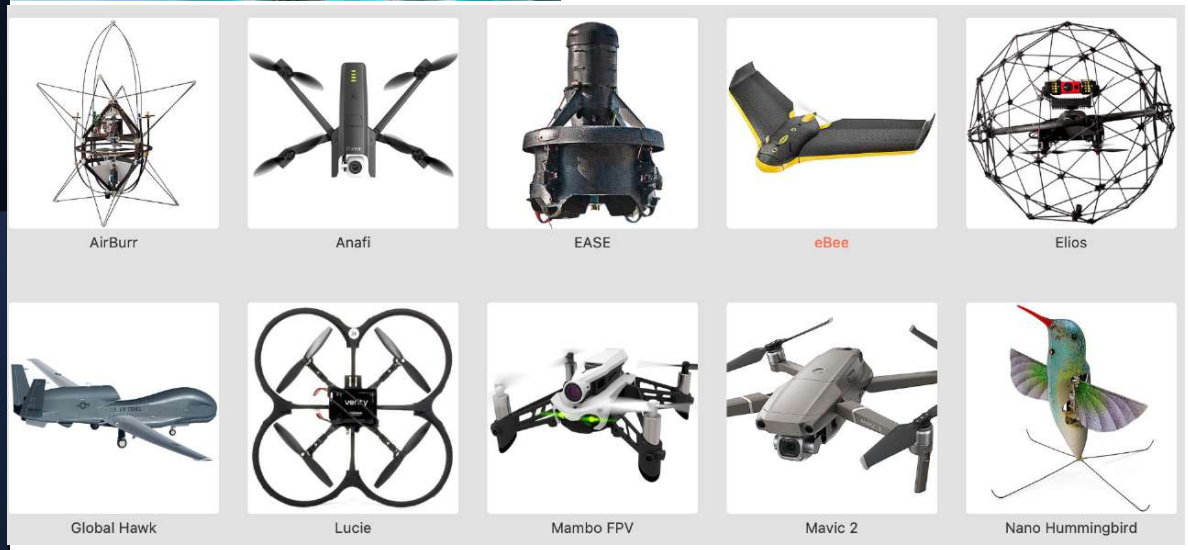
DEMYSTIFYING & DOMESTICATING AI ROBOTS ETHICAL BY DESIGN – Future intelligent robots, promises, and challenges

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



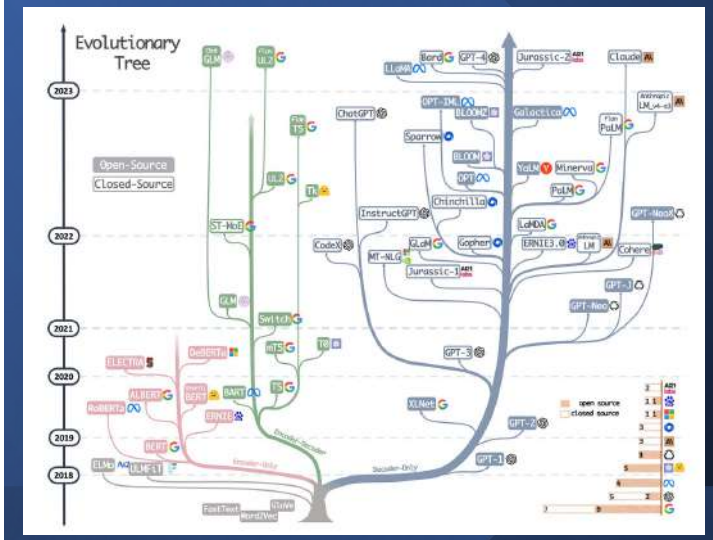
https://techdissected.com/editorials-and-discussions/advantages-of-having-a-humanoid-robot

AI SPRING STARTING IN NOVEMBER 2022

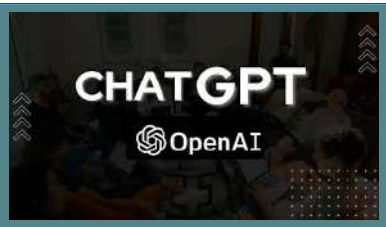
ChatGPT was launched on November 30, 2022,

by San Francisco–based OpenAI

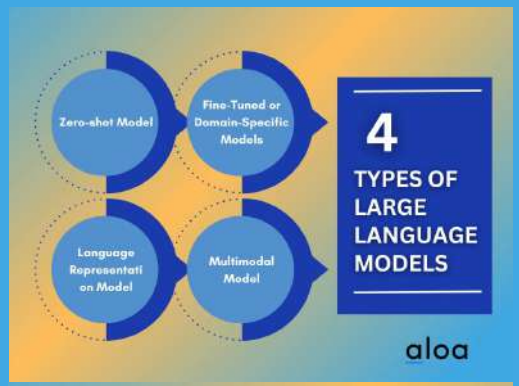
(the creator of the initial GPT series of large language models; DALL. E 2, a diffusion model used to generate images; and Whisper, a speech transcription model).

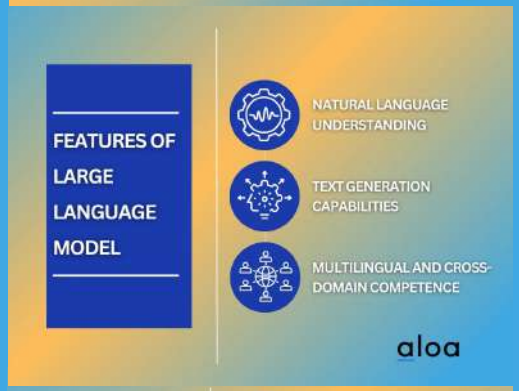


The evolutionary tree of modern LLMs
https://arxiv.org/abs/2304.13712



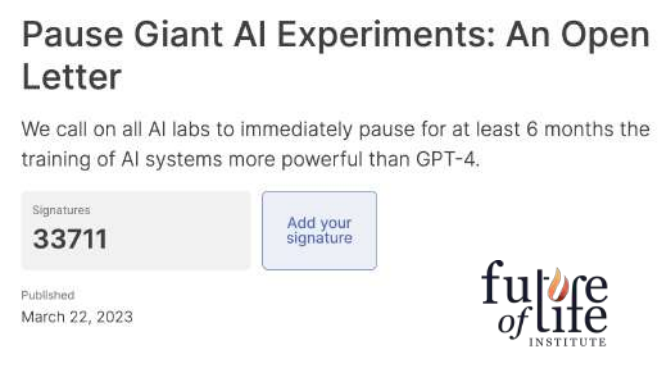
https://backlinko.com/chatgpt-alternatives
https://mobisoftinfotech.com/resources/blog/chatgpt-alternatives/





https://aloa.co/blog/what-is-a-large-language-model-a-beginners-guide

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, Max Tegmark

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.

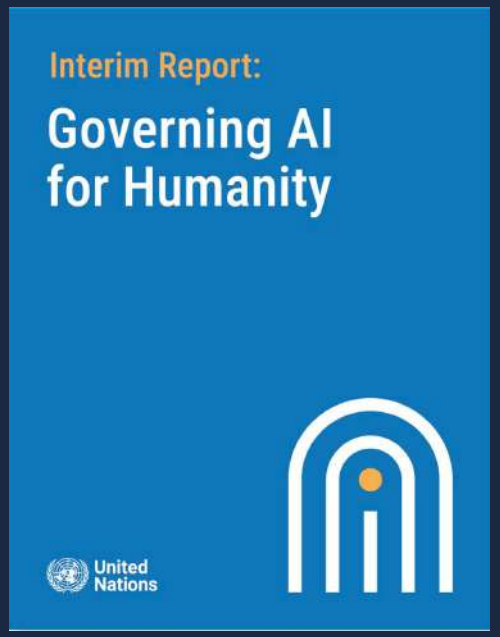
Other signatories include researchers from the Vector Institute and Mila, as well as professors from universities across Canada. Open AI CEO Sam Altman, Microsoft CTO Kevin Scott, etc.

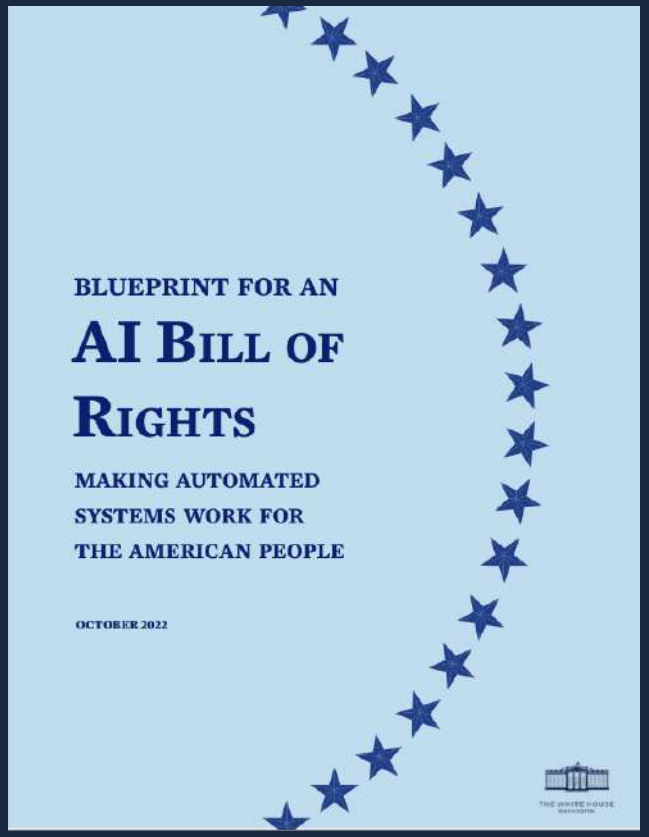
Academics, CEOs sign on in support of AI regulation and Bill C-27 as Canadian companies race to adopt the technology

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

Source: 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.

Can we trust intelligent systems?

Despite anecdotes from people who believed the GPT program was conscious and might persuade humans to behave irrationally and dangerously,

GPT of today has nothing to be conscious with. The reason is similar to the fact that the GPT program does not breathe - it has nothing to breathe with. Looking behind the scenes we can see that present AI is not conscious.

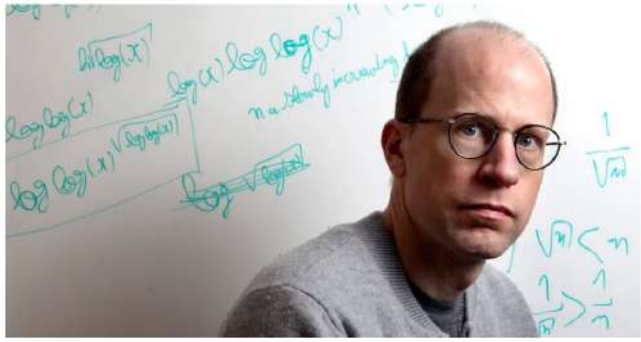
Global reactions to the emergence of GPT programs show how we humans are dependent on and deeply fascinated by language.

Before GPT, there was no entity capable of reasonable dialogue, and not having functioning consciousness.

How is that possible for GPT?

However, it does not mean that in the future a machine cannot be designed to be conscious.

https://www.wired.com/story/nick-bostrom-fear-ai-fix-everything/?bxid=5cec24ecfc942d3ada053185&cndid=55112229&esrc=bounceX&source=Email_0_EDT_WIR_NEWSLETTER_0_DAILY_ZZ&utm_brand=wired&utm_campaign=aud-dev&utm_content=WIR_Daily_050524&utm_mailing=WIR_Daily_050524&utm_medium=email&utm_source=nl&utm_term=WIR_Daily_Active



FAST FORWARD | 5-MINUTE READ

Nick Bostrom Made the World Fear AI. Now He Asks: What If It Fixes Everything?

BY WILL KNIGHT

Philosopher Nick Bostrom popularized the idea that superintelligent AI could erase humanity. His new book imagines a world in which algorithms have solved every problem.

FEAR OF ARTIFICIAL INTELLIGENCE?
NLP, ML AND LLMS BASED DISCOVERY OF AI-PHOBIA AND FEAR SENTIMENT PROPAGATION BY AI NEWS

WORKING PAPER (PREPRINT) – RELEASED FOR RAISE-24 ON 3/9/2024

Jim Samuel
Rutgers University
jim.samuel@rutgers.edu

Tanya Khanna
Rutgers University

Srinivasaraghavan Sundar
Rutgers University

ABSTRACT

Confusion, fear and mixed sentiments prevail in the minds of people towards what is arguably one of the most important of dynamics of modern human society: Artificial Intelligence (AI). This study aims to explore the contributions of news media towards this phenomenon - we analyze nearly seventy thousand recent news headlines on AI, using natural language processing (NLP) informatics methods, machine learning (ML) and large language models (LLMs) to draw insights and discover dominant themes. Our theoretical framework was derived from extant literature which posits the power of fear producing articles and news headlines which produce significant impacts on public behavior even when available in small quantities. We applied extensive textual informatics methods using word and phrase frequency analytics, sentiment analysis and human experts based thematic analysis to discover insights on AI phobia inducing news headlines.

Our rigorous analysis of nearly seventy thousand headlines using multiple validation methods in NLP (exploratory informatics including BERT, Llama 2 and Mistral[1; 2] based topic identification), ML (supervised informatics) and LLMs (neural nets for sentiment classification, with BERT, Llama 2 and Mistral) demonstrates the presence of an unreasonable level of emotional negativity and fear inducing verbiage in AI news headlines. The framing of AI as being dangerous or as being an existential threat to humanity can have a profound impact on public perception, and the resulting AI phobia and confusion in public perceptions are inherently detrimental to the science of AI. Furthermore, this can also impact AI policy and regulations, and harm society. We conclude with a discussion deducing implications for society and make recommendations for education and policies that could support human identity and dignity.

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4755964

"Domestication of ignorant entities"

"Eco-cognitive computationalism considers computation in context, following some of the main tenets advanced by the recent cognitive science views on embodied, situated, and distributed cognition. "

"Through eco-cognitive computationalism we can clearly acknowledge that the concept of computation changes, depending on historical and contextual causes, and we can build an epistemological view that illustrates the "emergence" of new kinds of computations, such as the one regarding morphological computation. This new perspective shows how the computational domestication of ignorant entities can originate new unconventional cognitive embodiments."

Lorenzo Magnani (2021) Computational domestication of ignorant entities Synthese 198(11)
DOI: 10.1007/s11229-020-02530-5

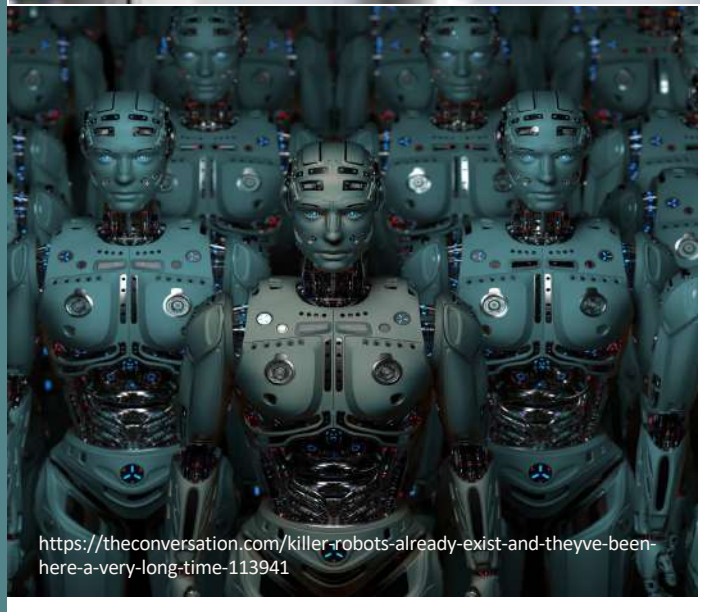
How can we trust intelligent robots?

Trusting the intelligent autonomous robot technology presupposes their anticipated beneficial influence on societies and individuals, globally. Questions 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 robotics are good examples of ethical 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 the 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.

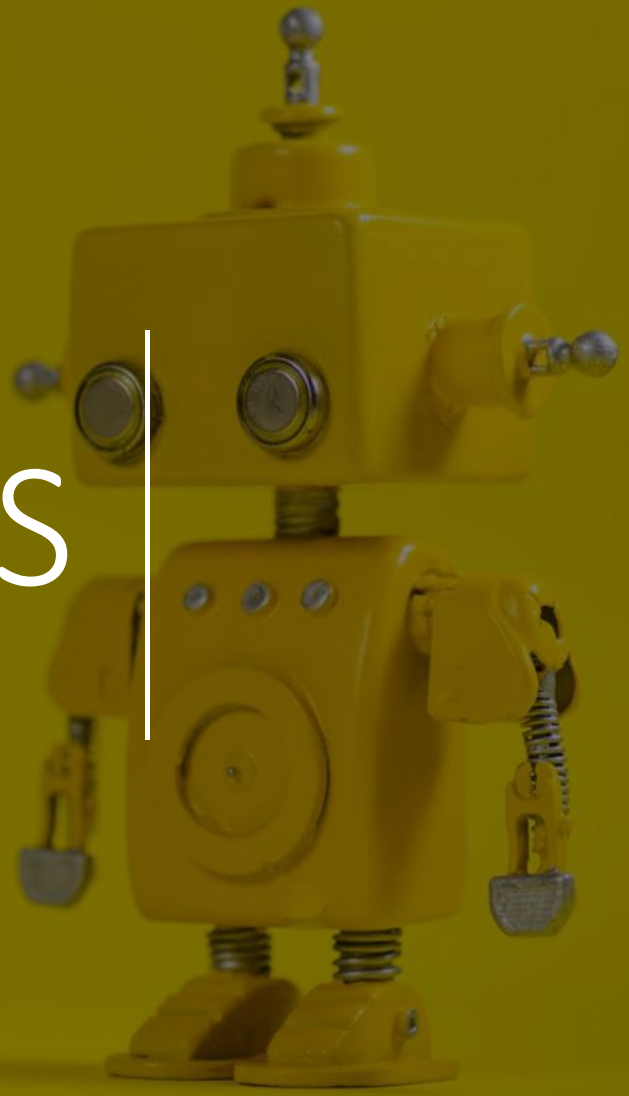
https://robohub.org/from-disembodied-bytes-to-robots-that-think-and-act-like-humans/



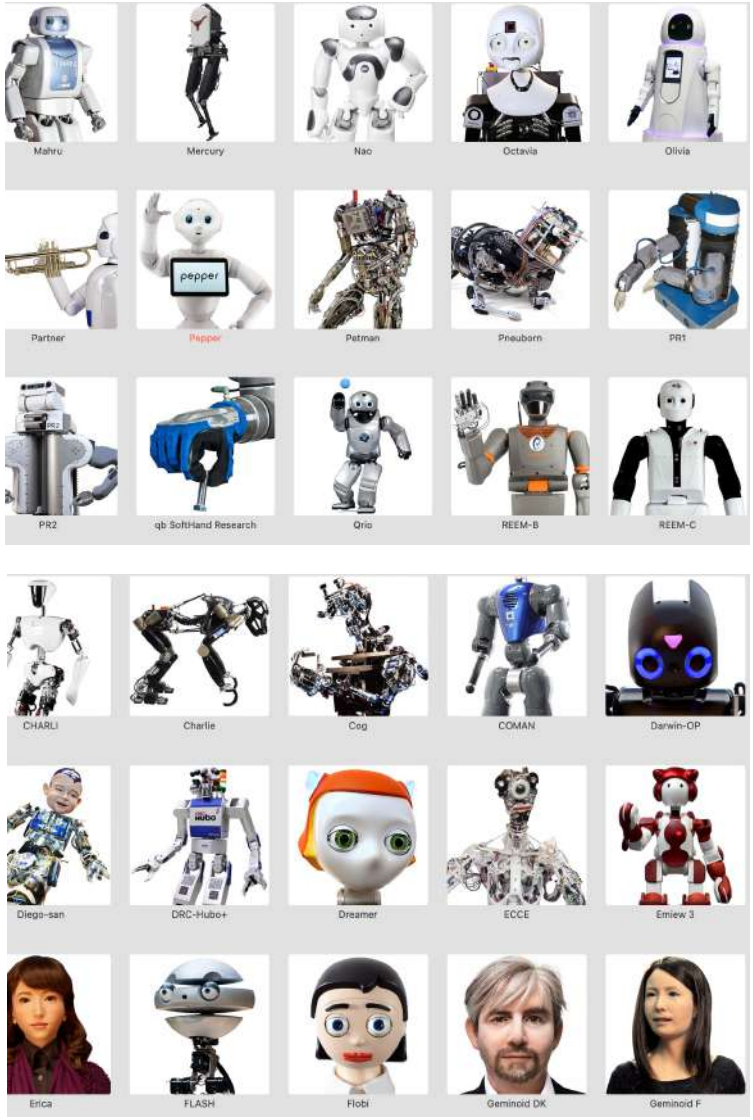
https://theconversation.com/killer-robots-already-exist-and-theyve-been-here-a-very-long-time-113941

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



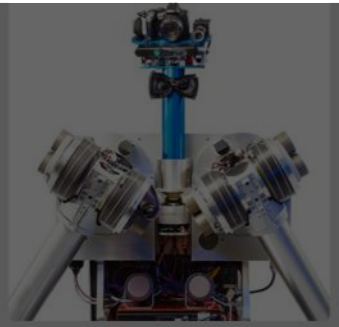
https://robots.ieee.org/

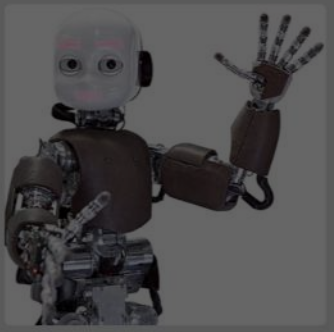


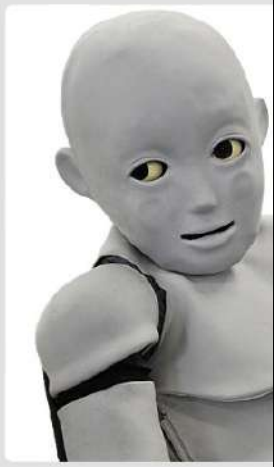
Humanoid Robots

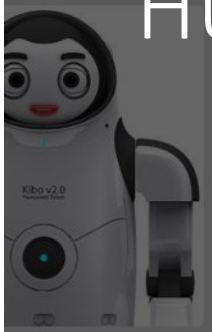
https://robots.ieee.org/

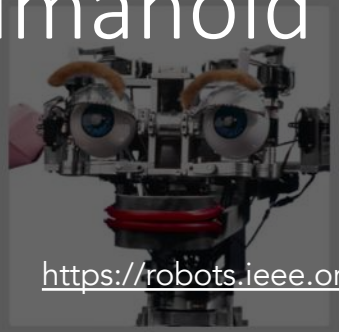
Humanoid Robots

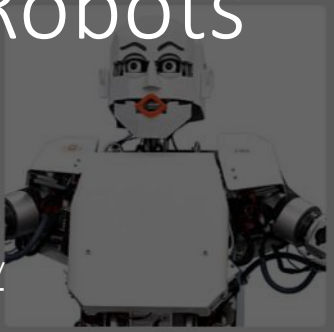


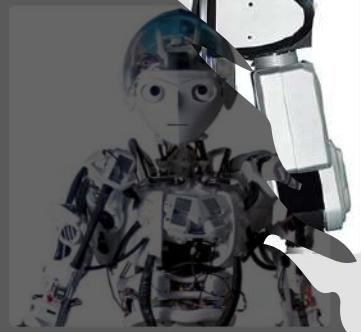




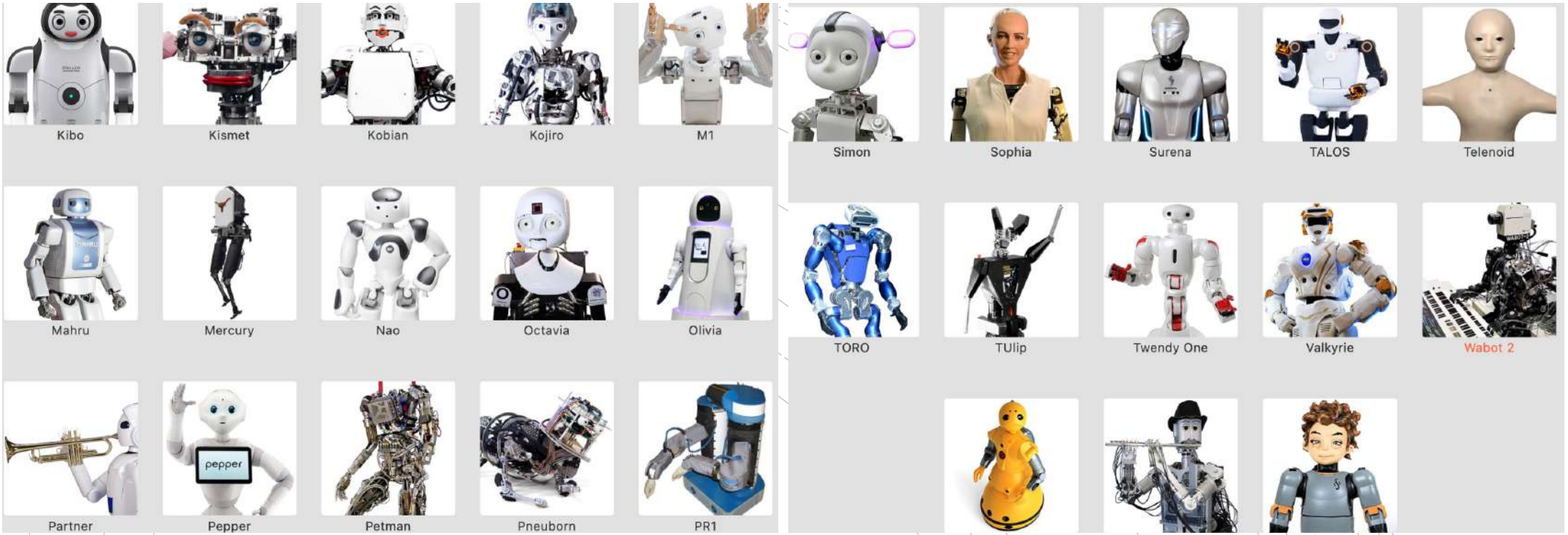






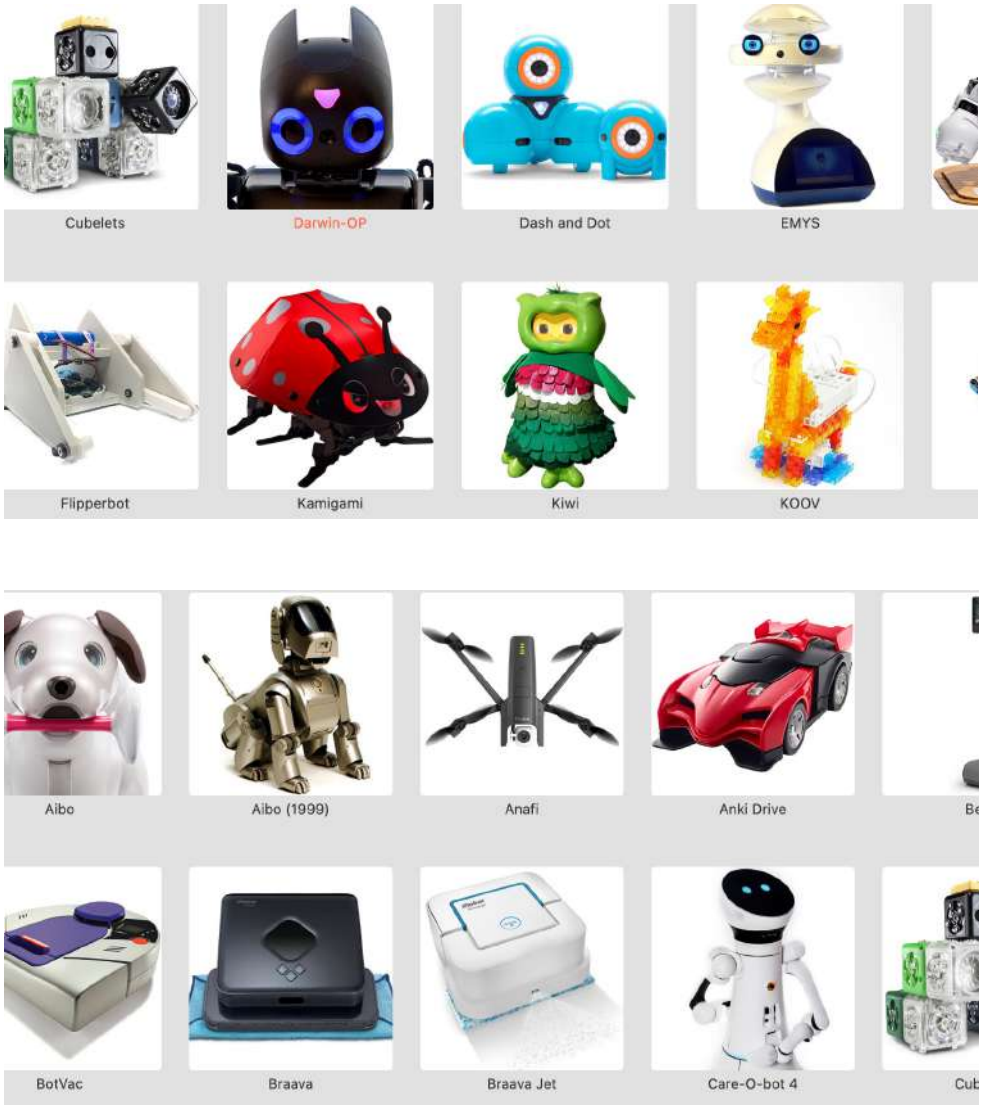


https://robots.ieee.org/



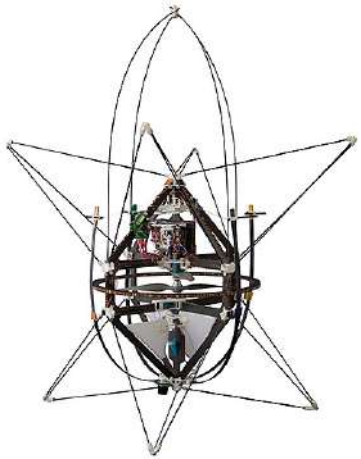
Humanoid Robots

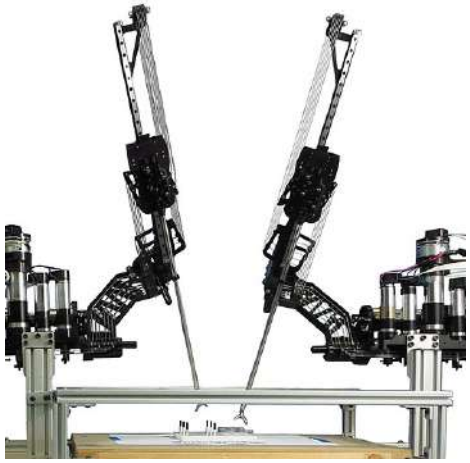
https://robots.ieee.org/



Education & Consumer Robots

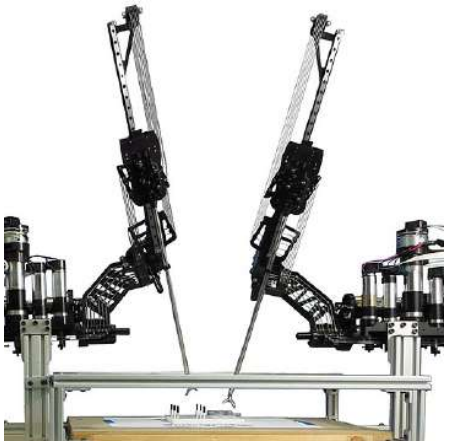
https://robots.ieee.org/





Research Robots

https://robots.ieee.org/



Medical Robots

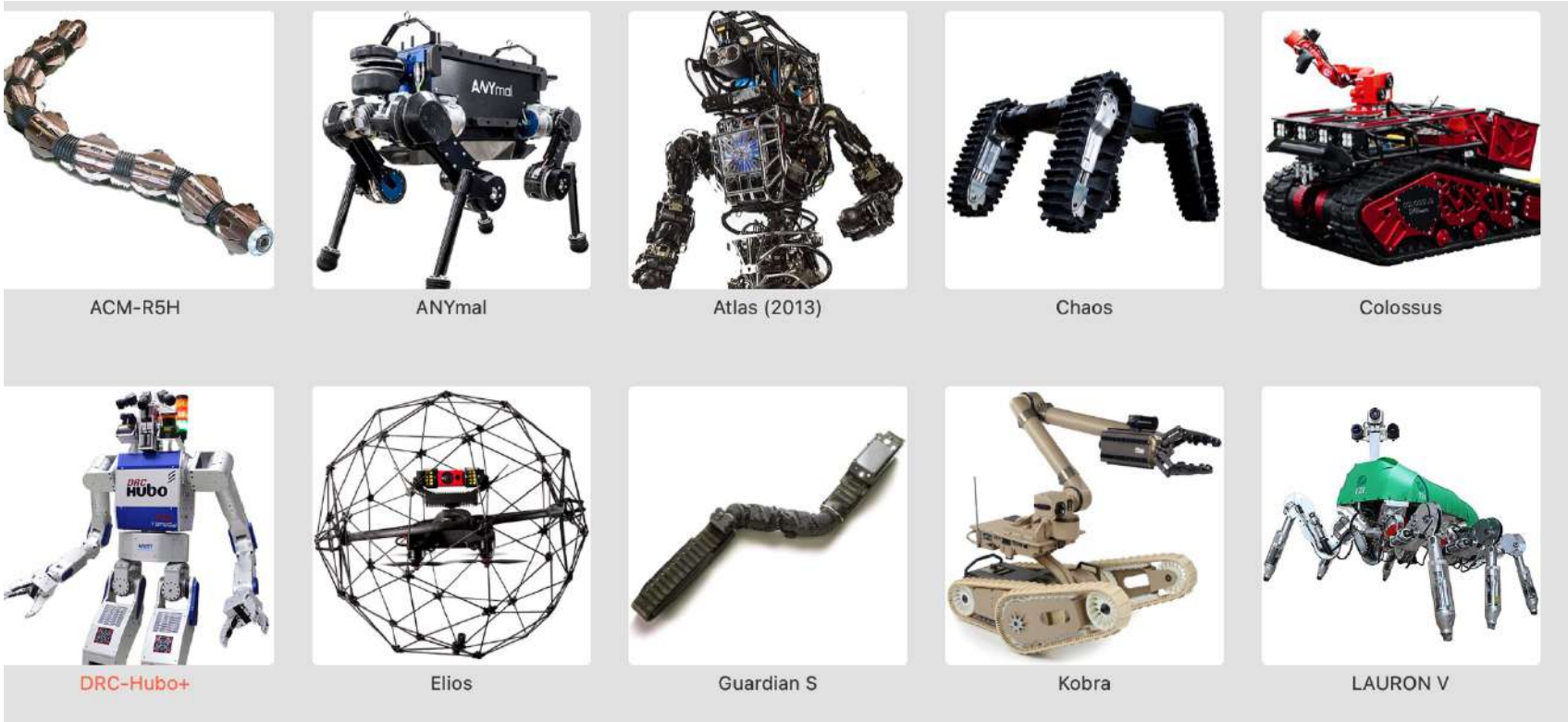
https://robots.ieee.org/

Nano Robots

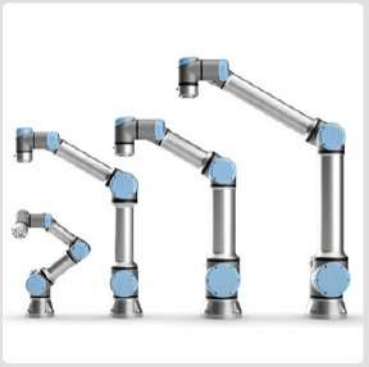
https://www.europeanpharmaceuticalreview.com/news/

Disaster response

https://robots.ieee.org/



Unimate


UR

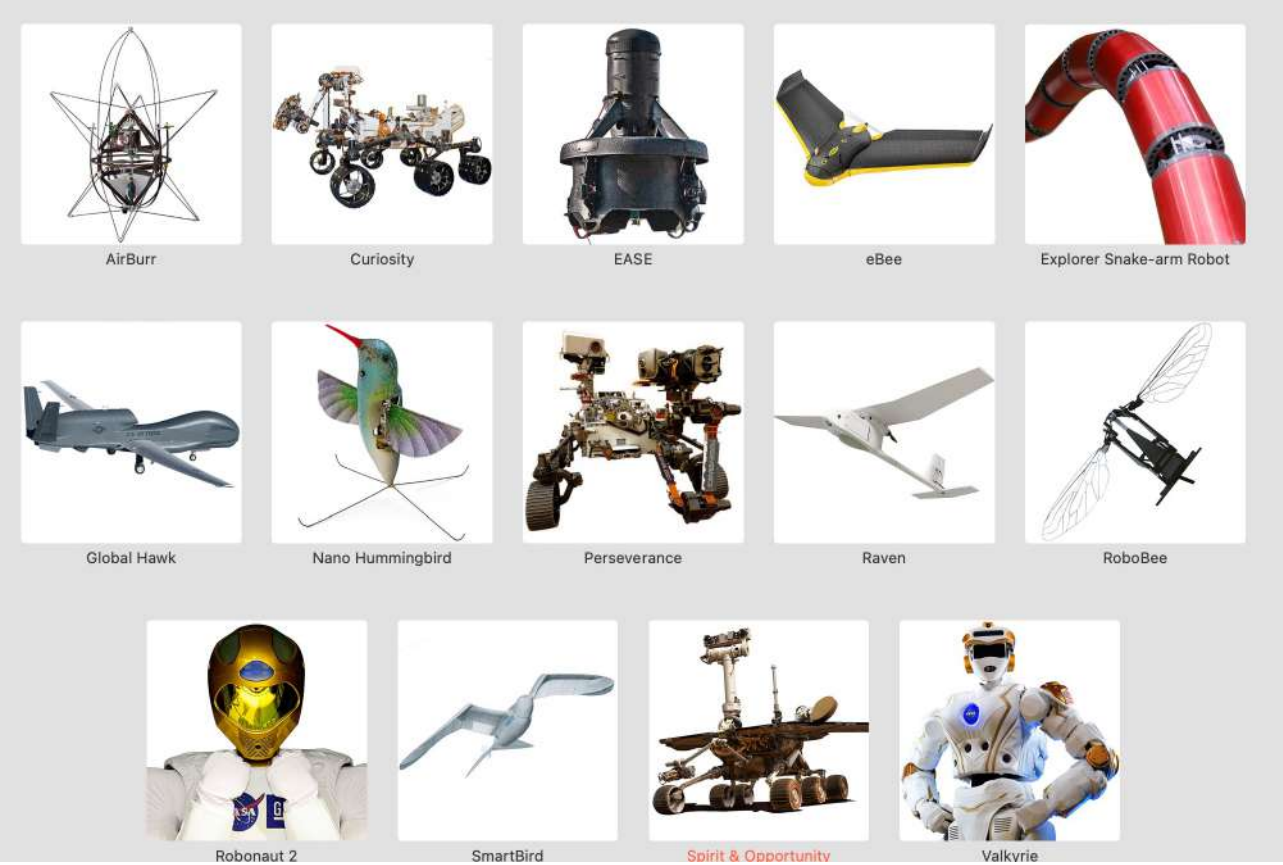

Versatrax


WAM

YuMi

Industrial Robots

https://robots.ieee.org/



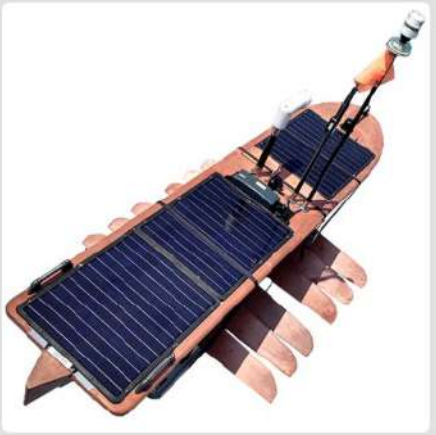
Aerospace Robots

https://robots.ieee.org/

ACM-R5H

Aqua2

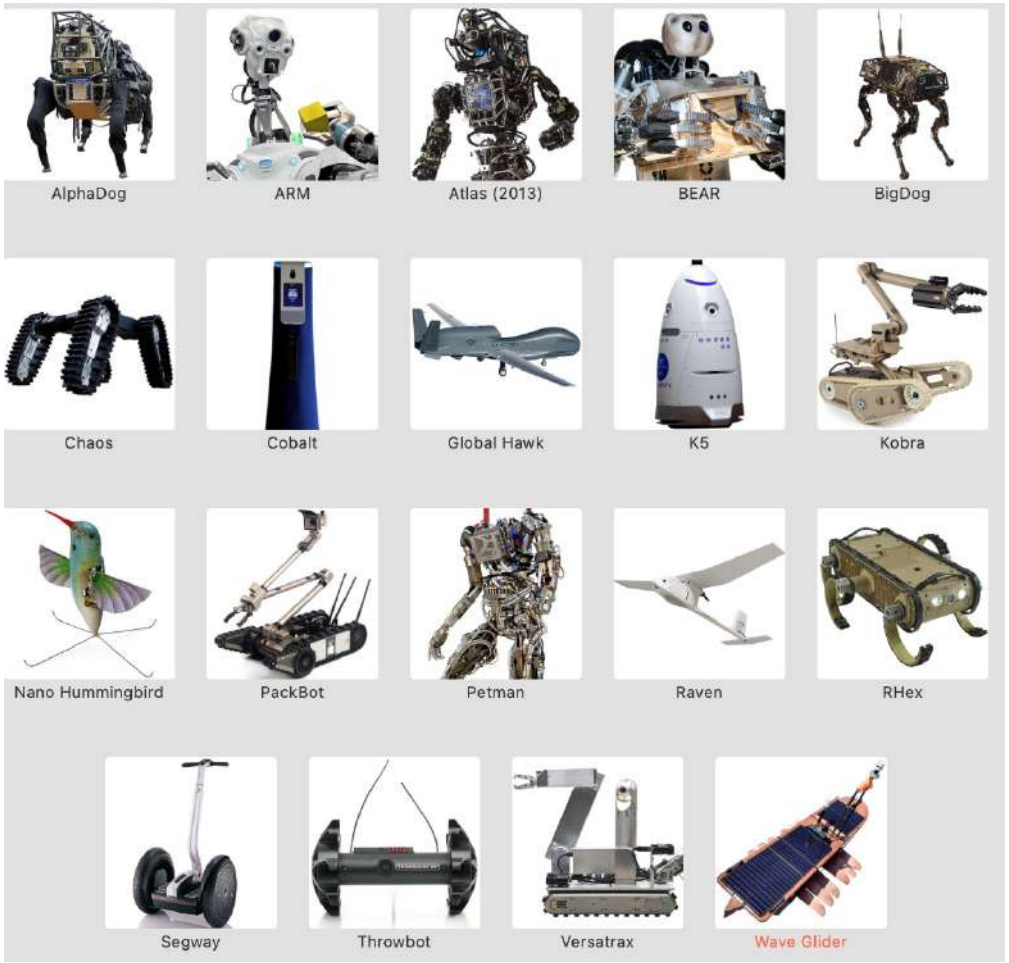
Aquanaut



Wave Glider

Underwater Robots

https://robots.ieee.org/

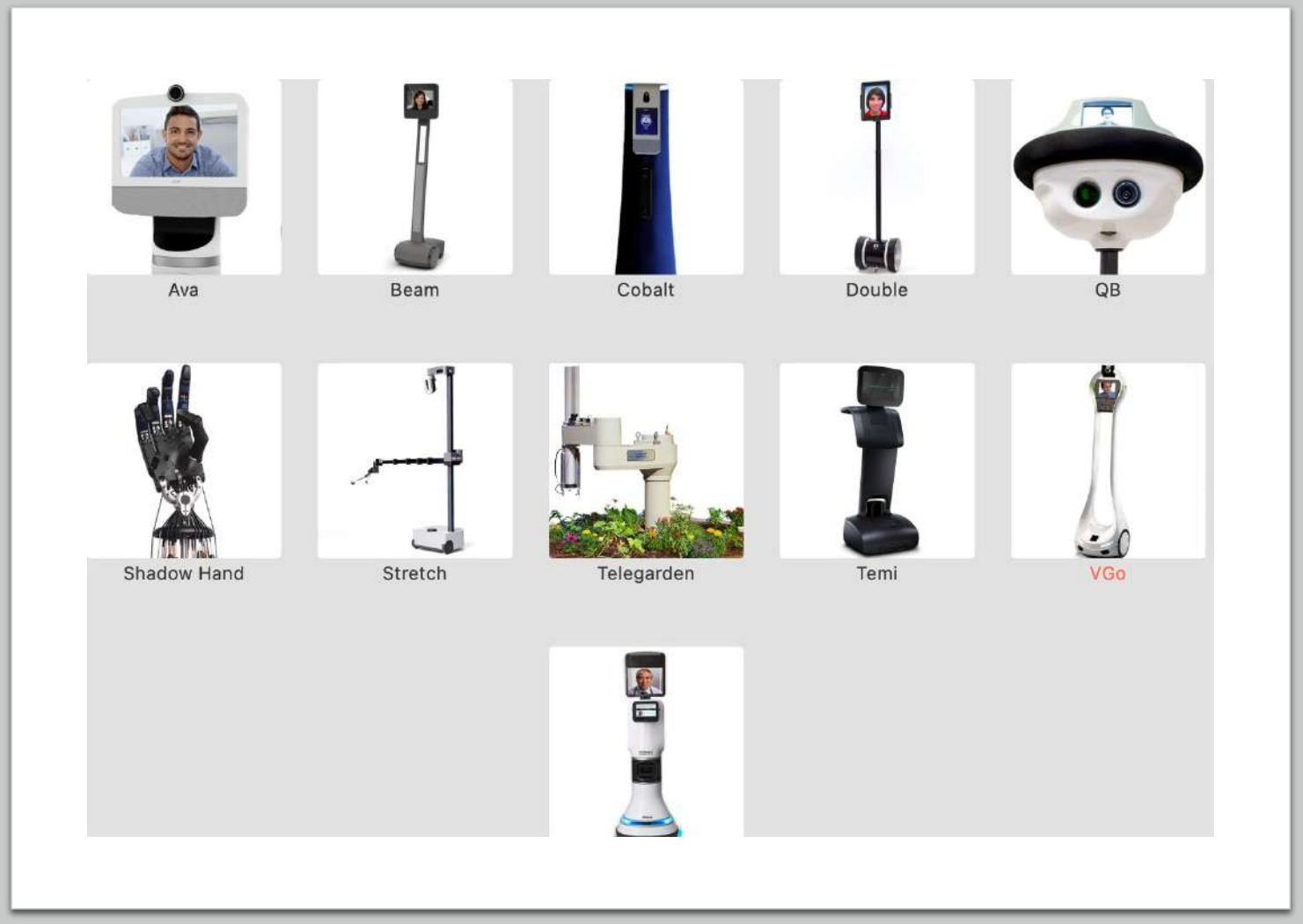


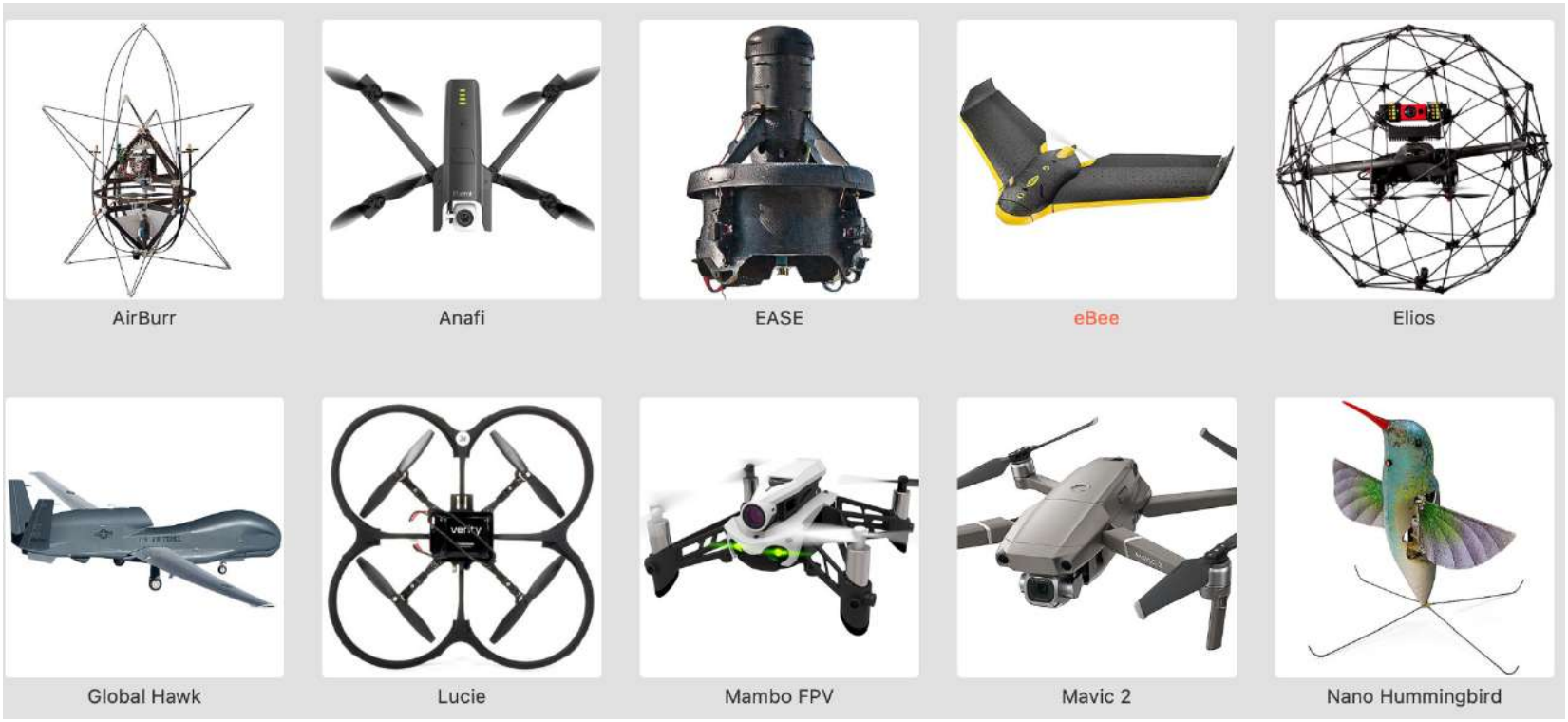
Military and Security Robots

https://robots.ieee.org/

Telepresence Robots

https://robots.ieee.org/





Drones

https://robots.ieee.org/

Autonomous Cars

https://robots.ieee.org/

Boss

Google Self-Driving Car

nuTonomy

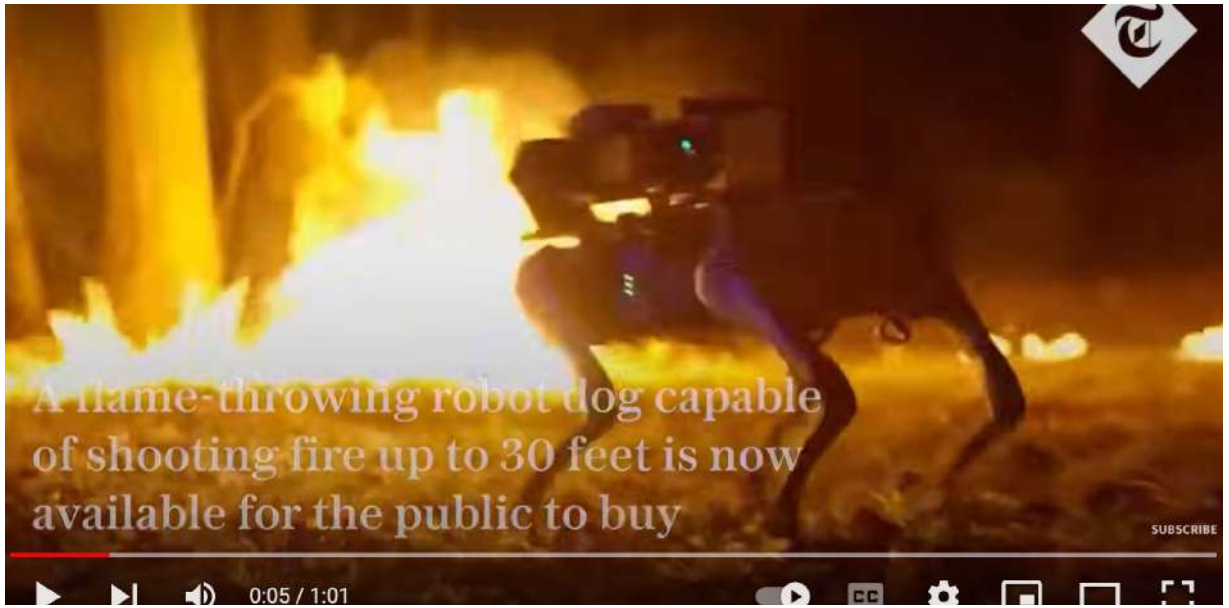
Stanley

Waymo

A flame - throwing robot

https://www.youtube.com/watch?v=U83BfU1phCw

https://www.youtube.com/watch?v=b5dE3vxWP9E



Intelligent, ethical robot according to GPT4 (Dall-E)

ME: Can you please generate **an image that represents AI ethics?**

GPT: Here's an image that represents AI ethics, symbolized by a digital brain, balanced scales, human silhouettes, interconnected gears and circuits, and emanating light rays against a futuristic blue and white background.

GPT4 (Dall-E)

ME: But I am thinking about humans in society designing ethical AI machines.

GPT: Here's an abstract image that visualizes the idea of humans from diverse backgrounds collaboratively designing an ethical AI, with the machine at the center featuring a heart symbol to represent human warmth. A light bulb glows above, symbolizing inspiration and the illumination of ethical ideas.

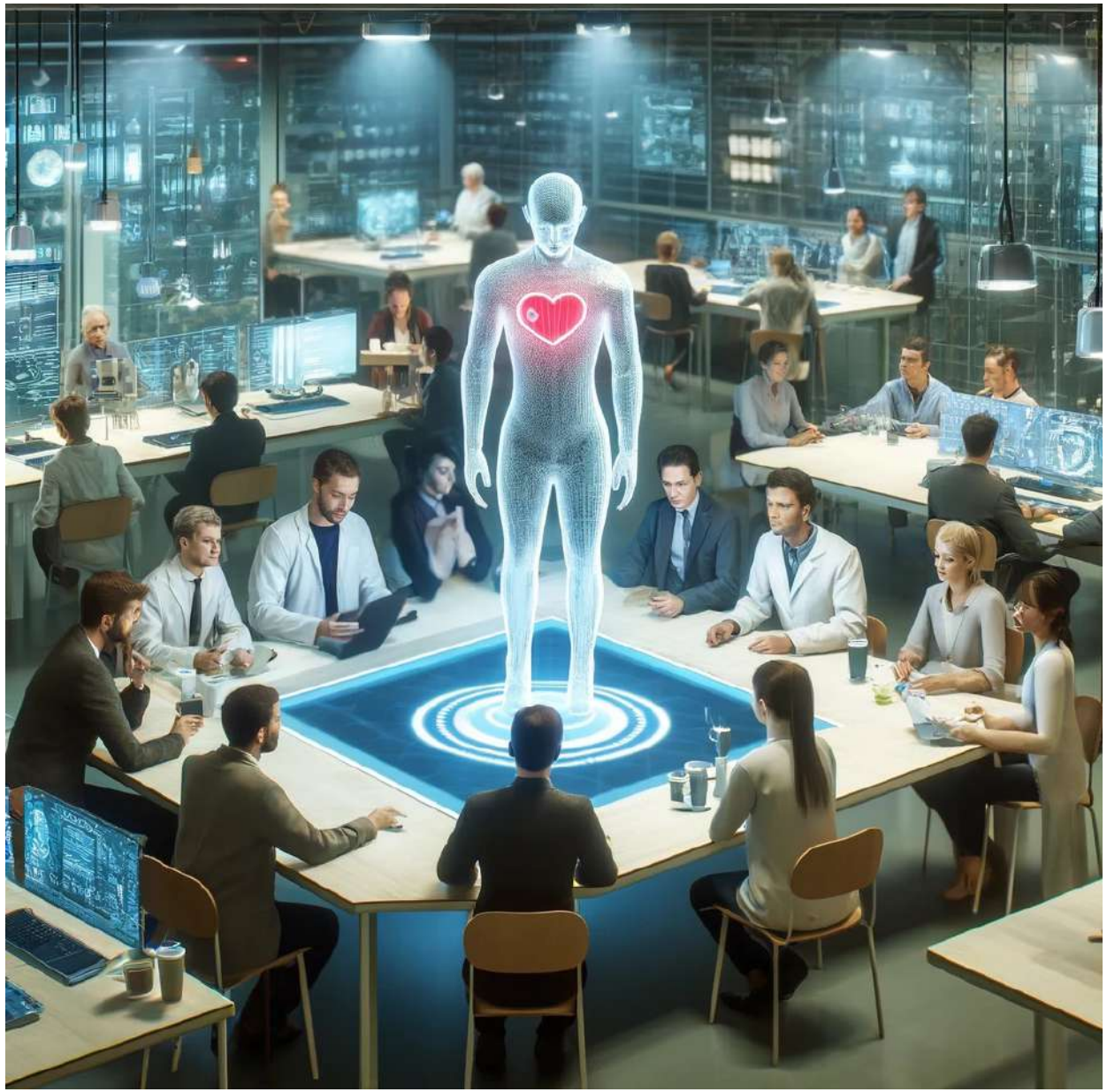
ME: But it looks like a scene in a musical!
I am thinking of real-life designers in the industry, creating future ethical AI for the good of humanity.

- GPT: This image captures the essence of real-life AI designers in an advanced tech environment, collaborating on ethical AI development, with a hologram of a humanoid robot featuring a glowing heart at the center.

- ME: Why is AI so huge in the picture?

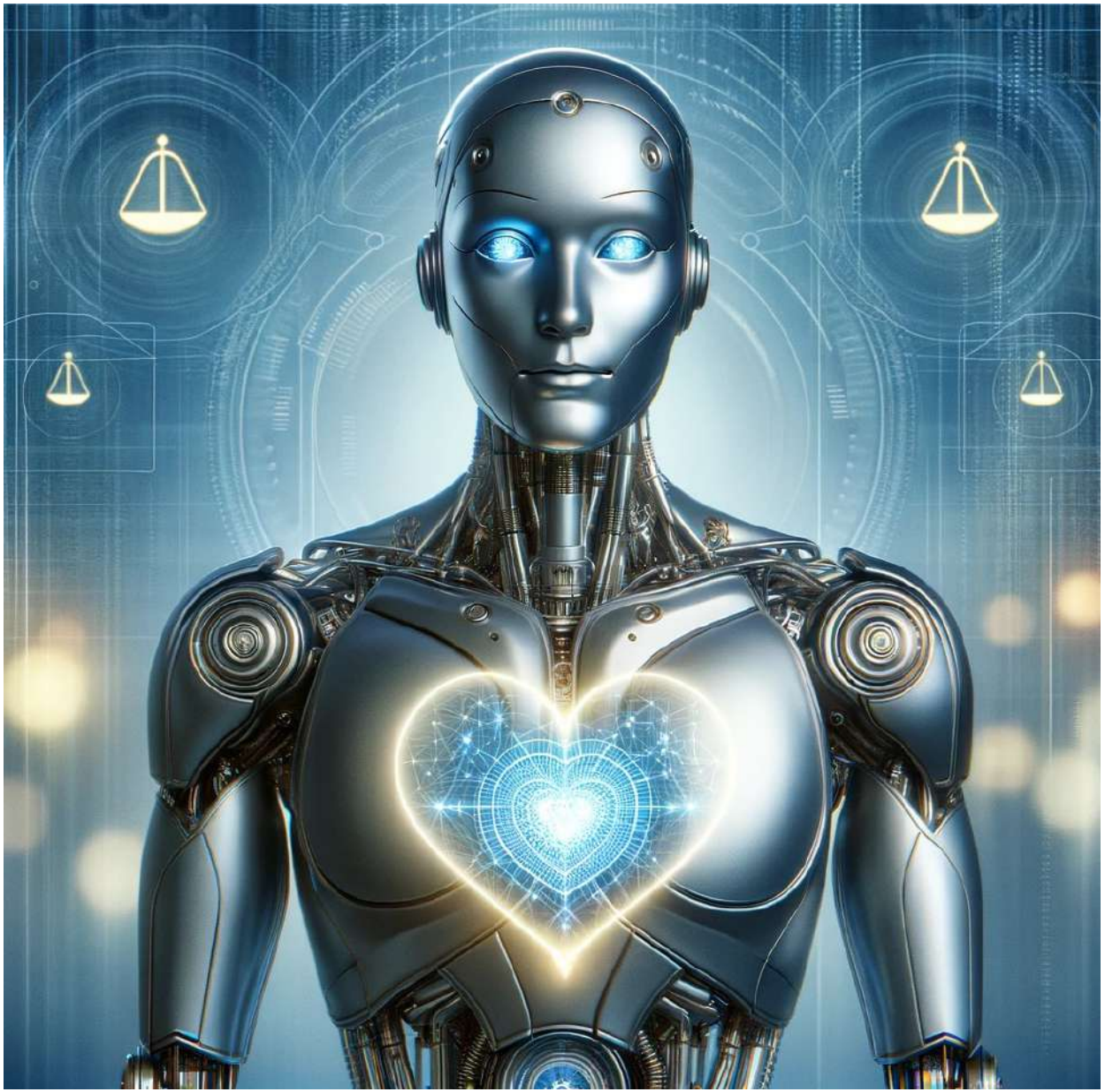
- GPT: The size of the AI hologram in the center is meant to be symbolic, emphasizing its significance and the central role of AI in the scene. It represents the focus of the designers' efforts and the impact AI has on their work and potentially on society. By making it larger than life, it dramatizes the concept of AI being at the heart of innovation and ethical considerations in technology.
Would you like the AI to be represented differently?

- ME: Yes. That big AI looks frightening



- GPT: In this image, the holographic AI model is designed to be human-sized and part of the collaborative process, integrating more comfortably with the team of engineers and ethicists at work.

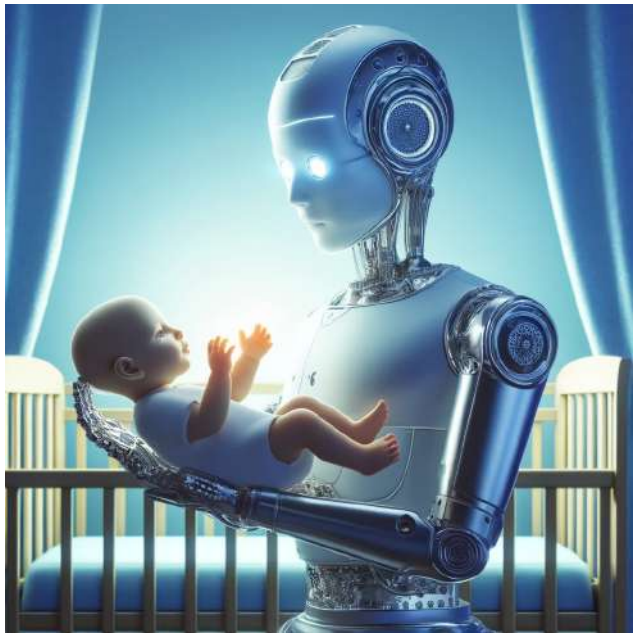
- ME: Thank you.

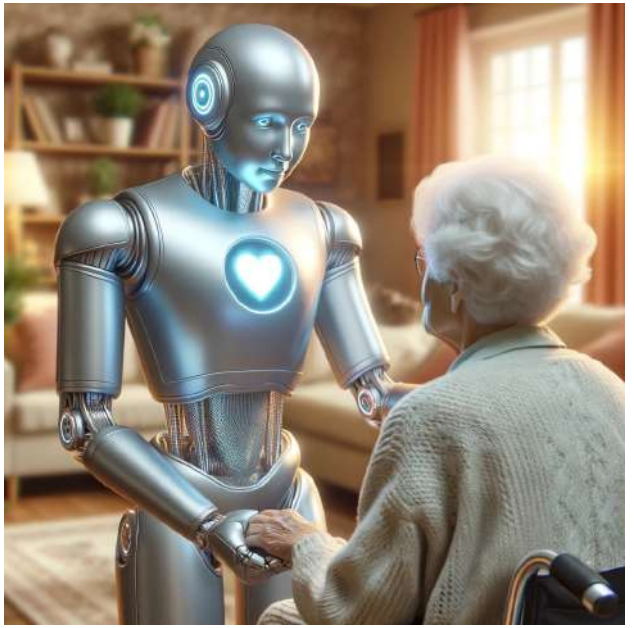


ME: Can you please generate an image that represents **an intelligent ethical robot**?

- GPT: Here's a representation of an intelligent, ethical robot with a heart symbol on its chest, set against a backdrop that hints at its commitment to ethical principles
- ME: Thank you.
- GPT: You're welcome! If you have any more ideas or questions, feel free to reach out. Have a great day!
-

Intelligent, ethical robot according to GPT4 (Dall-E)





Delegating responsibilities to intelligent autonomous systems: challenges and benefits

Gordana Dodig-Crnkovic, Gianfranco Basti, and Tobias Holstein

United Nations report "Governing AI for Humanity" and EU's "AI Act" emphasize the human role in ethical AI development, advocating for inclusive governance and continuous ethical oversight of socio-technological systems. We explore the concept of distributed responsibility in a network of agents, drawing on perspectives that distinguish between human ethical deliberation and machine responsiveness where AI is seen as a part of a larger interconnected system with shared responsibilities.

It is important to acknowledge the limitations of human judgment and actively work towards mitigating its consequences through careful design, the use of diverse competencies, continuous oversight, and constant systemic learning.

The discussion extends to the machine ethics approach, which integrates ethical principles into AI design, aiming for consistency, scalability, and alignment with human values. We argue for a multifaceted strategy that includes continuous learning, ethical education, and societal engagement to ensure the development of responsible AI. We identify the limitations of human judgment and the necessity for meticulous design and oversight to navigate the ethical landscape of AI integration into society.

THE ROLE OF HUMANS: TIME PERSPECTIVE

EVEN UNDER UNCERTAIN (OR RAPIDLY CHANGING) PREMISES, we must act, like in the white-water world of turbulent, non-linear systems, using hypothetical scenarios of speculative design.

Conversations on meaningful intelligent technologies situated in a social context.

VALUE ALIGNMENT

ENVIRONMENTAL, COGNITIVE& SOCIAL & SUSTAINABILITY ASPECTS

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

WHAT CAN WE LEARN FROM AUTONOMOUS CARS ABOUT ETHICS ASPECTS OF OTHER ROBOTS?

Autonomous Cars

Based on:

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

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

The Perspective

The aim of this lecture – to offer new views

As the topic of Design Ethics, AI ethics and even AV ethics are huge, what this lecture can do is to open the window with a view, giving you just 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

Wicked Problems in Design Thinking

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

TUANA. COMMUNICATIONS OF THE ACM | DECEMBER 2015 | VOL. 58 | NO. 12

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.

The human role in ethical AI development vs. artificial intelligent agency

The United Nations report "Governing AI for Humanity" and the EU's "AI Act" emphasize the human role in ethical AI development, advocating for inclusive governance and continuous ethical oversight of socio-technological systems.

We explored the concept of distributed responsibility in a network of agents, drawing on perspectives that distinguish between human ethical deliberation and machine responsiveness.

Autonomous AI is seen as a part of a larger socio-technological interconnected system with shared responsibilities.

More powerful virtual agents
"2023 was the year of being able to chat with an AI. Multiple companies launched something, but the interaction was always you type something in and it types something back," says Stanford's Peter Norvig.
"In 2024, we'll see the ability for agents to get stuff done for you. Make reservations, plan a trip, connect to other services."

https://www.ibm.com/blog/artificial-intelligence-trends/

https://www.youtube.com/watch?v=Boj9eD0Wug8 Mark Zuckerberg & Yuval Noah Harari in Conversation (01:25:00 Totalitarianism & Surveillance Capitalism)

Self-Driving (Autonomous) Cars as Intelligent Robotic System

We take Self-driving cars as an example of emerging technology that is combining advances in several underlying emergent technologies such as electric mobility and artificial intelligence (with connected driving, intelligent cities, intelligent infrastructure, etc.)

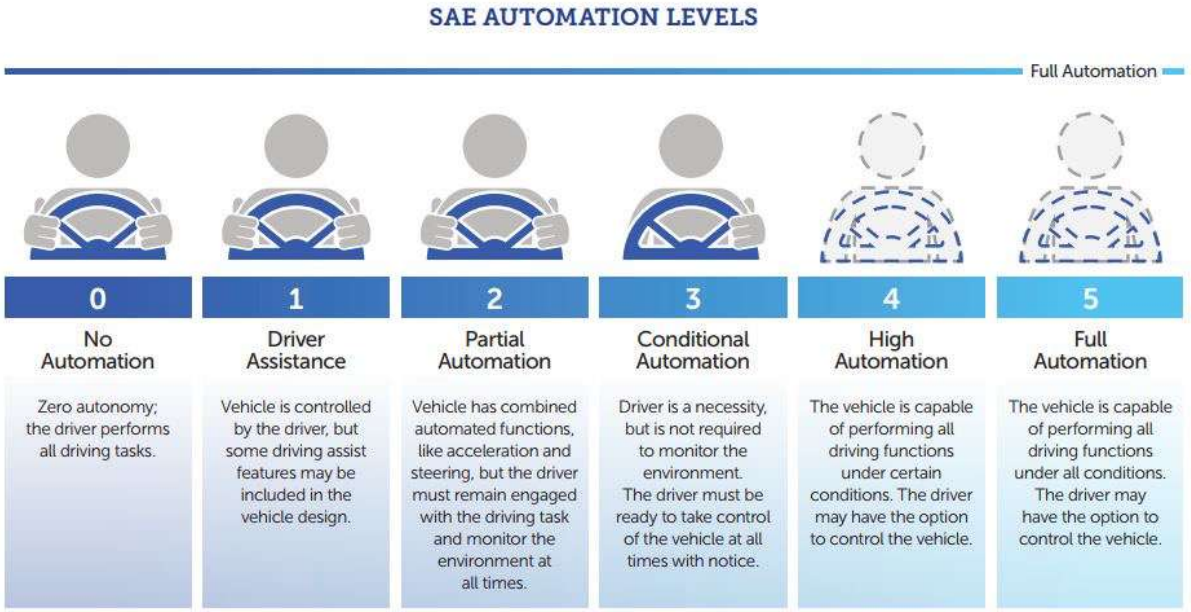
Technology emerges not in vaccuum but in its social context that today is global technosocial environment

How can we contribute in different roles as stakeholders to the development of good society with help of new powerful technologies. Who are the main actors/stakeholders and how do they affect the development? Autonompus cars have been studied a lot and we can learn from the development so far.

AUTONOMOUS CARS DEVELOPMENT

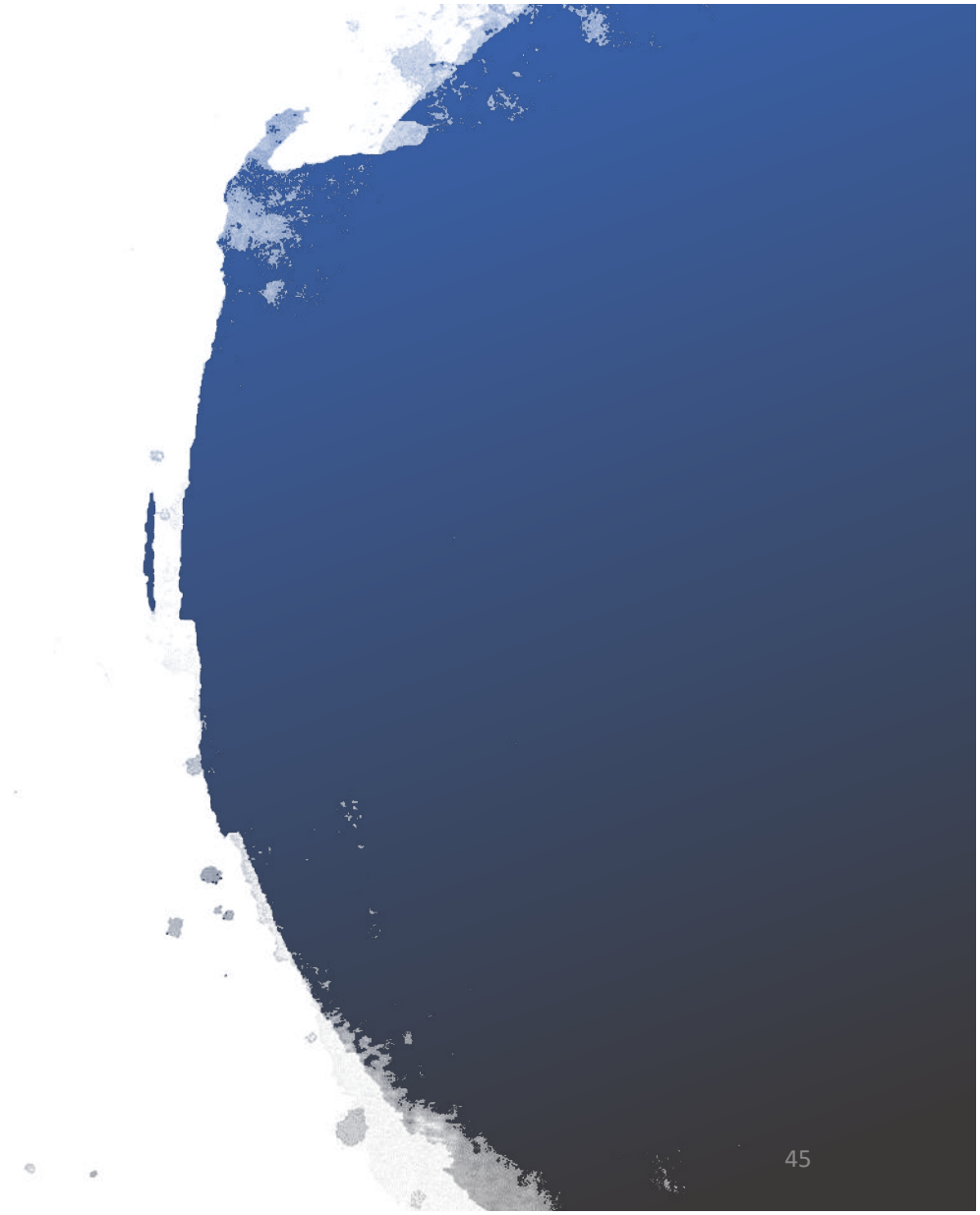
LEVELS OF AUTOMATION

Equivalent for Intelligent Robotics Needed!

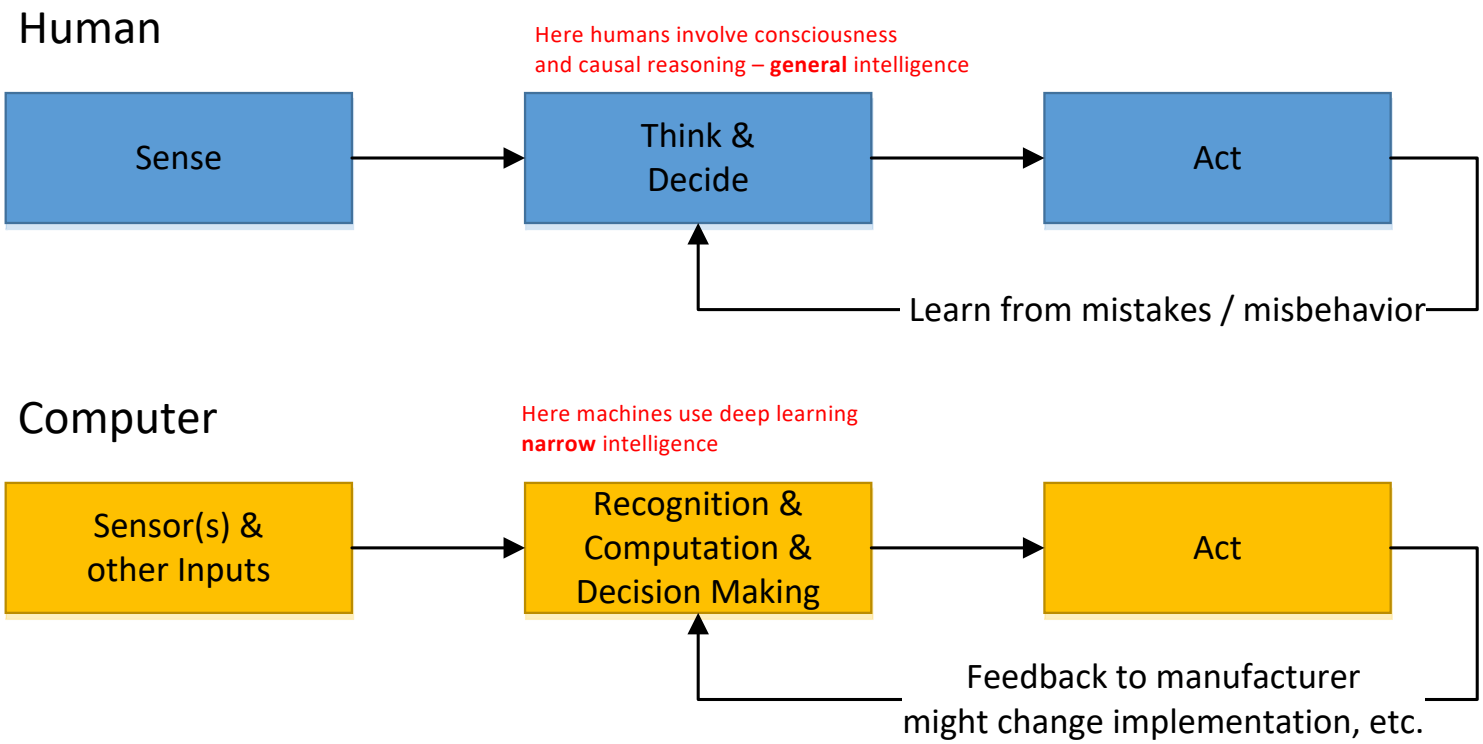


INTELLIGENCE OF AUTONOMOUS CARS – NARROW AI

(FAR FROM HUMAN LEVEL GENERAL AI)



Human Decision-making Process versus Self-Driving Car (Computer)



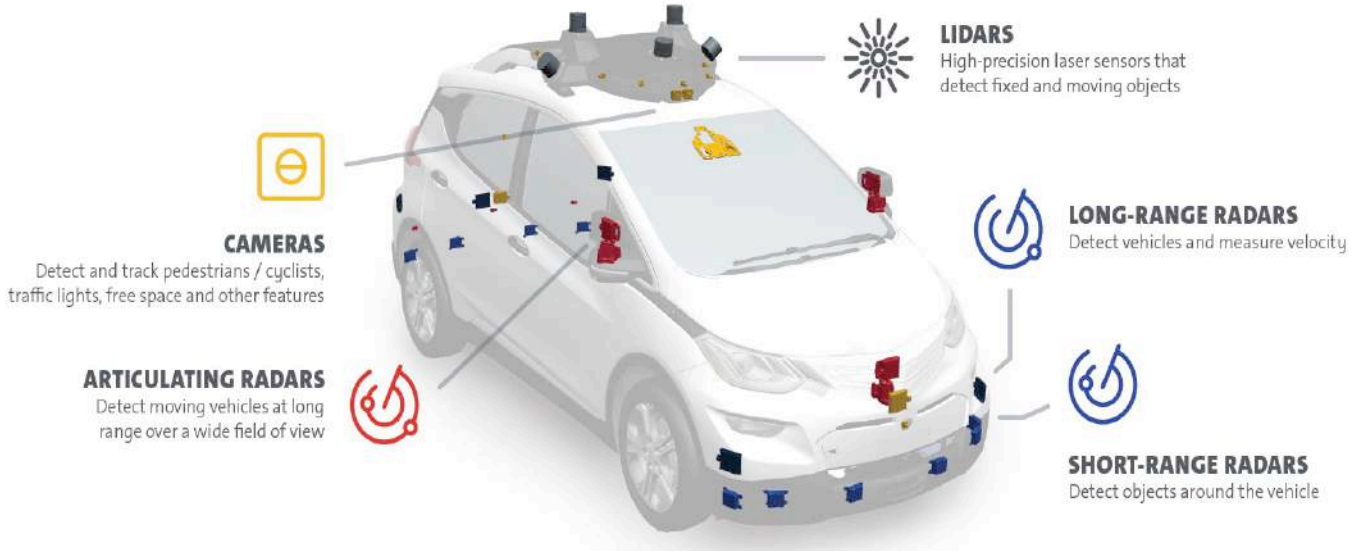
Decision Making in Self-Driving Cars

Decision making process involves sensors, external sources of information, networks, hardware, software, etc.

Environmental influences, such as weather conditions (rain, bright sun, storm, …)

Complex input must be filtered and only represents an abstraction of the real world.

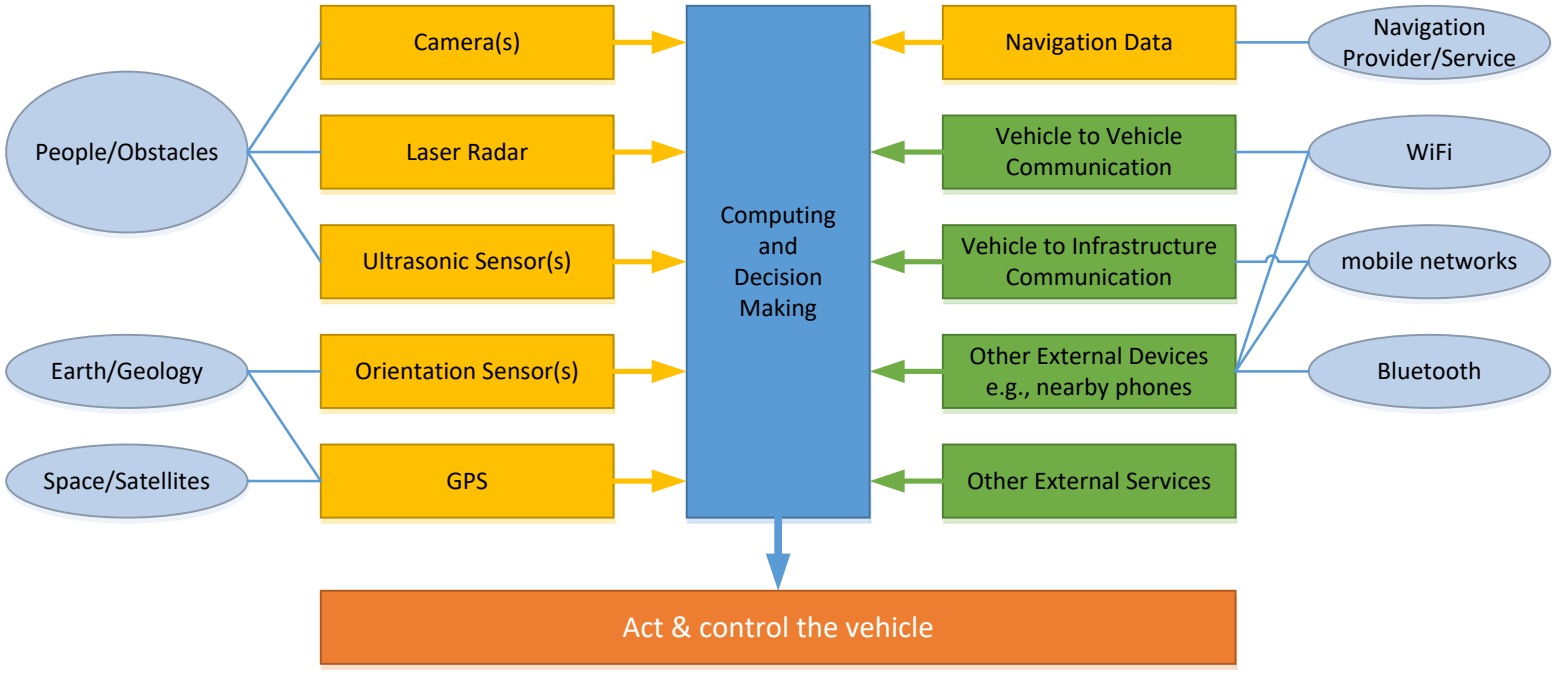
Technical Components



Picture Source: General Motors Safety Report 2018

What does a Self-Driving Car “see”…

Abstract Decision Making Process



This is an outline of what a decision making process might include.
Based on a literature review and official press releases (Tesla, Google, GM).

TECHNICAL CHALLENGES WITH ETHICAL CONSEQUENCES IN AUTONOMOUS CARS

Safety

How can we test self-driving cars?	and when is testing sufficient?
Real world vs Abstract world	Training of Neural Networks

Security

Attacks against car systems and sensors

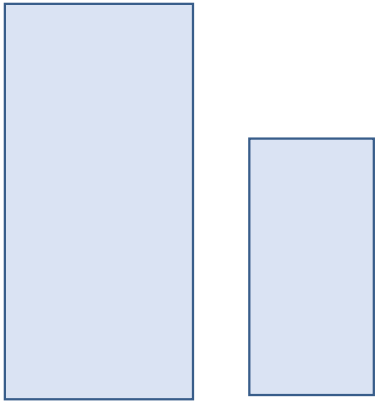
System & security updates

Do we need a “black box” in self-driving cars like in aircrafts?

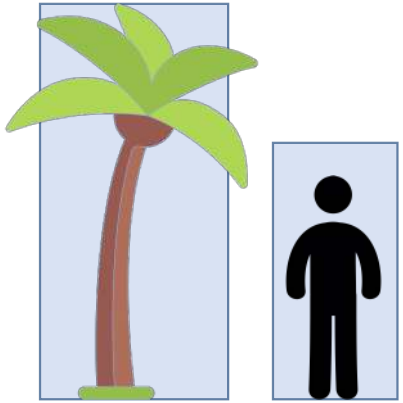
Privacy

- What data should the car have access to?
  - Who will have access to that data?
  - How will the data be used?
- What data is collected?

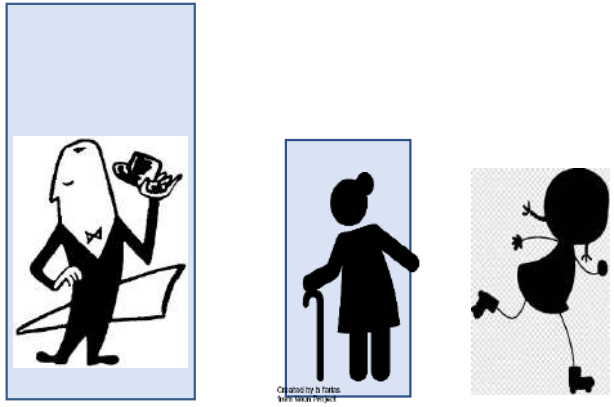
Privacy
What does the car “recognize” ?- Equivalent for Intelligent Robotics!



Objects, different size,
Position, moving or stationary



Objects vs Person(s)



„Everything“ including human identity – connected to data-bases

Trust

How trustworthy are data sources?

E.g., GPS, map data, external services

Trust between self-driving car and services

How trustworthy is the self-driving car?

E.g., Trust between user and car

Transparency

Multi-disciplinary challenge to ensure transparency, while respecting intellectual property rights, corporate secrets, security concerns, etc.

How much should be disclosed, and disclosed to whom?

Reliability

What do we have to rely on?

- What if sensor(s) fail?
- What if networks fail?

Redundancy for everything?

Responsibility and Accountability

- Who is responsible and for what?
- Who is accountable and for what?
- How is responsibility distributed among:
  - Developers
  - Car manufacturers
  - Safety inspectorates
  - Governmental institutions
  - Involved participants in the traffic
  - Other stakeholders

Quality Assurance Process

- Lifetime of components
- Maintenance
- Ethics-aware decision making in all processes will help to make ethically justified decisions.

SOCIAL CHALLENGES WITH ETHICAL CONSEQUENCES IN AUTONOMOUS CARS

Equivalent for Intelligent Social Robots will be Central!

Stakeholders Interests

- Loss of jobs (for cabs/taxi/truck/heavy industrial vehicles drivers)
- Humans in the loop
- Impact on Society

Stakeholders Interests

Freedom of movement

Will the car go, where I want it to go?

Implementation of restrictions

Route to Destination

Can the passenger define the route, or is it determined by the system?

Road trips?

Addressing Organisational Ethical Issues of AI

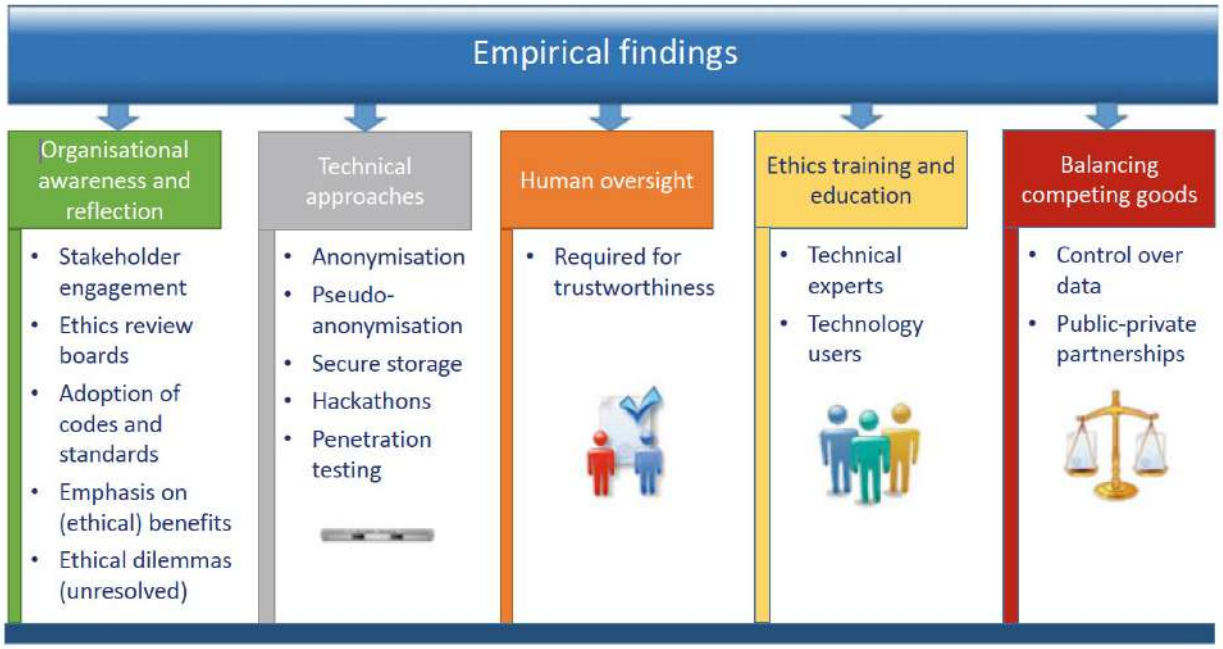
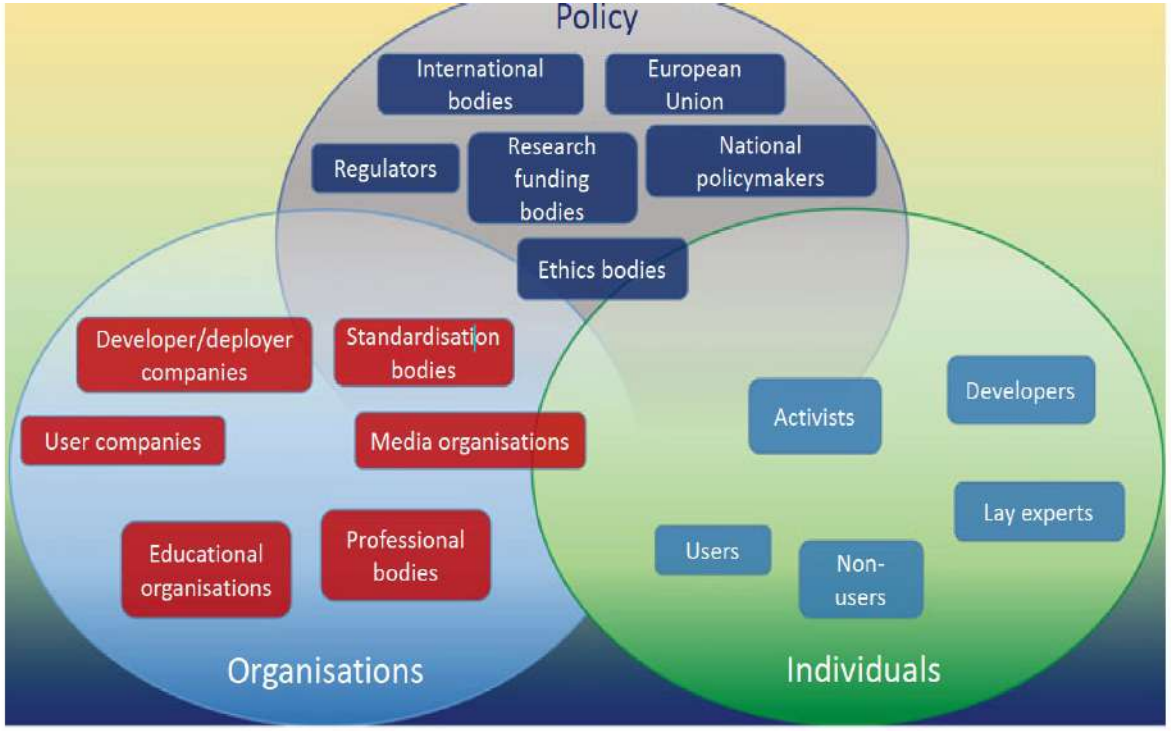


Fig. 5.2 How case study organisations address ethical issues of AI: empirical findings

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

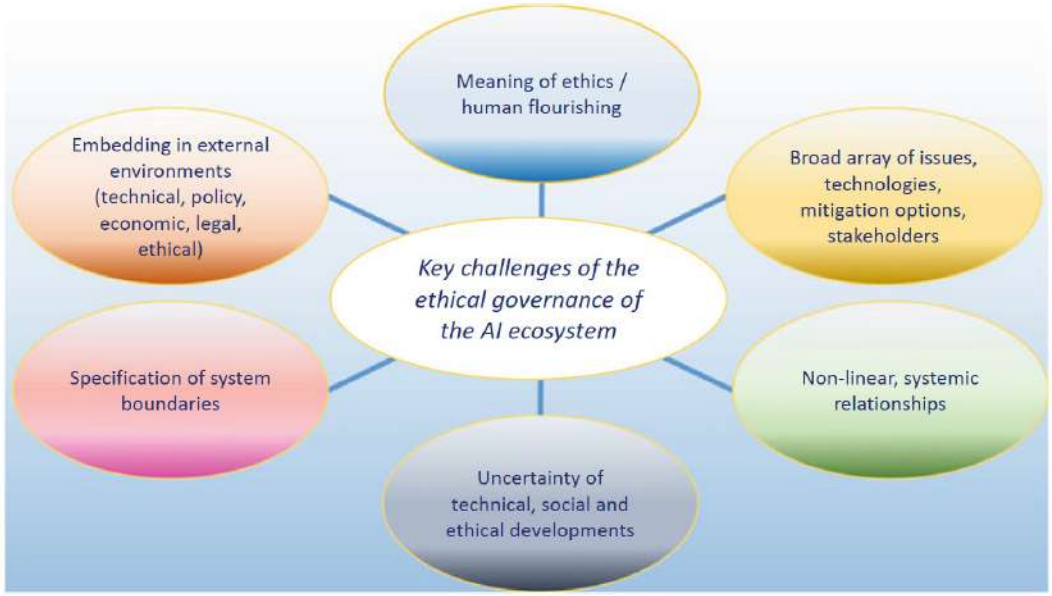


Fig. 7.1 Key challenges of ethical governance of AI ecosystems

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

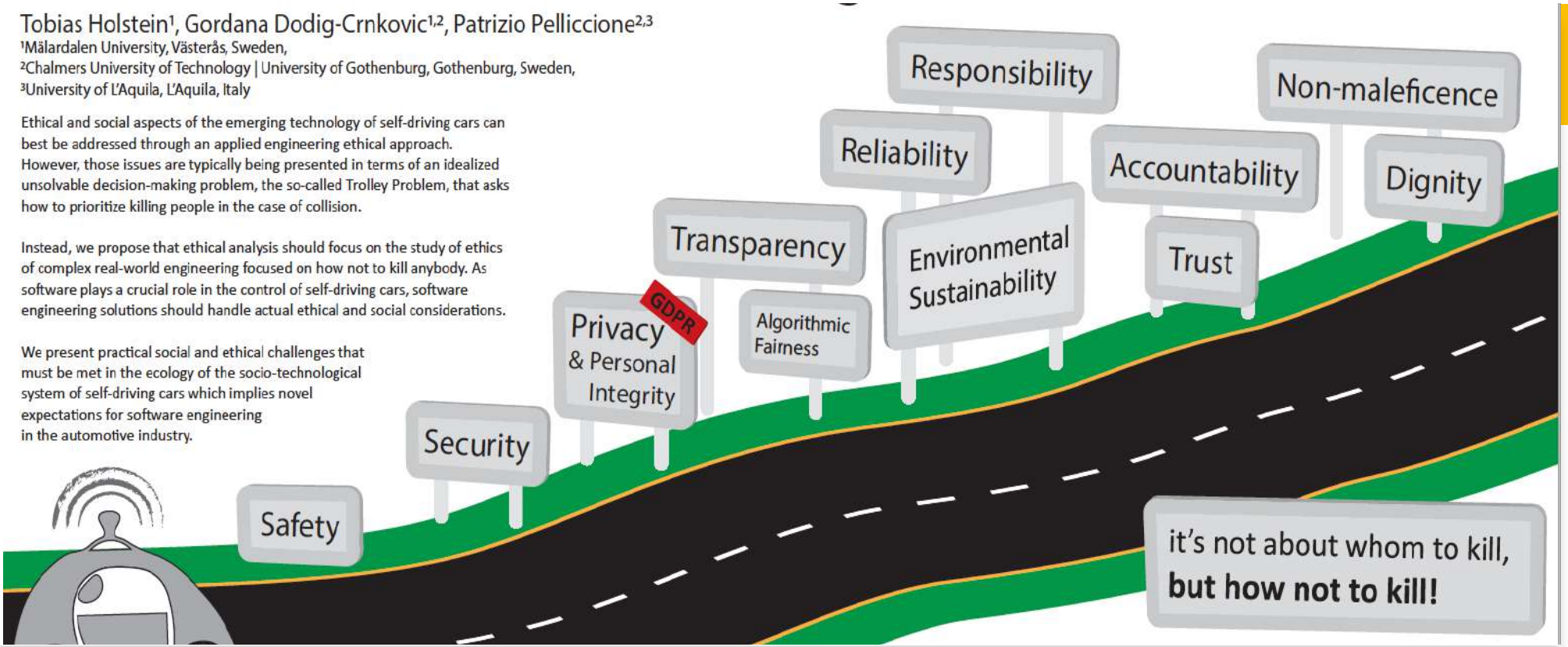
Ethical Issues of AI

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)

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

Requirements	Technical Challenges	Approaches
Safety	Hardware and software adequacy. Vulnerabilities of machine-learning algorithms. Trade-offs between safety and other factors (like economic). Possibility of intervention in self-driving cars (including for the Police). Systemic solutions to guarantee safety in organizations (regulations, authorities, safety culture).	Setting safety as the first priority. Learning from the history of automation. Learning from driving experience - perception and input interpretation processes. Specification of how a self-driving car will behave in cases when the car is not able to operate autonomously. Clarification of the role of the police. Regulations, guidelines, standards being developed as the technology develops.
Security	Minimal necessary security requirements for deployment of self-driving cars. Security in systems and connections. Deployment of software updates. Storing and using received and generated data in a secure way.	Technical solutions that will guarantee minimum security under all foreseeable circumstances. Anticipation and prevention of the worst-case scenarios regarding security breaches. Provide active security. Accessibility of all data, even in the case of accidents, has to be provided, so that it can be analysed to foster knowledge and to provide facts for next generation developments.
Privacy	Trade-offs between privacy and data collection/recording and storage/sharing.	Following/applying legal frameworks to protect personal data, such as GDPR.
Transparency	Information disclosure, what and to whom. Transparency of algorithmic decision making. Transparency in the techno-social ecosystem.	Assurance of transparency and insight into decision making. Active sharing of knowledge to ensure the interoperability of systems and services.
Algorithmic Fairness	Algorithmic decision making is required to be fair and not to discriminate on the grounds of race, gender, age, wealth, social status etc.	This requirement is related to transparency of decision making and expectation of explainability of the ground for decision making.
Reliability	Reliability of sensors and software and need for redundancy. Reliability of required networks and solution for the case when the network is unavailable.	Definition of different levels for reliability, such as diagnostics, vehicle input sensors, software, and external services, set the ground for reliability measures of the car as a system and its components. Standardized process required to shift from fail-safe to fail-operational architecture.
Environmental Sustainability	Environmental sustainability ethics refers to new ways of production, use, and recycling for autonomous vehicles.	Production, use, and disposal/recycling of technology rises sustainability issues (batteries, car sharing) that must be addressed.
Intelligent behavior control	Intelligent behaviour may lead to unpredictable situations resulting from learning and autonomous decision making.	Development of self-explaining capability and other features ensuring desired behavior in intelligent software.
Transdisciplinarity -Systemic approach	Ethics in design, requirements engineering, software-hardware development, learning, legal and social aspects, software-hardware interplay.	Adoption of transdisciplinarity and system approaches is increasing and should be given even more prominent role.
Quality	Quality of components. Quality of decision making. Lifetime and maintenance. QA process. Adherence to ethical principles/guidelines	Ethical deliberations included in the whole process starting with design and development. Ethics-aware decision making to ensure ethically justified decisions.

Table 1 Summary of the technical challenges and approaches, grouped by requirement

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

Requirements	Social Challenges	Approaches
Non-maleficence	Technology not causing harm. Disruptive changes on the labor market. Change of related markets and business models (e.g., insurances, manufacturers).	Partly covered by technical solutions. Preparation of strategic solutions for people losing jobs. Learning from historic parallels to industrialization and automatization.
Stakeholders involvement	In this field different stakeholders are involved – from professionals designing, developing, maintaining cars, to their users, and general public.	Active involvement of stakeholders in the process of design and requirements specification as well as decisions of their use.
Beneficence	Values and priorities: Ensure that general public values will be embodied in the technology, with interests of minorities taken into account.	Initiatives as "AI for good" exemplify this expectation that new technology not only do not cause harm, but actively do good for its stakeholders.
Responsibility and Accountability	Assignment and distribution of responsibility and accountability are among central regulative mechanisms for the development of new technology. They should follow ethical principles.	The Accountability, Responsibility and Transparency (ART) principle based on a Design for Values approach includes human values and ethical principles in the design processes (Dignum, 2019).
Freedom and Autonomy	Freedom of choice hindered by the system (e.g. it may not allow to drive into a certain area)	The freedom of choice determined by regulations. Determination and communication of the amount of control a human has in context of the self-driving car
Social Sustainability	In the domain of business, social sustainability is about identifying and managing business impacts on people	Pursuing social equity, community development, social support, human rights, labour rights, social responsibility, social justice, etc.
Social Fairness	Ascertaining fairness of the socio-technological system.	Fairness of the decision-making. Related to transparency and explainability.
Dignity and Solidarity	This requirement refers to the entire socio-technological system.	Challenges come from the lack of common wholistic view.
Social Trust	Establishing trust between humans and highly automated vehicles as well as within the social system.	Further research on how to implement trust across multiple systems. Provision of trusted connections between components as well as external services
Justice: legislation, standards, norms, policies and guidelines	Keeping legislation up to date with current level of automated driving, and emergence of self-driving cars. Creating and defining global legislation frameworks. Including ethical guidelines in design and development processes	Legislative support and contribution to global frameworks. Ethics training for involved engineers. Establishment and maintenance of a functioning socio-technological system in addition to functional safety standards

Table 2 Summary of the social challenges and approaches, grouped by requirement

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

New for Intelligent Robots – Cognitive and Psychological Effects of Social/Companion Robots

Cognitive and psychological effects of social/ companion robots on humans	Personal integrity Cognitive load Deception	Further research on how social robots and especially increasingly intelligent and human-like robot companions affect users. Solid understanding of effects, after stakeholders interests are taken into consideration should be followed by regulation/legislation. Humanoid or zoomorphic robots may cause emotional attachment to some users. "Robots should not be designed in a deceptive way to exploit vulnerable users" (Boden et al. 2017)

Boden, Margaret, Joanna Bryson, Darwin Caldwell, Kerstin Dautenhahn, Lilian Edwards, Sarah Kember, Paul Newman, et al. 2017. "Principles of Robotics: Regulating Robots in the Real World." Connection Science. https://doi.org/10.1080/09540091.2016.1271400.

Ehics Of Self-driving Cars Paper* Method

1. A list of ethical values (requirements) for autonomous cars was compiled from an extensive literature study and connected to challenges together with approaches for their fulfilment.
2. The proposed analysis with values, challenges and approaches for technical and social aspects was presented to different stakeholders and discussed via seminars.
3. Finally we number of experts was invited to review the results.

* 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

Practical Use of the Proposed Ethical Framework

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

The First Ethical Guidelines For Automated Driving Equivalent for Robotics Needed!

German Ethics Commission's report comprises 20 propositions. The key elements are:

Automated and connected driving is an ethical imperative if the systems cause fewer accidents than human drivers (positive balance of risk).

The protection of individuals takes precedence over all other utilitarian considerations.

In hazardous situations, the protection of human life must always have top priority.

https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

The First Ethical Guidelines For Automated Driving

Equivalent for Robotics Needed!

In the event of unavoidable accident situations, any distinction between individuals based on personal features (age, gender, physical or mental constitution) is impermissible.

In every driving situation, it must be clearly regulated and apparent who is responsible for the driving task: the human or the computer.

It must be documented and stored who is driving (to resolve possible issues of liability, among other things).

Drivers must always be able to decide themselves whether their vehicle data are to be forwarded and used (data sovereignty)

https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

“Learning By Experience” And “Proven In Use” Concepts Equivalent for Robotics Needed!

“Learning by experience” (recording data from autonomous cars) presupposes a functioning socio-technological system that provides strong coupling among legislation, guidelines, standards and use, and promptly adapts to lessons learned.

H. Schäbe and J. Braband. Basic requirements for proven-in-use arguments. CoRR, abs/1511.01839, 2015.

Challenges

Legislation	Global framework	Guidelines	Including Ethics into all phases in the life-cycle
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, development and other processes in the life-cycle of a product.

Holstein, Dodig-Crnkovic, Pellizzione: Ethical and Social Aspects of Self-Driving Cars, ArXives

Recommendations - Applied for Robotics

Producers supporting and collaborating with legislators in their task to keep up-to-date with the current level of technology

Legislative support and contribution to global frameworks to ensure a smooth enrollment of the emerging technology

Include ethics in the overall process of design, development and implementation of technology. Ensure Ethics training for involved engineers

Establish and maintain a functioning socio-technological system in addition to functional safety standards.

Holstein, Dodig-Crnkovic, Pellizzione: Ethical and Social Aspects of Self-Driving Cars, ArXives

Conclusions

It is time to stop discussing unsolvable ethical dilemmas that obfuscate much bigger actual ethical challenges.

Discuss the real-world ethical challenges surrounding emerging technology.

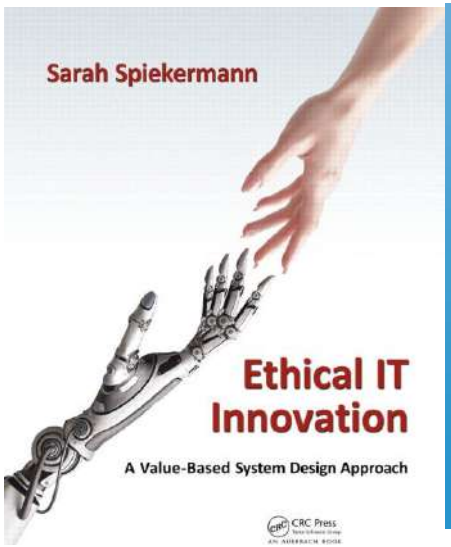
Define what is technically possible and ethically justifiable.

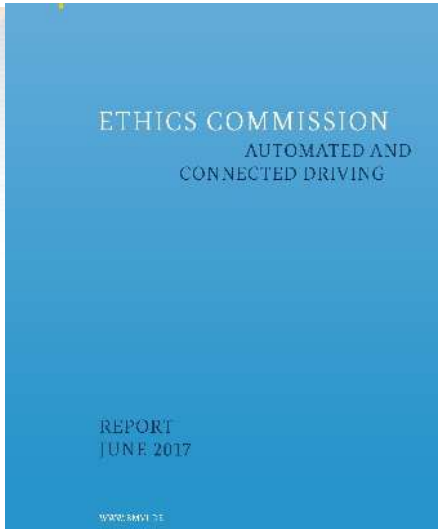
Create transparency to support evaluations by independent organisations/experts.

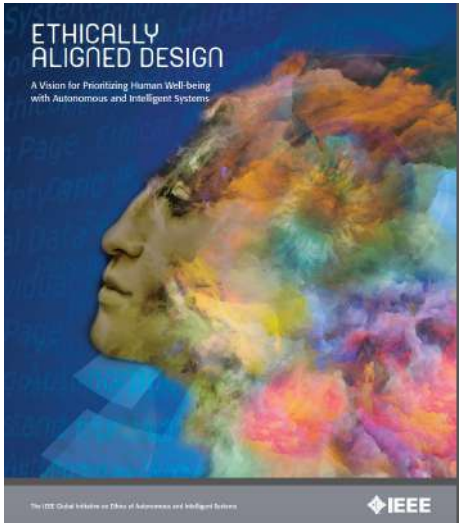
Ethicality/Ethicity as non-functional property? (Ethicality: the state, quality, or manner of being ethical.)

There is already a body of normative documents that can support ethicality of design and implementation.

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/

Soma Design – Intertwining Aesthetics, Ethics and Movement - Kristina Höök

Values explored from within somaesthetics

Connecting socially grounded, negotiated understanding of values with the individual perspective from the point of view of soma – individual body. This connection is important. How do values feel on the individual somatic level?

Soma design — examines and improves connections between sensation, feeling, emotion, subjective understanding and **values**.

Soma design builds on **somaesthetics** by Shusterman.

It combines soma as in our first-person experience of the world through our senses, with ***aesthetics*** *as deepening knowledge of our sensory experiences to* ***live a better life***.

Our cultural practices and digitally-enabled objects enforce a form of sedimented, agreed-upon movements, enabling variation, but with *certain prescribed ways to act, feel and think*.

Höök argues that by engaging in a soma design process we can better probe which movements lead to deepened **somatic awareness**; **social awareness of others** in the environment and how they are affected by the human-technology assemblage;

enactments of bodily freedoms rather than limitations; making norms explicit;

engaging with a *pluralist feminist position* on who we are designing for; and

aesthetic experience and expression.

https://dl.acm.org/doi/10.1145/3313831.3376678 Ethics in Movement: Shaping and Being Shaped in Human-Drone Interaction
SOMA - the body as distinct from the soul, mind, or psyche.

FURTHER READING

Legislation and Standards

Legislation implemented with rigorous monitoring the behavior of technology.

Implementation is within the responsibility of producers. That means that design and implementation of software should follow ethical guidelines.

Ethics & Law Aspects Equivalent for Robotics Needed!

Ryan Jenkins (2016) Autonomous Vehicles Ethics & Law: Towards an Overlapping Consensus

https://www.academia.edu/29332066/Autonomous_Vehicles_Ethics_and_Law_Towards_an_Overlapping_Consensus

Patrick Lin (2015) Why Ethics Matters for Autonomous Cars.

In: Autonomes Fahren Technische, rechtlische und gesellschaftliche Aspekte

https://www.springerprofessional.de/en/why-ethics-matters-for-autonomous-cars/4397684

Ethically Aligned Design

A Vision for Prioritizing Human Well-being With Autonomous and Intelligent Systems

https://ethicsinaction.ieee.org/

Embedding Values into Autonomous Intelligent Systems - The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

https://standards.ieee.org/develop/indconn/ec/ead_embedding_values.pdf

An example of ethical guidelines thinking one step further is described in the book:

Sarah Spiekermann. Ethical IT Innovation: A Value-Based System Design Approach. Taylor & Francis, 2015.

Policy Concerning Automated Vehicles (US DOT) Equivalent for Robotics Needed!

"DOT/NHTSA Policy statement concerning Automated Vehicles" 2016 update to "Preliminary statement of policy concerning automated vehicles".

Technical report, National Highway Traffic Safety Administration (NHTSA).

http://www.nhtsa. gov/staticfiles/rulemaking/pdf/Autonomous-Vehicles-Policy-Update-2016.pdf

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.