IEEE International Conference on Robot and Human Interactive Communication - RO-MAN 2022
The 2nd Workshop on Design Centred HRI and Governance

https://krinuts7.wixsite.com/ieee-roman-2022/agenda

Robot-Human Interactions in the Case of Robotic Autonomous Cars

A Speculative Design Perspective

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.html

THE AIM

Developing intelligent autonomous cars that we can trust and like

Workshop on Design Centered HRI and Governance

Developing intelligent autonomous robots that we can trust and enjoy presupposes they meet our expectation on values with anticipated beneficial influence on the societies and individuals, globally with respect to **ELSI (Ethical, Legal and Social Implications)**.

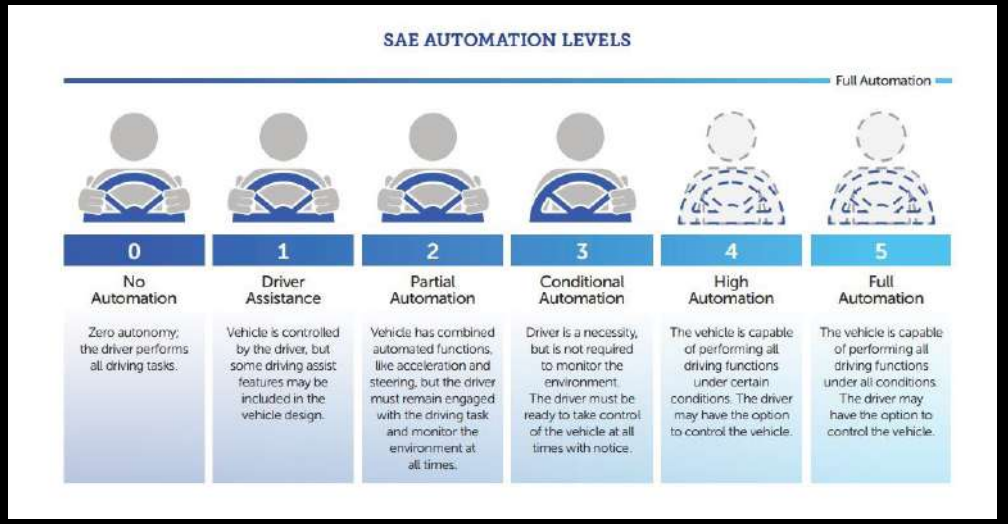
Values with questions of good and bad, right and wrong, and values, in general, are studied within the field of ethics.

In the emerging technology variety of stakeholders, including the legal system with other societal and governmental actors, companies and businesses, collaborate bringing about shared view of ELSI of the future. When it comes to design, future is addressed by speculative design (Anthony Dunne & Fiona Raby, 2013)

With respect to the problem being addressed we talk about Design Unbound. Designing for Emergence in a White Water World (Ann Pendleton-Jullian and John Seely Brown, 2018)

DEVELOPMENT OF SELF-DRIVING CARS

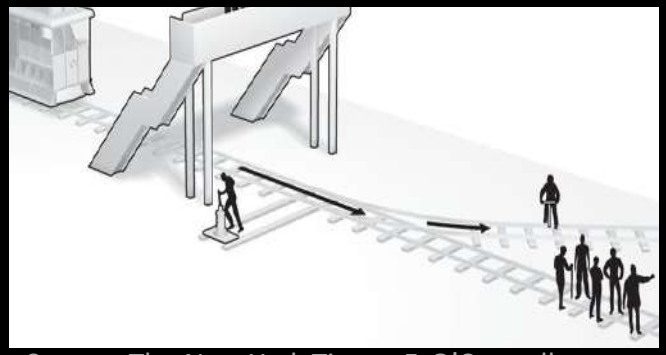
LEVELS OF AUTOMATION IN SELF-DRIVING CARS



The highest level of autonomous driving (Level 5 of 5), where a car can drive without human supervision:

No Steering-wheel or other primary driving controls => the former driver becomes solely passenger.

THE "TROLLEY PROBLEM"- THE PSEUDO-PROBLEM DOMINATING FOR LONG TIME THE DEBATE ABOUT AUTONOMOUS CARS



Source: The New York Times; F. O'Connell

Current discussions about self-driving cars repeatedly take form of decision-making problem borrowed from philosophy

THE TROLLEY PROBLEM:
whom WILL THE SELF-DRIVING CAR kill when it has to decide?

Ethical thought experiment defined by philosopher Philippa Foot in "The Problem of Abortion and the Doctrine of the Double Effect," pp. 5-15, *Oxford Review,* 5, (1967). Focus on the difference between responsibility of acting vs. non-acting.

Many different variants, such as the use of personas to include an emotional perspective. But there is always a single decision to make:

Whom to kill? **The Trolley Problem is Unsolvable by Construction**

THE REAL-WORLD ENGINEERING AND INTERACTION DESIGN PROBLEM IS
NOT WHOM TO KILL
BUT HOW NOT TO KILL!

OR IN GENERAL: HOW NOT TO HARM AND HOW TO DO GOOD, CONTRIBUTING TO HUMAN WELL-BEING

How to do good?
Values

"Value" is defined broadly as property that a person or a group considers important and desirable.

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.

VALUE-SENSITIVE/ VALUE-BASED DESIGN
helps us to establish criteria of good technology

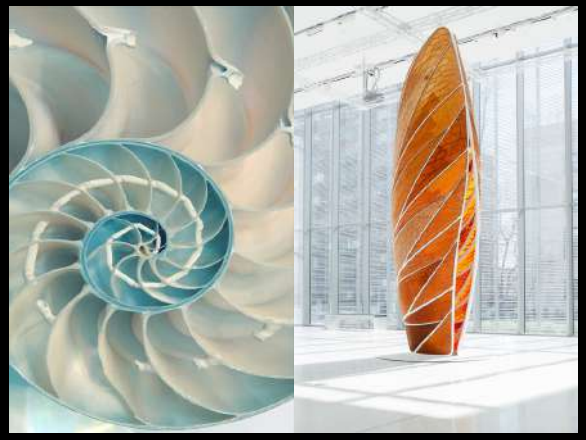
- Value-sensitive design (VSD) holds that artefacts are value-laden and design is value-sensitive. Thus there is need to identify early implicit values embedded in new technologies typically by focusing on the use cases.
- "Value" is defined broadly as property that a person or a group considers important and desirable. Designers unintentionally or intentionally inscribe their values in the design objects, thus shaping them accordingly.

- The design is typically carried out iteratively by combining the following approaches supporting the values:
  - conceptual (conceptions of values for users and stakeholders)
  - empirical (how values are realized in practice)
  - technical (design of technology),
  - research all of which is followed by
  - assessment

Luciano Floridi, Josh Cowls, Thomas C. King, Mariarosaria Taddeo (2020) How to Design AI for Social Good: Seven Essential Factors. Science and Engineering Ethics. https://doi.org/10.1007/s11948-020-00213-5

UN GOALS 2030

Doing good in the long run.
What does it mean?



EMERGING
TECHNOLOGIES
AND
SPECULATIVE
INTERACTION
DESIGN

I would like to draw your attention to the time perspective and Speculative design view of technology

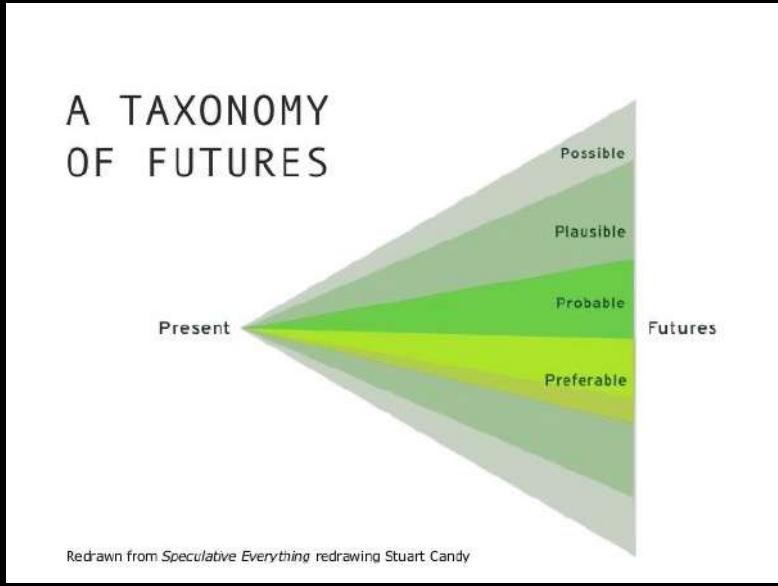


Table of Contents:

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 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 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.

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.

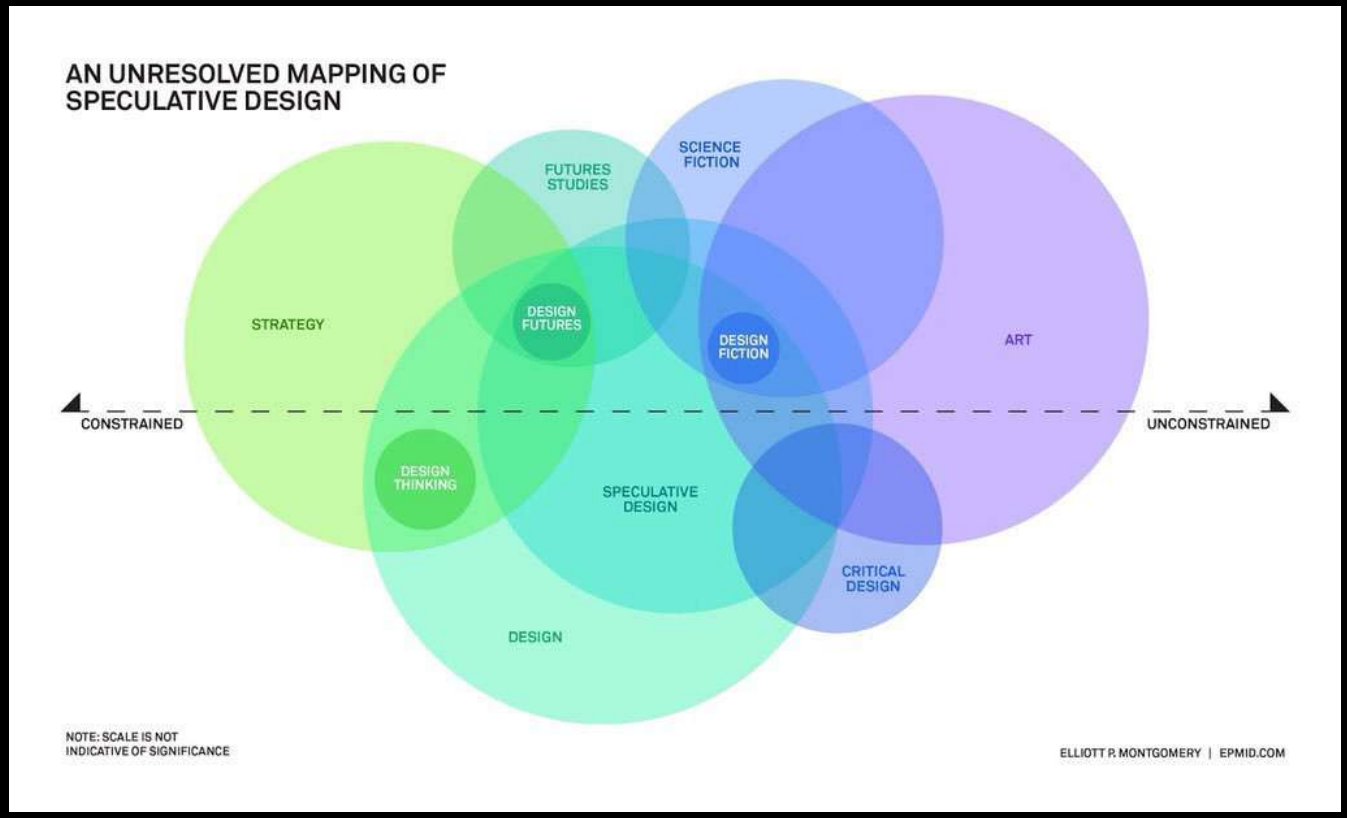
CRITICAL DESIGN QUESTIONS CULTURAL, SOCIAL AND ETHICAL

"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, 2013

http://dunneandraby.co.uk/content/bydandr/42/0

SPECULATIVE DESIGN AND ITS CONTEXT



APPROACHING HUMAN-ROBOT INTERACTIONS FOR EMERGING TECHNOLOGIES

Ponder over emerging technologies and their potential as “design material”

Think of present and future challenges you want to address

Remember UN sustainability goals

Think “what if”

Choose among variety of possible technologies, designs, and goals

SPECULATIVE LARGE-SCALE DESIGN

Dunne and Raby emphasize the potential of speculative design for large-scale social and political issues, such as democracy or sustainability or the alternatives to the existing economic models.

…

It should be kept in mind, that the purpose of speculative design should not be utopian or dystopian science fiction visions of the future, but dialogue on what the future can be.

DESIGNING FOR EMERGENCE IN A WHITE WATER WORLD

Design Unbound. Designing for Emergence in a White Water World

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

“navigating today's hyper-connected, rapidly changing, and radically contingent white water world.”

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

WICKED PROBLEMS IN DESIGN THINKING

What are Wicked Problems?

Wicked problems are problems with many interdependent factors making them seem impossible to solve. Because the factors are often incomplete, in flux, and difficult to define, solving wicked problems requires a deep understanding of the stakeholders involved, and an innovative approach provided by design thinking. Complex issues such as healthcare and education are examples of wicked problems.

The term “wicked problem” was first coined by Horst Rittel, design theorist and professor of design methodology at the Ulm School of Design, Germany. In the paper “Dilemmas in a General Theory of Planning,” he describes ten characteristics of wicked problems:

There is no definitive formula for a wicked problem.
Wicked problems have no stopping rule, and there’s no way to know your solution is final.
Solutions to wicked problems are not true-or-false; they can only be good-or-bad.
There is no immediate test of a solution to a wicked problem.
Every solution to a wicked problem is a "one-shot operation”
Wicked problems do not have a set number of potential solutions.
Every wicked problem is essentially unique.
Every wicked problem can be considered a symptom of another problem.
There is always more than one explanation for a wicked problem .Explanations depend on the individual perspective.
Planners/designers have no right to be wrong and must be fully responsible for their actions.

https://www.interaction-design.org/literature/topics/wicked-problems

WICKED PROBLEMS & DESIGNING FOR EMERGENCE

Design theorist and academic Richard Buchanan connected design thinking to wicked problems in his 1992 paper "Wicked Problems in Design Thinking." http://www.jstor.org/stable/1511637

Design thinking's iterative process is extremely useful in tackling ill-defined or unknown problems—reframing the problem in human-centric ways, creating many ideas in brainstorming sessions, and adopting a hands-on approach in prototyping and testing.

https://www.interaction-design.org/literature/topics/wicked-problems

ETHICAL VALUES FOR AUTONOMOUS INTELLIGENT SYSTEMS

Reliability	AIS solutions should be sufficiently reliable for the purposes for which they are being used. Users need to be confident that the collected data is reliable, and that the system does not forward the data to anyone who should not have it.
Safety	Safety is an emerging property of a socio-technical system, which is created daily by decisions and activities. Safety of a system should be verified where applicable and feasible. Need to consider possible liability and insurance implications.
Security	AI should be secure in terms of malicious acts and intentional violations (unauthorized access, illegal transfer, sabotage, terrorism, etc.). Security of a system should be verified where applicable and feasible.
Accountability	Decisions and actions should be explained and justified to users and other stakeholders with whom the system interacts.

Dodig-Crnkovic, G., Holstein, T., & Pelliccione, P. (2021). **Future Intelligent Autonomous Robots, Ethical by Design. Learning from Autonomous Cars Ethics**. https://arxiv.org/abs/2107.08122

ETHICAL VALUES FOR AUTONOMOUS INTELLIGENT SYSTEMS

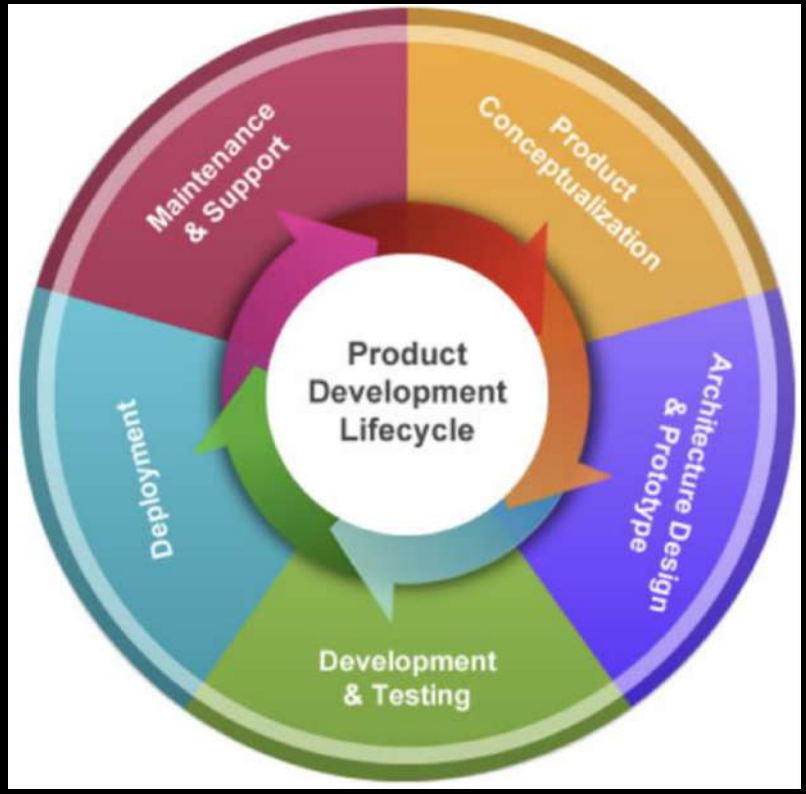
Explicability	Also 'explainability'; necessary in building and maintaining citizen's trust (captures the need for accountability and transparency), and the precondition for achieving informed consent from individuals.
Sustainability	The risks of AIS being misused should be minimized: Awareness and education. "Precautionary principle": Scientific uncertainty of risk or danger should not hinder to start actions of protecting the environment or to stop usage of harmful technology.
Role of technology in society	Governance: Society should use AIS in a way that increases the quality of life and does not cause harm to anyone. Depending on what type of theory of justice a society is committed to, it may stress e.g., the principle of social justice (equality and solidarity), or the principle of autonomy (and values of individual freedom and choice).

Dodig-Crnkovic, G., Holstein, T., & Pelliccione, P. (2021). **Future Intelligent Autonomous Robots, Ethical by Design. Learning from Autonomous Cars Ethics**. https://arxiv.org/abs/2107.08122

Ethical Framework in the Design of Emergent Technology

Ethical requirements must be fulfilled in all phases in the life-cycle of a product (autonomous car/robot) 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

Conclusions on Design Centered HRI with Ethical, Legal and Social Implications

We need a broad inter-disciplinary discussion about the values of the future intelligent, autonomous and social robotics .

Discuss the real-world ethical challenges and opportunities of emerging technologies.

Assure transparency to support evaluations by independent organisations/experts

Address context-dependencies of a given technology in order to identify value preferences. Stakeholder's involvement is essential as well as agreed upon values. Systemic view is vital

We propose Ethicality as non-functional property in requirements engineering.

Education in ELSI for engineering students and adequate information of all stakeholders is essential.

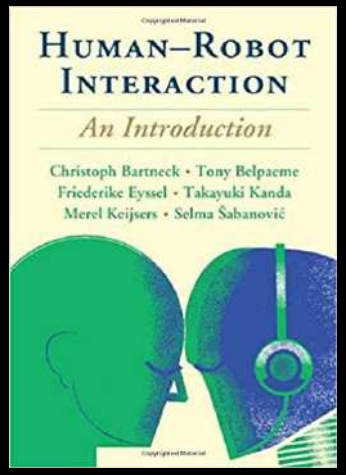
ELSI are wicked problems that typically have no "final solution" but the best current solution given constraints, that is constantly adapted to technological and social development.

WHAT THIS LECTURE CAN OFFER TO YOU IS JUST OPENING A WINDOW WITH A VIEW

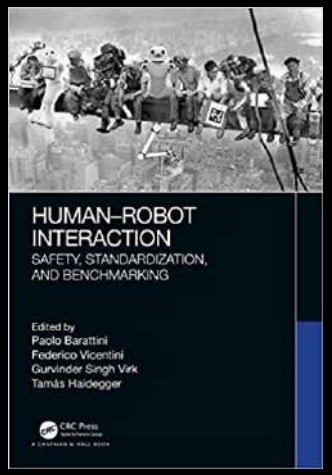
References

- 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.

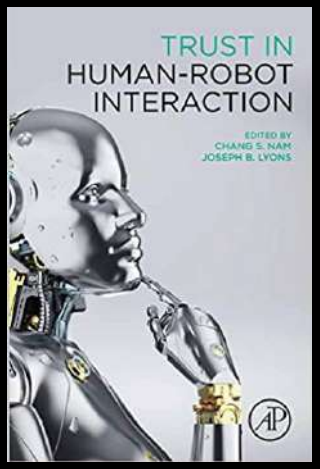
ROBOT-HUMAN INTERACTION LITERATURE



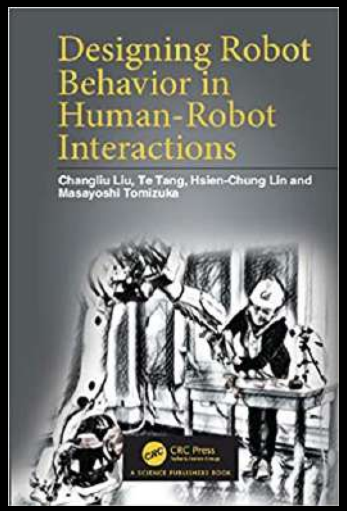
https://www.amazon.com/Human-Robot-Interaction-Introduction-Christoph-Bartneck/dp/1108735401?asin=B0845ZM9M4&revisionId=fixed_format&format=4&depth=2



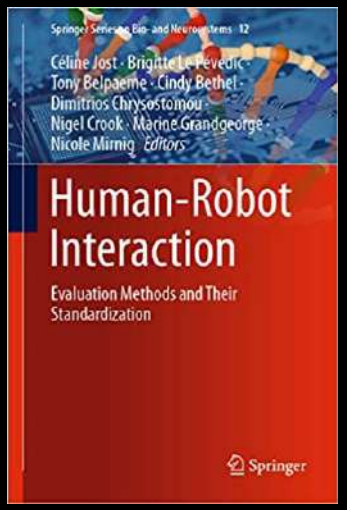
https://www.amazon.com/Human-Robot-Interaction-Safety-Standardization-Benchmarking-ebook-dp-B07QM9VTR8/dp/B07QM9VTR8/ref=mt_other?_encoding=UTF8&me=&qid=1628325169



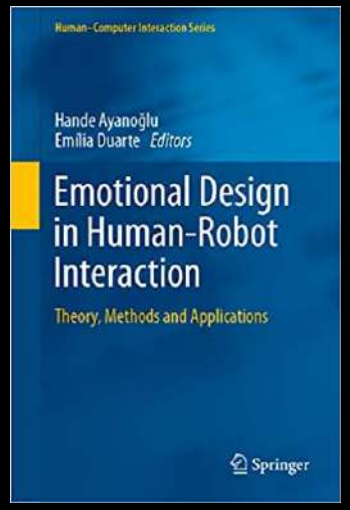
https://www.amazon.com/Trust-Human-Robot-Interaction-Chang-Nam-ebook/dp/B08NVM7SGF/ref=sr_1_3?dchild=1&keywords=Human-Robot+Interaction&qid=1628325169&s=books&sr=1-3



https://www.amazon.com/Designing-Robot-Behavior-Human-Robot-Interactions/dp/0367179695/ref=sr_1_4?dchild=1&keywords=Human-Robot+Interaction&qid=1628325169&s=books&sr=1-4



https://www.amazon.com/Human-Robot-Interaction-Evaluation-Standardization-Neurosystems-ebook/dp/B088LYNPYB/ref=sr_1_7?dchild=1&keywords=Human-Robot+Interaction&qid=1628325169&s=books&sr=1-7



https://www.amazon.com/Emotional-Design-Human-Robot-Interaction-Human-Computer-ebook-dp-B07FR6QZPQ/dp/B07FR6QZPQ/ref=mt_other?_encoding=UTF8&me=&qid=1628325169