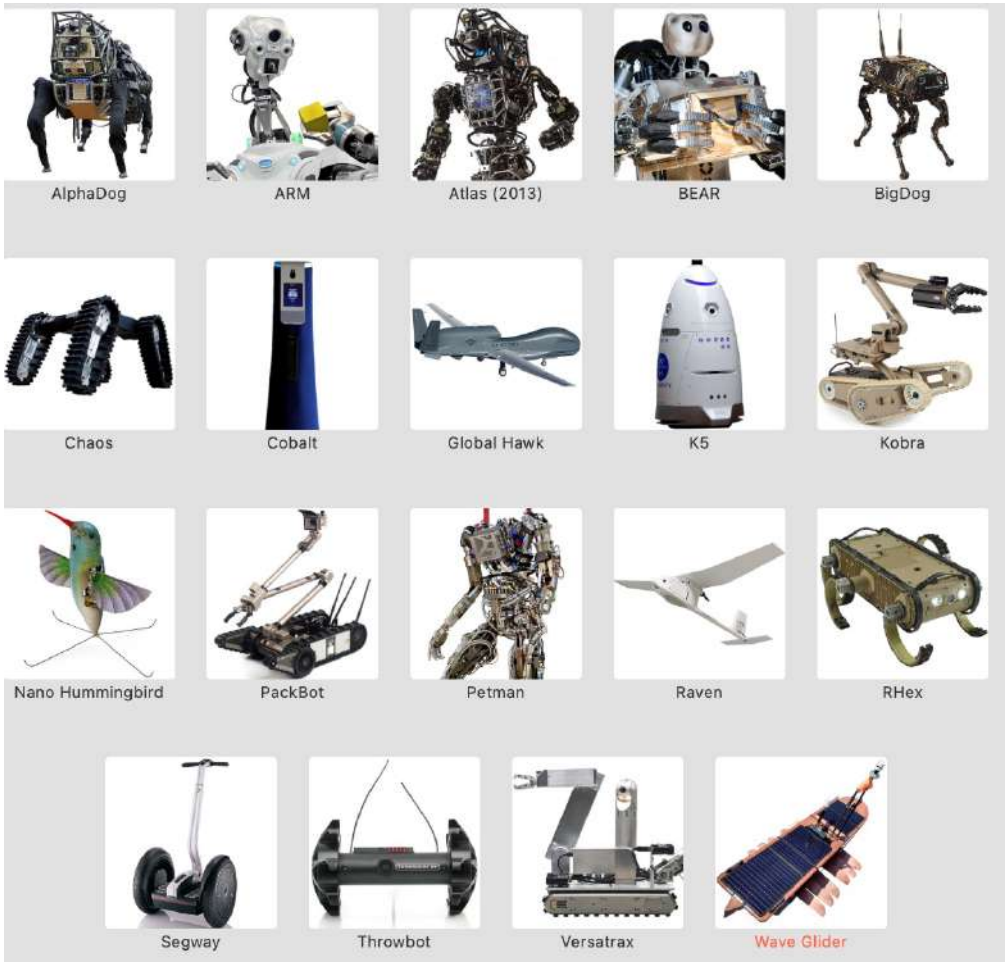
Underwater Robots

https://robots.ieee.org/

Telepresence Robots

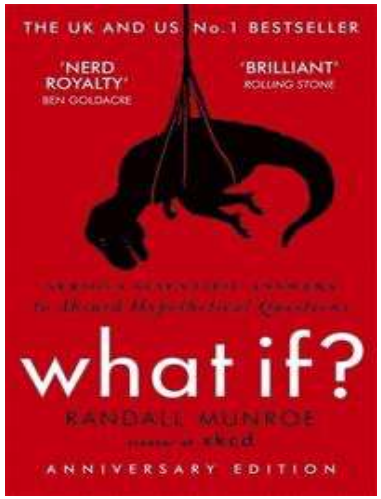
https://robots.ieee.org/





Military and Security Robots

https://robots.ieee.org/



WE ARE DISCUSSING POSSIBLE FUTURES

THROUGH SOCIALLY DISRUPTIVE TECHNOLOGIES

DESIGN FOR POSSIBLE 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. The familiar and engaging nature of the designed output is intended to facilitate discourse with a broad audience: from experts in the field such as scientists, engineers and designers to the consumers and users of technological products and systems.

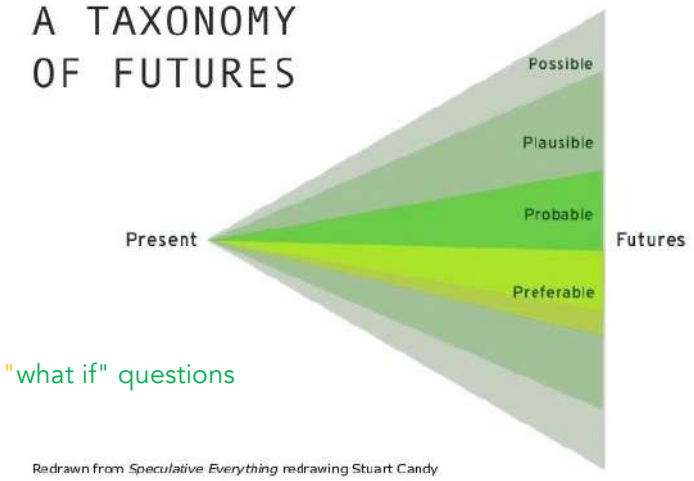
Auger Loizeau

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

SPECULATIVE EVERYTHING

Based on design thinking not SF

(Different methods and goals)!



"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.

SPECULATIVE DESIGN CREATES SPACE TO…

Arrange emerging (not yet available) technological ‘elements’ to hypothesise future, products and artefacts

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

Develop new perspectives on big systems

SPECULATIVE DESIGN FACILITATES…

Asking ‘What is a better future (with respect to 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.

SPECULATIVE DESIGN ASKS…

What would life be like if we had such technologies?

It can act as a cultural and behavioural litmus test, trying out applications before they happen and allowing for adjustments to be made.

Its agenda is to facilitate a more democratic and considered approach to technological development.

ADDRESSING CHALLENGES AND OPPORTUNITIES OF THE FUTURE

We use Speculative Design to describe work that uses design (products, services, scenarios) to address challenges and opportunities of the future.

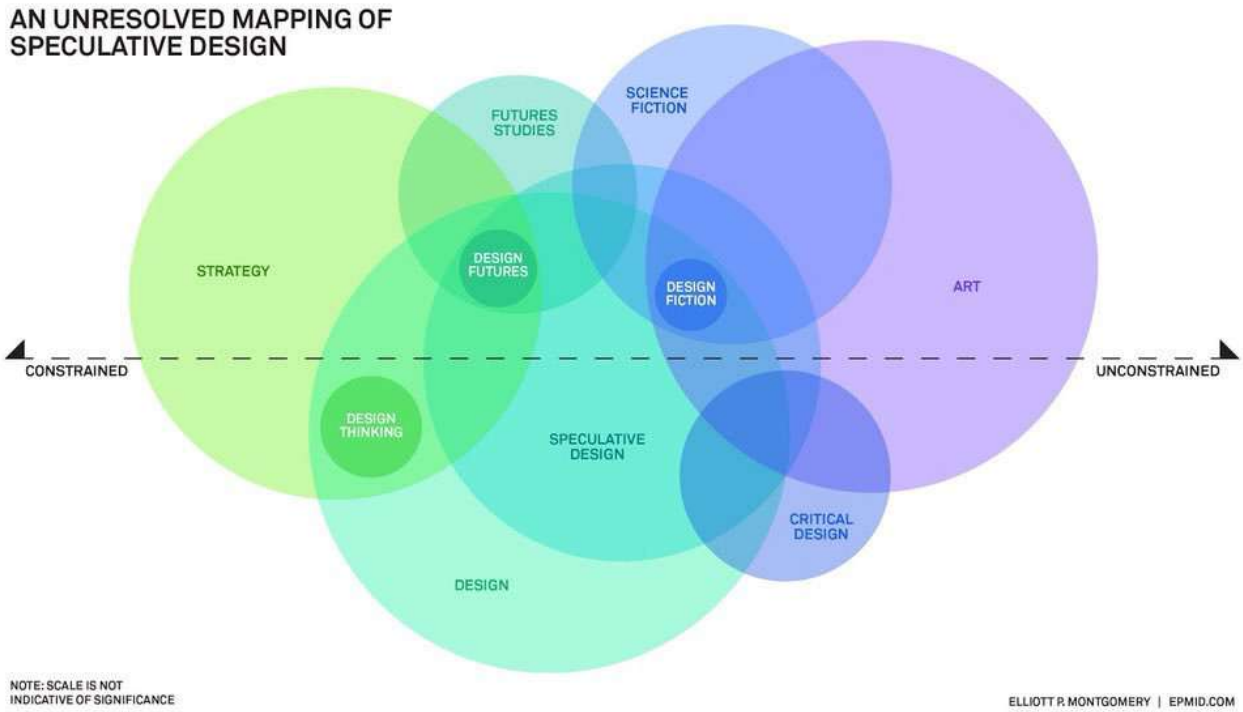
We tend to look 5-10+ years forward and speculate on how things could be and what future we want or don't want based on these scenarios.

CRITICAL DESIGN

"Let's call it critical design, that questions the cultural, social and ethical implications of emerging technologies. A form of design that can help us to define the most desirable futures and avoid the least desirable."

Anthony Dunne & Fiona Raby

SPECULATIVE DESIGN AND ITS CONTEXT



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

Design Thinking and Wicked Problems

Design Unbound. Designing for Emergence in a White Water World

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

https://mitpress.mit.edu/books/design-unbound-designing-emergence-white-water-world-volume-1

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

https://www.youtube.com/watch?v=-U8h4wNBfCQ

https://www.desunbound.com/

A White Water World – Complex & Dynamic

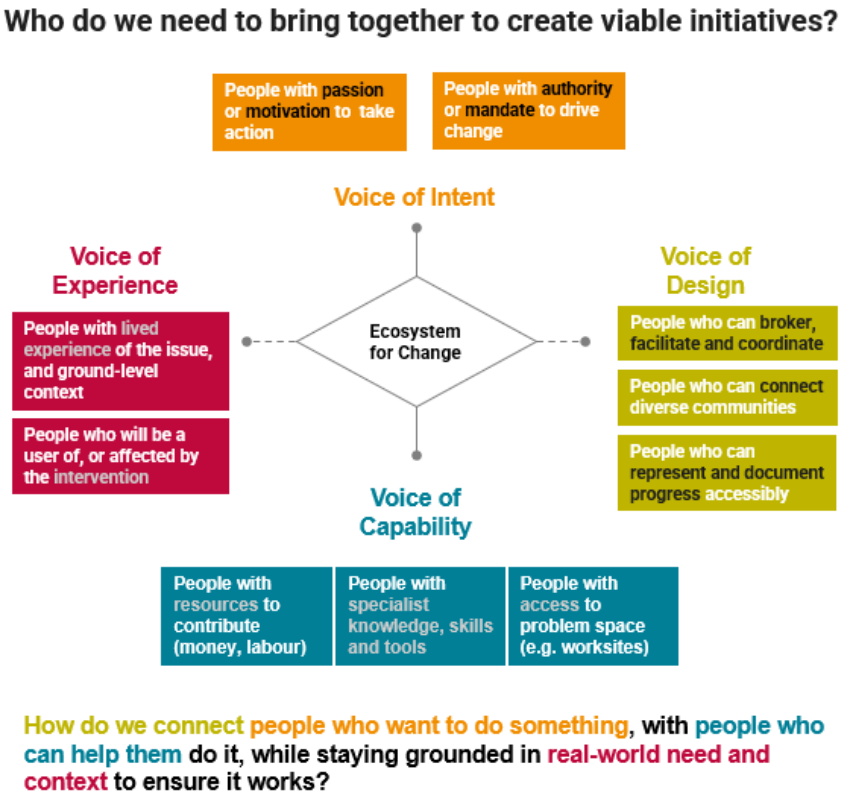
"we are forcing the past as a solution set. But the past as a solution set is not a viable option. We need a new tool set."

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. The tools presented are not the tools of a carpenter but tools that are directly associated with a new kind of practice that is the offspring of complexity science and architecture.

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

COMPLEXITY & SYSTEMIC THINKING

DECISION-MAKING & PLANNING IN AN EMERGENT WORLD



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

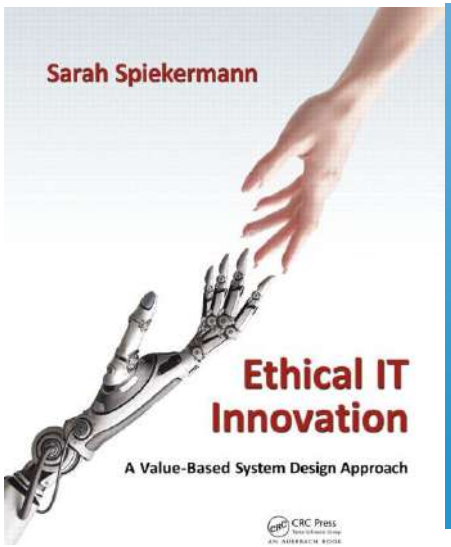
VALUE-BASED HUMAN-CENTRIC DESIGN

Values

Values serve as a guide to action and knowledge.

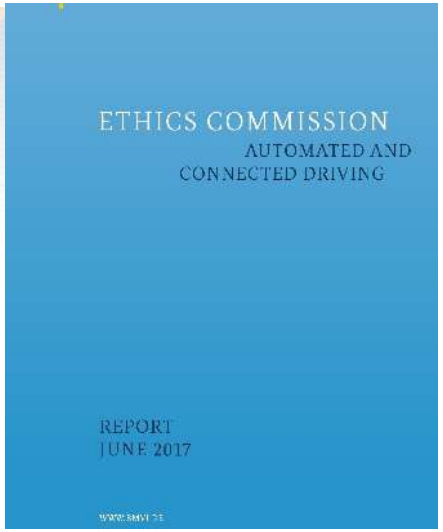
They are relevant to all aspects of scientific and engineering practice, including discovery, analysis, and application.

A VALUE-BASED DESIGN APPROACH



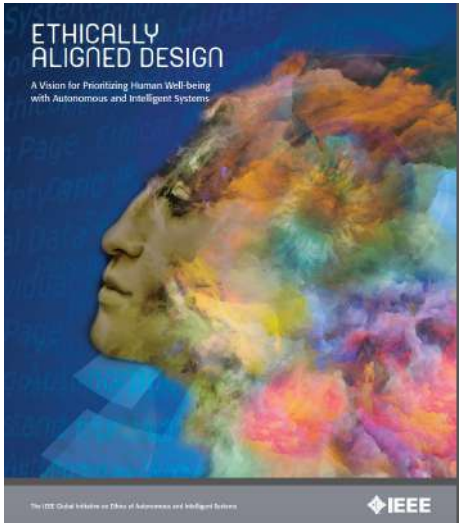
Sarah Spiekermann

Ethical IT Innovation:
A Value-Based System Design Approach



Ethics Commission: Automated and connected driving (Report by Federal Ministry of Transport and Digital Infrastructure of Germany [BMVI])

BMVI = Bundesministerium für Verkehr und digitale Infrastruktur



https://ethicsinaction.ieee.org/

EXPERIENCES FROM Autonomous Cars

-a special case of intelligent autonomous robot

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

Loss of jobs (for people in elderly/health care sector)

Humans in the loop

Impact on Society

Stakeholders Interests

Freedom of choice

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

Implementation of restrictions

To what extent will the user be in control?

What will be the role of AI?

What about GPT-level intelligence?

Believable conversational level and related consequences

Social Trust

Challenges

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

Approaches

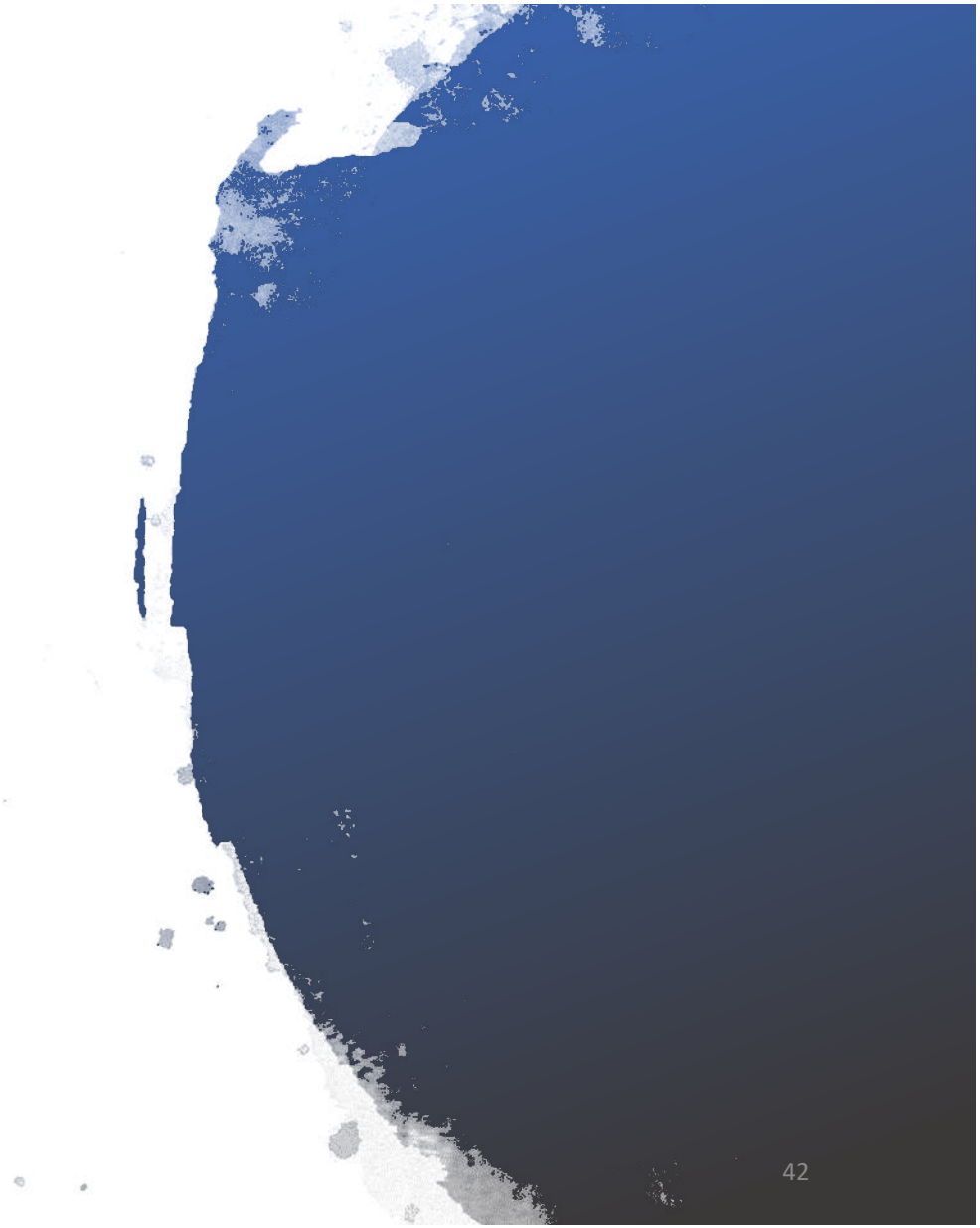
- Further research on how to implement trust across multiple systems
- Provision of trusted connections between components as well as external services

INTELLIGENCE OF EMERGENT TECHNOLOGIES

MAKES A DIFFERENCE

At present: NARROW AI

(STILL FAR FROM HUMAN LEVEL, GENERAL AI)



Addressing Organisational Ethical Issues of AI

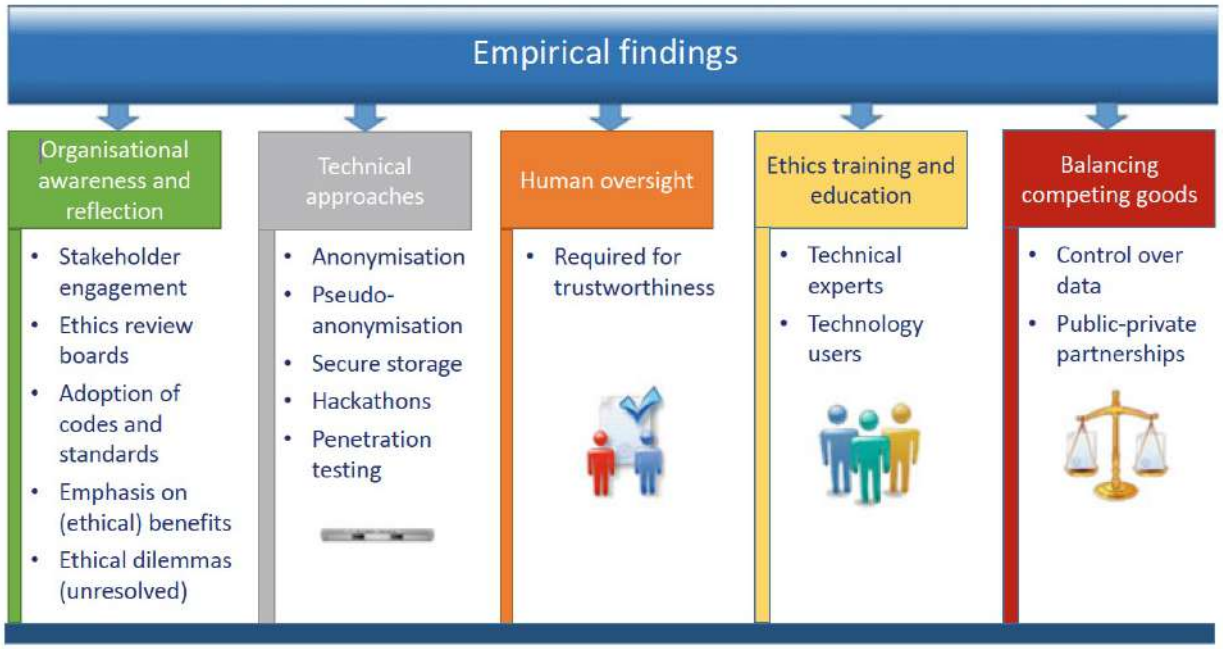
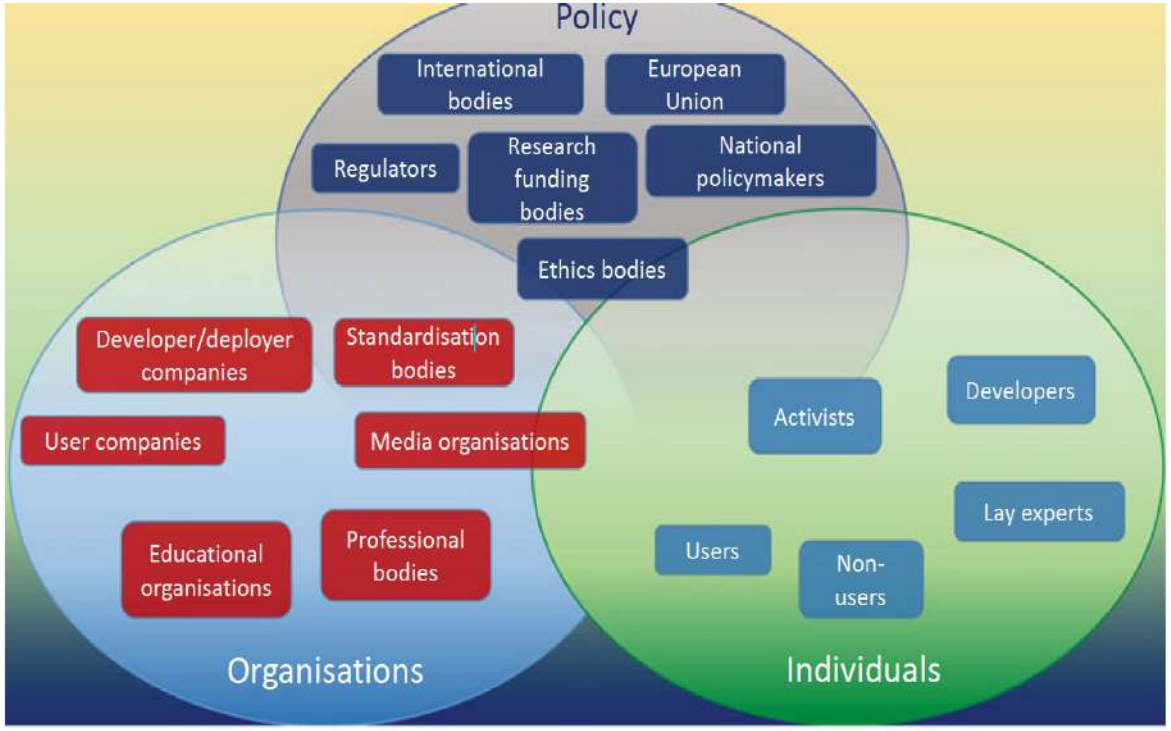


Fig. 5.2 How case study organisations address ethical issues of AI: empirical findings

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

Overview of AI stakeholders, Artificial Intelligence for a Better Future



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 SYSTEMS

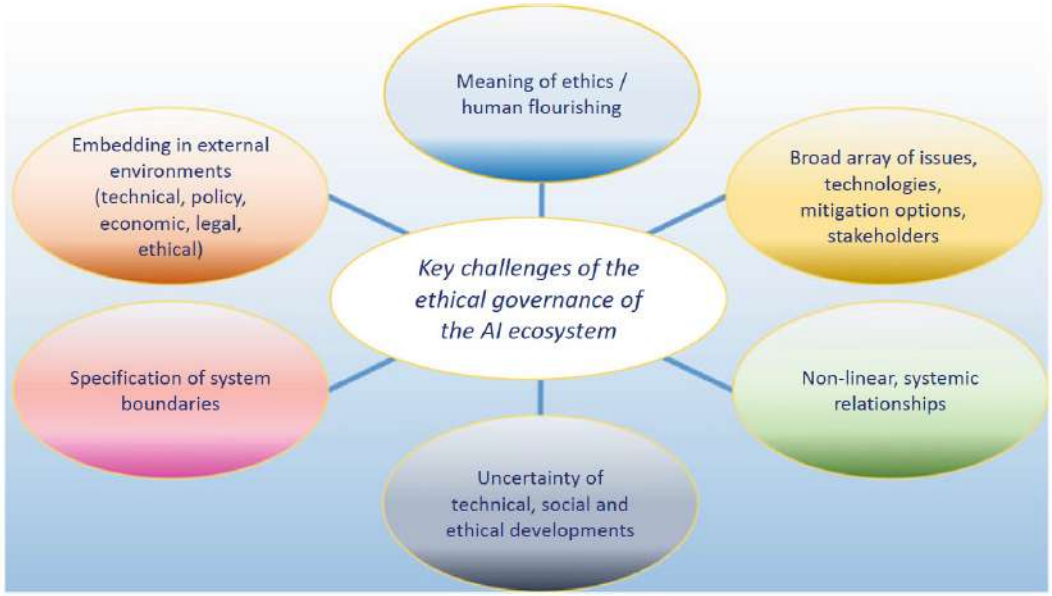


Fig. 7.1 Key challenges of ethical governance of AI ecosystems

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

Ethical Issues of AI

https://www.youtube.com/watch?v=uZMs9IePwMQ What if ChatGPT had a robotic body?

In an analogy with artificial intelligence, which is the ability of a machine to perform activities that would require intelligence in humans, artificial morality is considered to be the ability of a machine to perform activities that would require morality in humans.

The capacity for artificial (artifactual) morality, such as artifactual agency, artifactual responsibility, artificial intentions, artificial (synthetic) emotions, etc., comes in varying degrees and depends on the type of agent.

Table 4.1 Three categories of ethical issues of artificial intelligence

1. Issues arising from machine learning	
Privacy and data protection	Lack of privacy
	Misuse of personal data
	Security problems
Reliability	Lack of quality data
	Lack of accuracy of data
	Problems of integrity
Transparency	Lack of accountability and liability
	Lack of transparency
	Bias and discrimination
	Lack of accuracy of predictive recommendations
	Lack of accuracy of non-individual recommendations
Safety	Harm to physical integrity
2. Living in a digital world	
Economic issues	Disappearance of jobs
	Concentration of economic power
	Cost to innovation
Justice and fairness	Contested ownership of data
	Negative impact on justice system
	Lack of access to public services
	Violation of fundamental human rights of end users
	Violation of fundamental human rights in supply chain
	Negative impact on vulnerable groups
	Unfairness
Freedom	Lack of access to and freedom of information
	Loss of human decision-making
	Loss of freedom and individual autonomy
Broader societal issues	Unequal power relations
	Power asymmetries
	Negative impact on democracy
	Problems of control and use of data and systems
	Lack of informed consent
	Lack of trust
	Potential for military use
	Negative impact on health
	Reduction of human contact
	Negative impact on environment
Uncertainty issues	Unintended, unforeseeable adverse impacts
	Prioritisation of the "wrong" problems
	Potential for criminal and malicious use
3. Metaphysical issues	
	Machine consciousness
	"Awakening" of AI
	Autonomous moral agents
	Super-intelligence
	Singularity
	Changes to human nature

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

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:

1. Conceptualization/Design/Prototyping/ Construction/Development/Testing/Production
2. Deployment/Application/
3. Maintenance/Support
4. Oversight/Regulation

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

Challenges

Legislation	Global framework	Guidelines	Implementation of Ethics
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

Building Ethical Technology in an Ethical Way

Before the question of how to build ethical technology in an ethical way comes the question if it is possible. For example, the open question of intelligent autonomous weapons currently prompted ethicists and roboticists to propose a complete ban on intelligent autonomous weapons. Thus, the first question to ask is whether certain technology is acceptable at all

The ethics of intelligent autonomous robots must permeate design, application, production, and/or maintenance and oversight within the corresponding techno-social system, and must be based on learning from experience

Both studies from the literature and our own research emphasize the need for a system-level approach involving the entire software-hardware system as well as human, organizational, and social factors.

With the constantly evolving, emergent nature of intelligent technologies, a crucial aspect is their development includes anticipation and consideration of uncertainties. Speculative design with anticipatory ethics are necessary for emerging technologies

At present, there is a gap between general principles and their specific, context-dependent implementations when making multi-criteria decisions and identifying key ethical considerations. This issue can only be resolved through the collaboration of multidisciplinary teams with the appropriate expertise, working within the specific context in question.

Ethical principles, guidelines, and assessments, as well as regulatory documents, must be continually updated and developed in line with technological advancements and must involve input from all relevant stakeholders. Incorporating ethical considerations into the development and use of intelligent autonomous robots is essential for building trustworthy future technology systems.

DIGITAL HUMANISM

https://owncloud.tuwien.ac.at/index.php/s/vmZSxsuruhk77Iy

Digital technologies are changing human behaviour, significantly changing our society, and our environment. Digital humanism observes and describes these changes and aims at shaping and influencing the development of digital technologies and policies towards the values of human rights, democracy, participation, inclusion, and diversity. It is a broad concept combining technical and social innovation and ranging from research to politics. Aims to:

- Put the human in control
- Strengthen the social contract, lend a personal voice
- Mitigate risks of categorization and classification
- Provide personal data privacy
- Detect malicious behaviour
- Develop and deploy trustworthy systems
- Develop AI to work in partnership with humans

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/

DIGITAL HUMANISM – ROADMAP

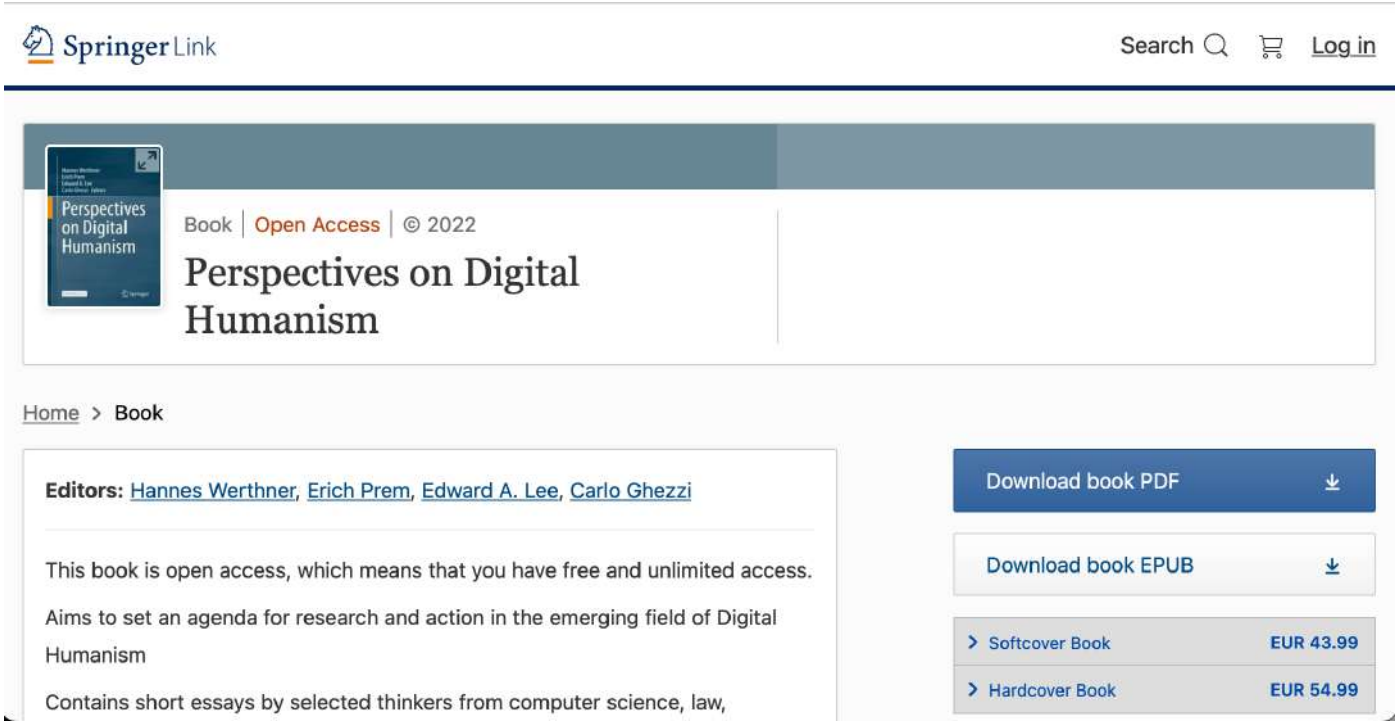
https://owncloud.tuwien.ac.at/index.php/s/vmZSxsuruhk77Iy

Support societal values such as:

- Morality
- Human dignity; the human as worthy of respect; the human condition in its contrast to machines
- Value-based online content curation
- Sustainability and environmental protection
- Democratisation and transparency in recommendation and personalization
- Varied stakeholder views in the design of digital technologies
- An inclusive society

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



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

Digital Humanism Lecture Series

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

https://www.youtube.com/@DigitalHumanism Youtube channel

Digital Humanism Manifesto

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

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

UNESCO Chair on Digital Humanism

TU WIEN Informatics

Peter Knees and Julia Neidhardt
Chairholder and Co-Chairholder

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

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

EUROPEAN VOICES IN THE ETHICS DEBATE
Peter-Paul Verbeek

https://ppverbeek.org/video-and-audio/

https://youtu.be/rUQKS1WbCZw AI at work Keynote by Peter-Paul Verbeek | Professor, University of Twente 28:00 "EU lead"

https://www.utwente.nl/en/news/2019/7/185302/peter-paul-verbeek-new-chairman-unesco-commission

Peter-Paul Verbeek, Distinguished Professor at the University of Twente, was appointed chairman of UNESCO's World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), 2019 → 2023. https://www.unesco.org/en/ethics-science-technology/comest

https://www.youtube.com/watch?v=S8a1DascnZg Moralizing Technology and the ethics of things

https://ppverbeek.org/

EUROPEAN VOICES IN THE ETHICS DEBATE
Philip Brey

https://scholar.google.com/citations?user=MZY_5kAAAAAJ&hl=en

https://www.youtube.com/watch?v=enGWtmIuUUM Ethics by Design in AI & Big Data systems | Philip Brey on SIENNA & SHERPA work

https://www.youtube.com/watch?v=VQmFUySR9Zo Ethics In A Digital Society | Introduction And Overview | The Other Society

https://www.itspmagazine.com/the-other-society-hosts-and-panelists/kevin-macnish

https://people.utwente.nl/p.a.e.brey

EUROPEAN VOICES IN THE ETHICS DEBATE
Vincent C. Müller

A. v. Humboldt Professor, Philosophy and Ethics of AI

Director, Centre for Philosophy and AI Research (PAIR)

Friedrich-Alexander Universität Erlangen-Nürnberg (FAU)

Turing Fellow, The Alan Turing Institute, UK

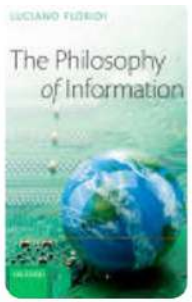
Visiting Professor, Technical University Eindhoven (TU/e) NL

https://plato.stanford.edu/entries/ethics-ai/ Ethics of Artificial Intelligence and Robotics

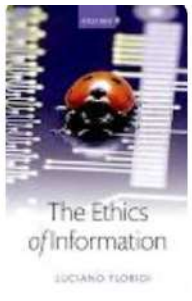
http://www.sophia.de/activities.htm

EUROPEAN VOICES IN THE ETHICS DEBATE
Luciano Floridi

https://en.wikipedia.org/wiki/Luciano_Floridi

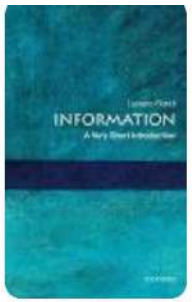


The Philosophy ...
2011

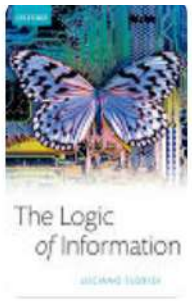


The Ethics of Information
2013

The Fourth Revolution:...
2014



Information: A Very Short...
2010



The Logic of Information:...
2019

Philosophy and...
1996

https://www.youtube.com/watch?v=YLNGvvgq3eg We live in the infosphere

https://www.youtube.com/watch?v=R2vxeEO87d4 Interview Prof. Luciano Floridi

https://www.oii.ox.ac.uk/people/profiles/luciano-floridi/

EUROPEAN VOICES IN THE ETHICS DEBATE
Mark Coeckelbergh

https://www.youtube.com/watch?v=JCgAvA5UIV4

Mark Coeckelbergh on Responsibility for AI

The Horizon Europe-funded https://bit.ly/3HtXE97 ERA Chair at the Institute of Philosophy of the Czech Academy of Sciences in Prague where I will help to set up a new international research Center of Environmental and Technology Ethics – Prague (CETE-P)(https://bit.ly/3Bw4APe).

Guest professor at WASP-HS (https://bit.ly/3Wh1whB) and University of Uppsala, where I will have the pleasure to work on our exciting research project "AI Design Futures" with Amanda Lagerkvist, Magnus Strand, and Virginia Dignum.

https://link.springer.com/article/10.1007/s11948-017-9953-8

Technology Games: Using Wittgenstein for Understanding and Evaluating Technology

https://philtech.univie.ac.at/team/mark-coeckelbergh/

EUROPEAN VOICES IN THE ETHICS DEBATE
Virginia Dignum

https://www.youtube.com/watch?v=_U8A5j7qdhg Responsible Artificial Intelligence: What is it and why care?

https://www.youtube.com/watch?v=Qmkqvy400gA

Virginia Dignum - Trustworthy AI Systems and Labour Markets (Regional Reshaping Work Stockholm 2020)

https://arxiv.org/pdf/2302.06655.pdf On the importance of AI research beyond disciplines

https://wasp-hs.org/blogposts/why-we-shouldnt-pause-research-on-ai-but-instead-prioritize-multidisciplinary-research-and-ai-governance/ Why We Shouldn't Pause Research on AI, but Instead Prioritize Multidisciplinary Research and AI Governance

https://www.umu.se/personal/virginia-dignum/

EUROPEAN VOICES IN THE ETHICS DEBATE
Sarah Spiekermann-Hoff

https://www.crcpress.com/Ethical-IT-Innovation-A-Value-Based-System-Design-Approach/Spiekermann/p/book/9781482226355#googlePreviewContainer Ethical IT Innovation. A Value-Based System Design Approach

https://www.researchgate.net/publication/318993631_IEEE_P7000_The_First_Global_Standard_Process_for_Addressing_Ethical_Concerns_in_System_Design

Wolfie Christl and Sarah Spiekermann (2016) Networks of Control. A Report on Corporate Surveillance, Digital Tracking, Big Data & Privacy. Facultas, Vienna

http://crackedlabs.org/dl/NetworksOfControl_PressInfoEN.pdf

https://www.wu.ac.at/ec/team-vo/sarah-spiekermann-hoff/

EUROPEAN VOICES IN THE ETHICS DEBATE
Bernd Carsten Stahl

https://super-morri.eu/speaker/bernd-carsten-stahl/

Professor of Critical Research in Technology and Director of the Centre for Computing and Social Responsibility at De Montfort University, Leicester, UK. My interests cover philosophical issues arising from the intersections of business, technology, and information. This includes ethical questions of current and emerging of ICTs, critical approaches to information systems and issues related to responsible research and innovation.

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

Artificial Intelligence for a Better Future – Open access

https://www.dmu.ac.uk/about-dmu/academic-staff/technology/bernd-carsten-stahl/bernd-carsten-stahl.aspx

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, Arhus University, Danmark, Ed. Ess and Hagengruber, July 201
- 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

References

- 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