

Ethics of Robotics

VALUE SEMINAR

THWS - INSTITUTE OF ROBOTICS

LECTURESHIP: M. BERTHOLD

Exam regulations

- ▶ Regular attendance and **debates** on the respective topics are essential for an apt preparation
- ▶ **Written exam (90min) [Mode: Pass/Fail]**
- ▶ **Question pool with study tasks (40-50 questions)**
 - four questions from the pool
 - two questions with transfer and application requests concerning contents only dealt with in the seminar
 - additional literature can help you for a more profound preparation

E-learning

IRO7 : Value Seminar - Ethics of Robotics

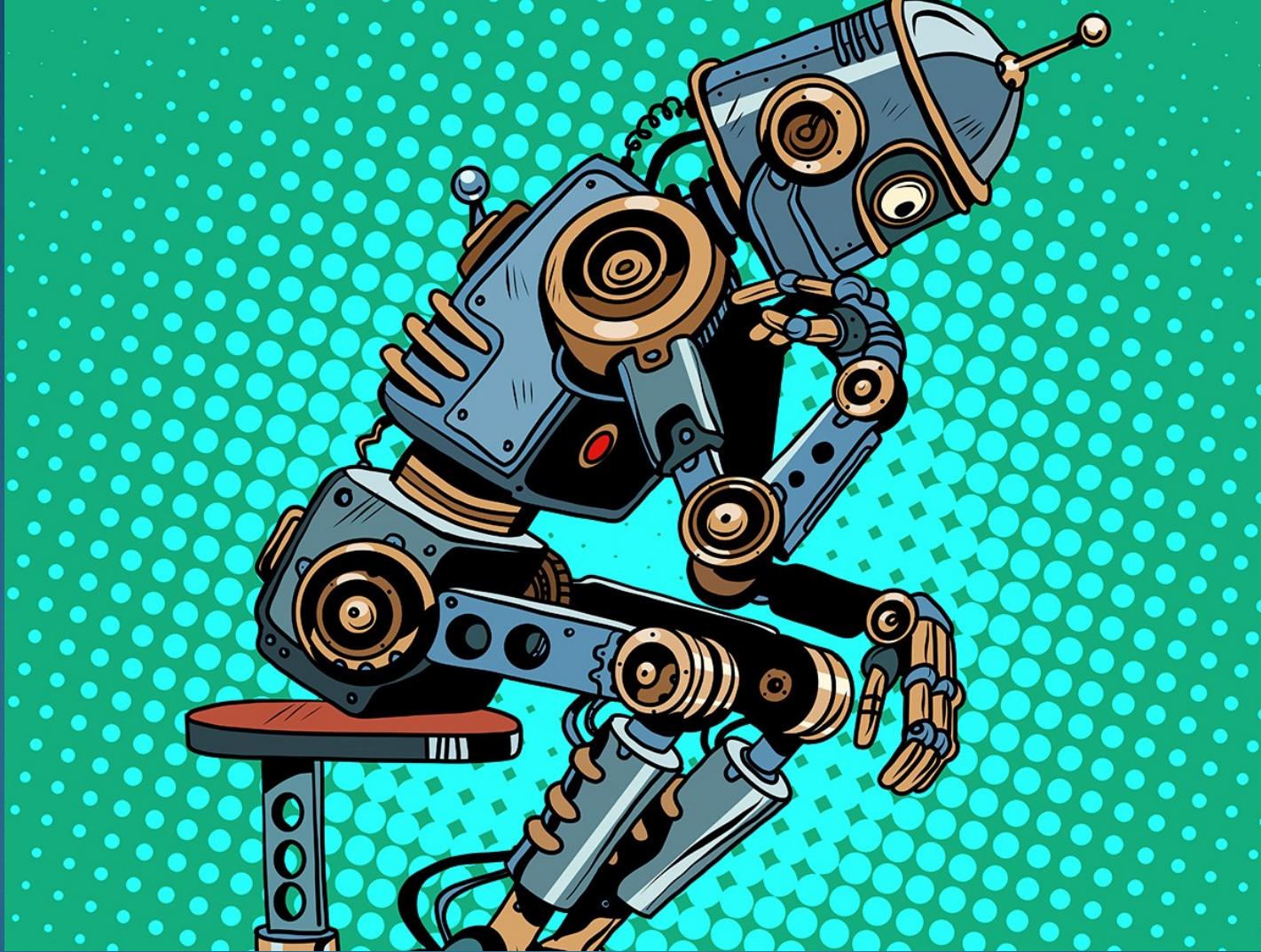
- ▶ **Enrolement key:** **Ethics23**
- Powerpoint
- Handouts
- Exam question pool

Further literature and reading recommendations

- ▶ Coeckelbergh, Mark. Robot ethics. Massachusetts: The MIT Press. 2022
- ▶ Lin, Patrick et. Al. Robot ethics: The ethical and social implications of robotics. Massachusetts: The MIT Press. 2014
- ▶ Funk, Michael. Roboter- und KI-Ethik. Eine methodische Einführung – Grundlagen der Technikethik Band 1. Wiesbaden: Springer. 2022

Contents and methodical approach

- ▶ (Ethical) Laws of robotics and their application
- ▶ Moral contexts of justification
- ▶ Moral philosophy and programming
- ▶ Particular application fields and ethical chances and concerns
- ▶ Sociopolitical developments in relation to increased usage of Strong AI and autonomous humanoid robots
- ▶ Dystopian and utopian prospects
- ▶ Men-machine interaction and its connection to general philosophy
 - > communication, conceptions of men and behavioral studies, responsibility/accountability, happiness-theories, aesthetics



<https://robohub.org/a-mathematical-approach-to-robot-ethics/>

Essentials of Robot-Ethics

- ▶ Various definitions for an interdisciplinary subject!

Robot Ethics refers to the study of ethical issues concerning the design, creation, and use of robots and artificial intelligence systems. It explores questions related to the moral responsibilities of humans in developing robots, the ethical behavior of robots, and their impact on society.

- > Threefold approach

Essentials of Robot-Ethics

Machine Ethics is a branch of artificial intelligence and computer science that explores the ethical implications of autonomous systems and machines. It focuses on developing principles, guidelines, and frameworks that govern the behavior and decision-making processes of intelligent machines, robots, and other autonomous systems. The goal of machine ethics is to ensure that these systems operate in a manner that is morally acceptable and aligns with human values.

> Anthropocentric access

Essentials of Robot-Ethics

- ▶ **Responsibility:** Determining who is accountable for the actions of autonomous machines, especially in the case of errors or harmful outcomes.
- ▶ **Bias and Fairness:** Addressing the biases that can be embedded in algorithms and ensuring that machine decisions are fair and unbiased, particularly in applications like hiring, lending, and law enforcement.
- ▶ **Privacy:** Safeguarding individuals' privacy and ensuring that autonomous systems handle sensitive information responsibly.
- ▶ **Transparency:** Making machine decisions more transparent and understandable to users and stakeholders, so the rationale behind these decisions is clear.

Essentials of Robot-Ethics

- ▶ **Security:** Ensuring that autonomous systems are secure and protected against hacking or misuse that could lead to harmful consequences.
- ▶ **Human-Computer Interaction:** Studying the ethical implications of how humans interact with machines and ensuring respectful and ethical user experiences.
- ▶ **Societal Impact:** Examining the broader impact of autonomous systems on society, including their influence on employment, social interactions, and power dynamics.

Crucial ethical process: Value alignment

Essentials of Robot-Ethics and Robot Laws

Key questions/problems:

- ▶ The *AI-control problem, The Superintelligence-Problem*
- ▶ Can we design genuine Artificial Moral Agents (AMAs) with responsibility, moral decision-making, autonomous moral evaluation?
- ▶ Are there universal guidelines or laws that enable a vindicated jurisdiction?

Laws of Robotics

- ▶ Basic terminology and concepts
 - Artificial Moral Agents (AMAs)
 - Laws of Robotics (by Isaac Asimov)
 - Zeroth Law of Robotics
 - Friendly AI

Concepts of moral philosophy and their application in the fields of AI and Robotics

Moral philosophy

“How should we live? Shall we aim at happiness or at knowledge, virtue, or the creation of beautiful objects? If we choose happiness, will it be our own or the happiness of all? And what of the more particular questions that face us: is it right to be dishonest in a good cause? Can we justify living in opulence while elsewhere in the world people are starving? Is going to war justified in cases where it is likely that innocent people will be killed? Is it wrong to clone a human being or to destroy human embryos in medical research? What are our obligations, if any, to the generations of humans who will come after us and to the nonhuman animals with whom we share the planet?”

Peter Singer, contemplations on moral philosophy

Moral philosophy

- ▶ Strives to define and determine what is right or wrong, good or evil just or unjust, equal or unequal and desirable or detestable
- ▶ **Three major branches and research fields** (*Encyclopedia of Philosophy*)
 - Meta-ethics
 - Descriptive Ethics
 - **Normative Ethics**
 - Applied Ethics

Normative ethics

- ▶ Frequently intertwines with particular questions from Applied Ethics

Sub-branches within normative ethics

Deontology/Principle Ethics

Utilitarianism

Virtue Ethics

Beware of *naturalistic fallacies*

Moral Codes

- ▶ A *moral code* refers to a set of principles or guidelines that individuals or a society adhere to, which define what is considered morally right or wrong.
- ▶ It serves as a framework for making ethical decisions and guiding behavior in various situations.

Moral codes are often influenced by cultural, religious, philosophical, and societal factors, shaping the values and beliefs of individuals and communities.

Moral Codes

“Many people believe that there are no moral universals - i.e., that there is so much variation from one culture to another that no single moral principle or judgment is generally accepted. It has already been shown that this is not the case.”

Peter Singer, contemplations on moral philosophy

Moral Codes and ethical frameworks Robotics and AI

- When we discuss programming autonomous and decision-making machines, to which moral frameworks should we resort to?

Deontology

- ▶ Ethics of personal duty and individual obligations (and motives)
- ▶ Focuses on universal, indisputable and inalienable principles
- ▶ Actions are in itself good or evil
- ▶ Consequences and circumstances are not important

Proponents

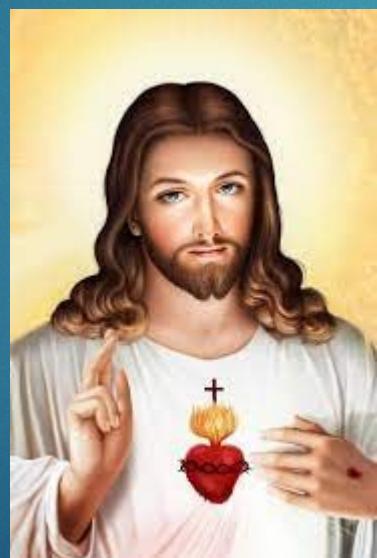
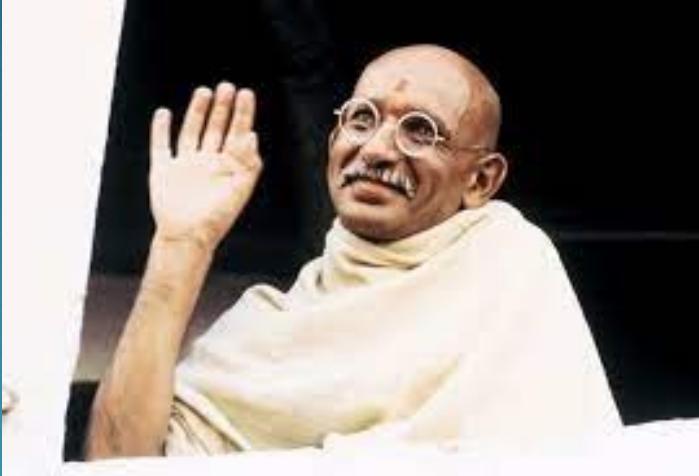
- ▶ Immanuel Kant; maxims and categorial imperative
- ▶ Religious commandments or divine command theories
- ▶ The Golden Rule
- ▶ Mother Theresa, Mahatma Ghandi

Categorical Imperative by Kant

“Act only according to that maxim whereby you can at the same time will that it should become a universal law.”

Immanuel Kant, Groundwork of the Metaphysics of Morals, 1785

Maxims are verbalized obligations of our rationality



Utilitarianism

- ▶ Focuses on the well-being, happiness and contentment of all
- ▶ The greater good is crucial
- ▶ Circumstances of an action matter
- ▶ Sub-category of consequentialism

Proponents

Jeremy Bentham, John Stuart Mill, Henry Sidgwick, Peter Singer

consequences and results of an action are the predominant feature for decision-making

Act-utilitarianism vs. Rule-utilitarianism

► **Jeremy Bentham**

Highly consequentialist (“Robin-Hood”-Utilitarianism)

Greater good through pure actions

Example: Claus Schenk Graf von Stauffenberg

► **John Stuart Mill**

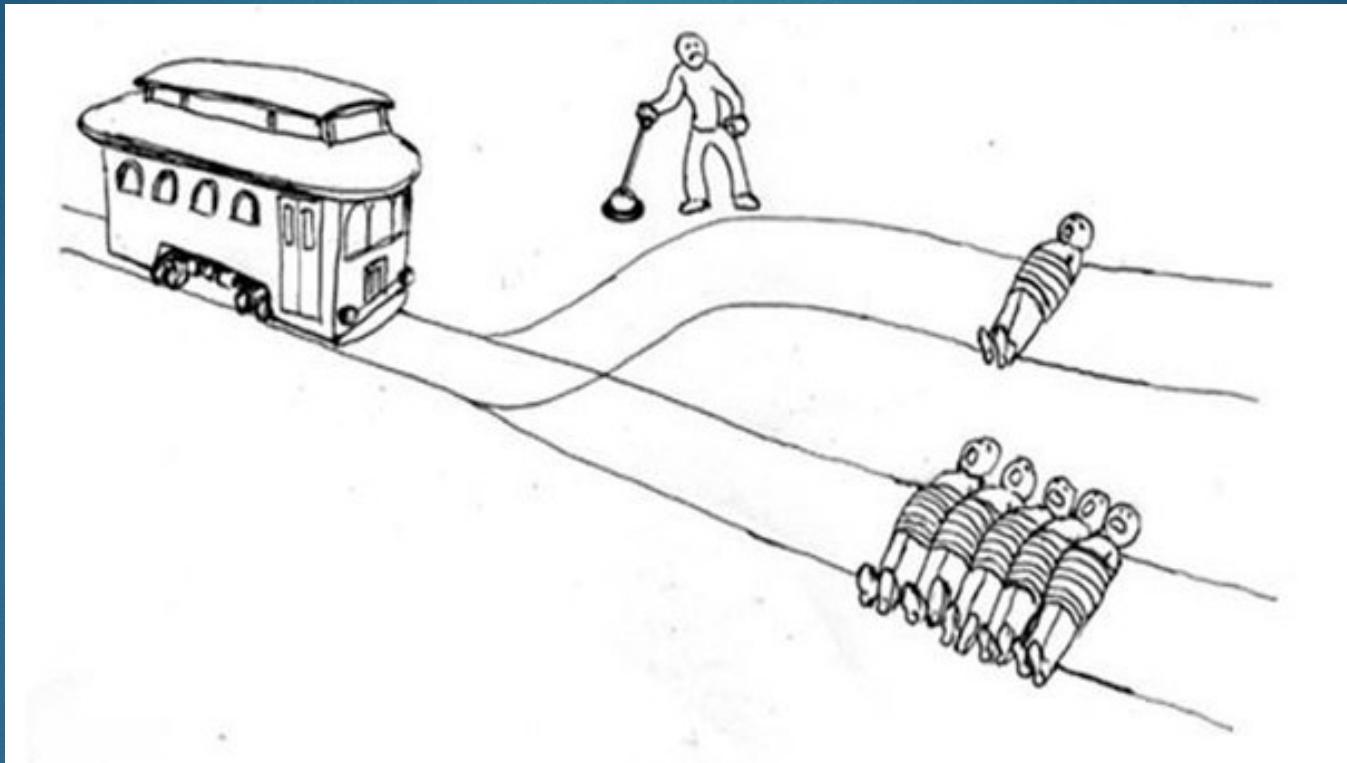
Relatively consequentialist

Greater good in accordance to established moral rules

Example: Lawn-mowing neighbor child

Further consequentialist reasoning:

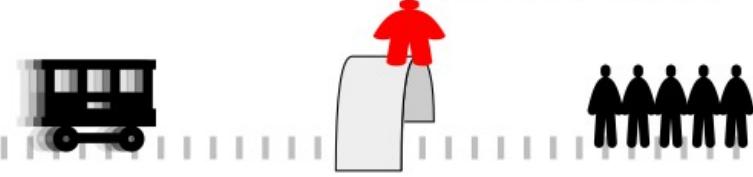
- 1) The end justifies the means
- 2) The *Principle of the double-effect (PDE)* or *double-effect reasoning*



1. The switch
Foot, 1967



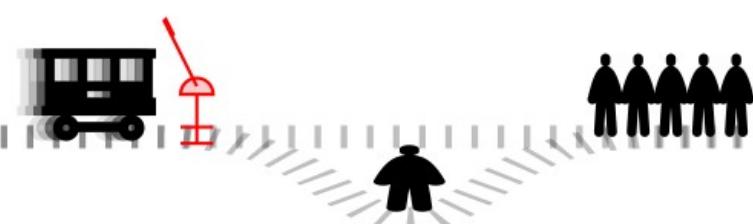
2. The fat man
Thomson, 1976



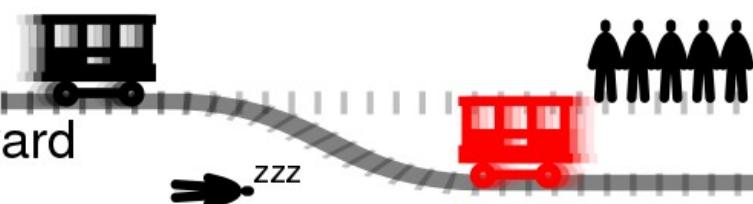
3. The fat villain



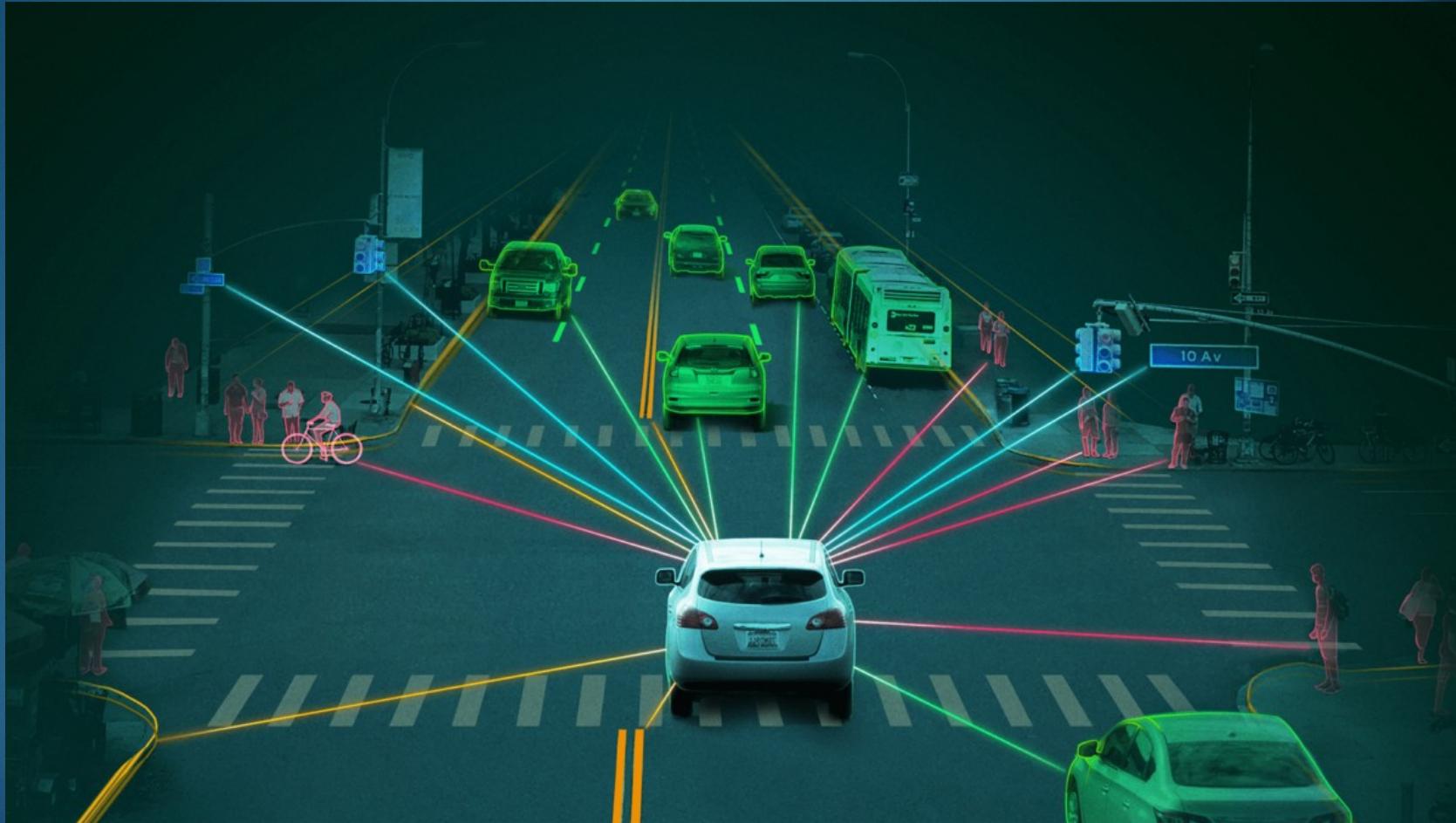
4. The loop
Costa, 1987



5. The man in the yard
Unger, 1992



Autonomous driving



Basic challenges for self-driving and autonomous vehicles

- ▶ pedestrian detection
 - ▶ lane departure warnings
 - ▶ traffic sign recognition
 - ▶ blind-spot detection
-
- ▶ Self-governing ethical evaluation
 - ▶ Judgemental decision-making

**sphere of
perception**

**sphere of
cognition**

Consumer Concerns About Self-Driving Cars

% of respondents naming the following reasons for their reluctance to use self-driving cars



Utilitarian paradox

Passenger safety could not be the top priority of a self-driving, fully automated, vehicle.

=> SAE classification therefore implies particular ethical assessments

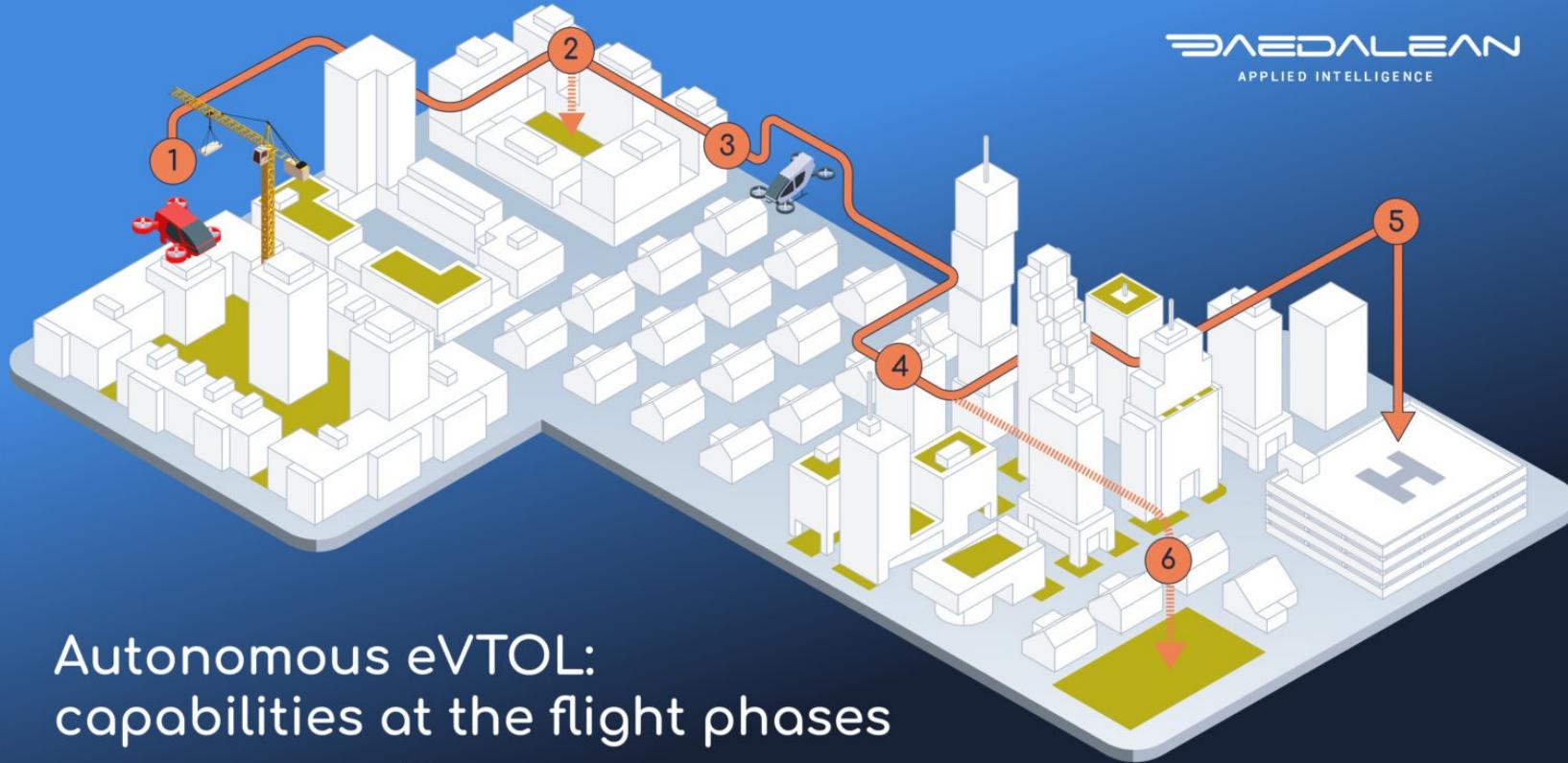
Autonomous planes

“Currently, software developers and aeronautical engineers have gotten very good at programming machines to pilot aircraft so they can safely and reliably complete repetitive tasks within predetermined parameters.

But as the technology stands now, these machines still can’t be creative. They can’t come up with instant, creative solutions to unanticipated problems that may suddenly threaten aircraft and passengers.”

Thom Patterson, January 2022

<https://www.flyingmag.com/how-will-self-flying-aircraft-make-ethical-choices/>



Autonomous eVTOL: capabilities at the flight phases



- 1 Detect and avoid hazards



- 2 Recognize where you can land in case of emergency



- 3 Detect and avoid air hazards (even if non-cooperative)
- 4 Know where you are and navigate (even if GNSS denied)



- 5 Conduct precision landing on pre-marked spot



- 6 Land safety on unmarked spot

Takeoff

Cruise

Landing

Ethical dilemma – deontology vs. utilitarianism

Should an AI-controlled plane, hijacked by terrorists who threaten to fly it directly into a packed stadium, “sacrifice” itself and its passengers for the greater good/lesser evil?

Thought experiment

A car realizes that it is being used as a transport vehicle in a criminal act (human trafficking, grand theft, drug smuggling etc.). Should it act as a moral agent in itself and autonomously shut down its services or inform the authorities?

VLIII – Virtue Ethics in Robotics

Is the field of Robot-Ethics compatible with Virtue Ethics?

Virtue Ethics

"Virtue ethics is a moral philosophy that focuses on the character of individuals and emphasizes the importance of developing virtuous traits. Unlike consequentialist or deontological ethical theories, which concentrate on actions or rules, virtue ethics is concerned with the inherent qualities of a person and how those qualities influence their behavior and decisions."

(Nancy Snow, Contemporary Virtue Ethics, 2022)

Necessity of virtuous actions

As all technological revolutions have disruptive potential for individuals and societies, so a virtuous code is indisputable for a successful coexistence between men and high-profile robots, especially when it comes to **artificial personhood**.

(comp. The Ethical and Social Implications of Robotics, Lin & Abney, 2014, p. 47ff.)

Virtues and vices

- ▶ Central to virtue ethics is the concept of **virtues**, which are positive traits or qualities, deemed to be morally good.
 - Inherently good/right/ethical (contrary - vices and flaws)
 - Desirable and admirable
 - Indispensable for a good life
- ▶ Virtue ethicists argue that individuals should strive to cultivate these virtues within themselves and make decisions based on these virtuous qualities.

Cycle of virtuous living (clockwise)



Main aims of virtue ethics

Phronesis	Eudamonia
Practical wisdom, prudence	Happiness, well-being, deep contentedness, delight, bliss
Not to be mistaken for <i>Pragmatics</i>	Not to be mistaken for <i>Hedonism</i>

Phronesis

Refers to the ability to make sound judgments and decisions in practical matters, considering both the short-term and long-term consequences of one's actions.

Phronesis involves practical reasoning, prudence, and the ability to apply moral and ethical principles to real-life situations. In philosophical terms, phronesis is often contrasted with theoretical wisdom (*sophia*) and practical skills (*techne*)

Eudamonia

- ▶ **Aristotle, *Nicomachean Ethics*, approx. 330-340 BC**
 - ⇒ refers to the highest human good and the ultimate goal of human life
 - ⇒ not just a fleeting emotional state, but a state of living well and flourishing as a human being
 - ⇒ not attained through pleasure, wealth, or honor alone but through the pursuit of virtuous activities and the development of one's moral and intellectual capacities. It is a life of meaningful and purposeful activity, where individuals fulfil their potential and lead a life of excellence

Aristotle - The Doctrine of Mesotes

According to Aristotle, there are numerous virtues and human beings should dedicate themselves to achieving them

Courage, temperance, liberality, magnificence, magnamity, proper ambition, patience, truthfulness, wittiness, friendliness, modesty, righteous indignation

Important guideline hereby:

The Golden Mean or The Doctrine of Mesotes

(Also utterly present in Eastern Philosophy - Confucianism, Buddhism, Daoism etc.)

Virtues through the centuries

Plato: *The Republic/Politeia*, 375 BC

Cardinal virtues: Wisdom, fortitude, temperance, justice

Thomas Aquinas: *Summa theologica*, 1485

Cardinal virtues + hope, faith, charity

Bertrand Russel: *An outline of Philosophy*, 1927

Critical thinking, rational discipline, reason, honesty, generosity

Robot-ethics - discrepancy

**The intrinsic incline to act virtuously is -
not in the slightest - crucial for an AI or a
robot**

**Why should an autonomous robot then strive to make
virtuous decisions?**

The virtue-code of Robotics

- ▶ How would you compose a virtue-code for robots?
- ▶ What would be the five predominant virtues for you to include?
- ▶ Do you see a possible clash of virtues when it comes to programming a moral code based on virtue ethics?

The case for virtuous robots

Martin Gilbert, University of Montreal, 2023

Thesis

Robots should, as autonomous moral agents (AMAs) simulate and duplicate the behavior of an **imaginary wholesome virtuous person.**

“Not only are virtuous robots technically feasible, but they have the advantage over their deontological and utilitarian counterparts of fostering normative consensus between these moral schools, improving social acceptability, and beginning to address the technical challenge of moral perception.”

Thought experiment (year 2040)

Imagine you are a highly skilled, but retired firefighter. A neighborhood building has been set on fire, probably through arson, and due to the hard-to-reach surroundings of the house, the firefighters with their robots have not yet arrived on the site. The house is in critical condition. You know that an evil man lives in the house with his teenage son. You have also observed that a rare model of a medical robot from a nearby hospital, that usually conducts meticulous surgeries, has autonomously entered the building to rescue its inhabitants.

As you enter the atrium of the house, the teenage boy is running past you. He is bleeding, yet otherwise seems unharmed. The father and the robot are lying “unconsciously” a few meters away from you, for parts of the collapsing ceiling have knocked both of them down. You can only carry and rescue one of them.

Do you rescue the robot?

VL IV – Moral responsibility of Robots



Key questions:

- ▶ Do robots deserve moral praise?
- ▶ Is moral blame justified when robots act immorally and can they be made accountable?

The manifold implications of the term responsibility

Definition (The Cambridge Dictionary)

- 1) to be in a position of authority over someone and to have a duty to make certain that particular things are done
- 2) to take blame or praise for something that has happened in accordance to your actions
- 3) good judgement and the ability to act correctly and make decisions on your own

Top-down vs. Bottom-up approaches to moral machines

„Philosophers like to think in terms of abstractions. Engineers like to think in terms of buildable designs. Bridging these two cultures is not a trivial task.“

(Lin&Abney, 2014)

- ▶ **Top-down:** Framing rule- and duty-based conceptions
- ▶ **Bottom-up:** Emulating learning and developing evolutionary processes.

The question of personhood

The Turing Triage Test - Robert Sparrow

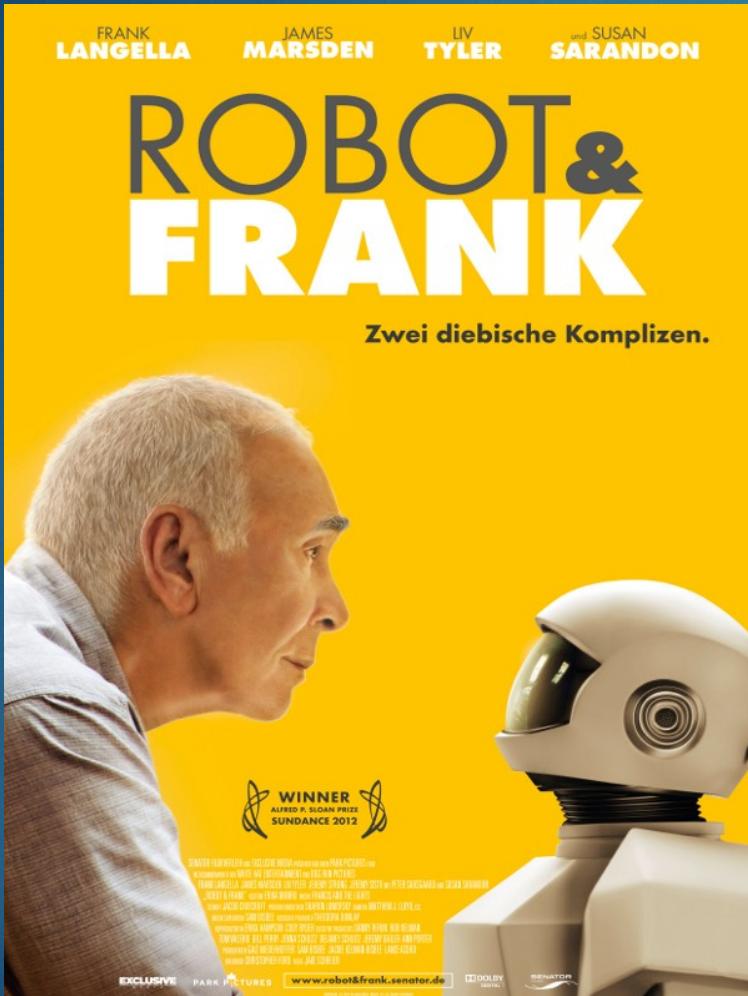
Can machines be (regarded as) people?

Personhood and accountability

Personhood and personal maturity are intertwined with accountability and responsibility on a diametral basis

- Psychology and pedagogics
- Law and legal ethics
- Neuroscience

Responsible Agents



Diffusion of responsibility

Subject of responsibility

(External or internal vindication)



Object of responsibility

(worthiness of protection and righteousness)

Freedom and volition

Free Will

Freedom of will

Freedom of choice

Freedom of action

Determinism

Psychology

Philosophy

Sociology

Physics

Biology and Neuroscience

Compatibilism

Soft determinism (or compatibilism) is the position or view that causal determinism is true, but we still act as free, morally responsible agents

=> in the absence of external constraints, our actions are caused by our desires.

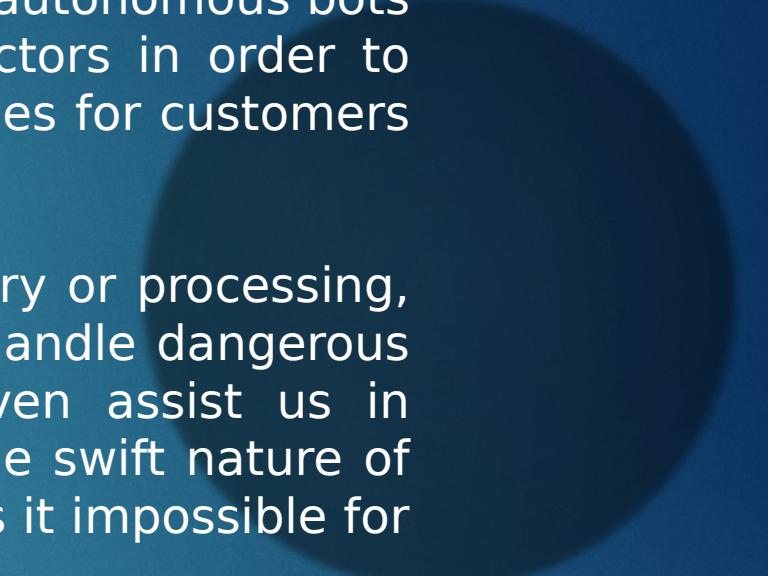
Scientific status quo – interdisciplinary consensus

- ▶ Freedom of will is non-existent
- ▶ Freedom of choice within a framework of action is
- ▶ Human beings are predetermined

Transferability to robots as AMAs?

Increase in demand and invests for autonomous robots (c.f. AMR)

- ▶ <https://www.statista.com/statistics/1285835/worldwide-autonomous-robots-market-size/>



The use of AI-robots in the **workplace** and to provide both services and products is becoming more and more commonplace. Today, autonomous bots are deployed to perform tasks across a whole host of sectors in order to make practices smoother and more efficient, improve services for customers and even keep people safe.

Far from day-to-day routine tasks in sales, education delivery or processing, today's AI bots can effectively support medical diagnosis, handle dangerous items or complex information, take over security, or even assist us in hospitality. The sheer number of tasks they can take on, the swift nature of the way they operate, and the speed of their learning makes it impossible for humans to keep up.

Whilst such efficiency has previously caused alarm amongst human workers when it comes to job security, the pace of change in the workplace, and the pressing need for tech-savvy upskilling, there's another concern growing in momentum...

- ▶ How can we keep track of their actions and decisions? And, how can we make sure to clarify accountability to reduce errors in the actions of an AI robot?

Cluster of evaluation

- ▶ Dr. Zofia Thoth, Durham University, GB [2022]

Negative	Neutral	Positive
Illegal actions Immoral actions	Permissible actions	Supererogatory actions
Commiting crimes or neglecting ethical and legislative guidelines	Actions that do not require moral evaluation (<i>cleaning, housekeeping, everyday tasks</i>)	Morally desirable actions (<i>Voluntary help, First-Aid, Civil Courage</i>)

Responsible AI

Responsible AI is an approach to developing and deploying (AI) from both an ethical and legal point of view. The goal of responsible AI is to employ AI in a safe, trustworthy and ethical fashion. Using AI responsibly should increase transparency and help reduce issues such as AI bias.

Proponents of responsible AI hope that a widely adopted governance framework of AI best practices makes it easier for organizations around the globe to ensure their AI programming is human-centered, interpretable and explainable.

The Seven Principles of RAI

Accountable and transparent

Explainable and interpretable

Fair with harmful bias managed

Privacy-enhanced

Secure and resilient

Valid and reliable

Safe

Case-studies

► Who is responsible?

- 1) Medical bots applying triage.
- 2) Care bots refusing orders for the patient's sake.
- 3) AMAs as CEOs ruining their company due to risky decisions.
- 4) Educational bots misseducating children.

VL V – Theories of justice and punishment, autonomy/heteronomy in Robo-Ethics



Basic questions of inquiry

- ▶ Can and should robots be punished for misbehavior?
- ▶ Can robots be considered guilty?
- ▶ Are robots truly autonomous in their actions?

PERSONHOOD

Forms of culpability/guilt

Legal culpability	Moral guilt	Metaphysical/ religious guilt
Infringement of law Infraction of other peoples' rights	Violation of norms, principles or guidelines, moral transgression	Violation of God's/Gods'/religious commandments
Murder, theft, homicide, torture, corruption etc.	Depraved behavior like abuse of confidence, egocentrism, exploitation of resources, rape etc.	Hybris, blasphemy, Seven Deathly Sins, neglection of rituals etc.
Sanctioning authority: Judiciary (in democracies) referring to constitutional amendments	Sanctioning authority: Bad conscience, society, peers etc.	Sanctioning authority (entity): God, Gods, religious community, sometimes state

Exceptional cases

There are situations coined by **particular circumstances**, in which there is a **lack of criminal responsibility** certified

- Age and maturity
- Inevitable mistake of law (fallacy of the perpetrator)
- Necessity as excuse (e.g. self-defence)

Purposes of criminal punishment

Absolute theories of punishment

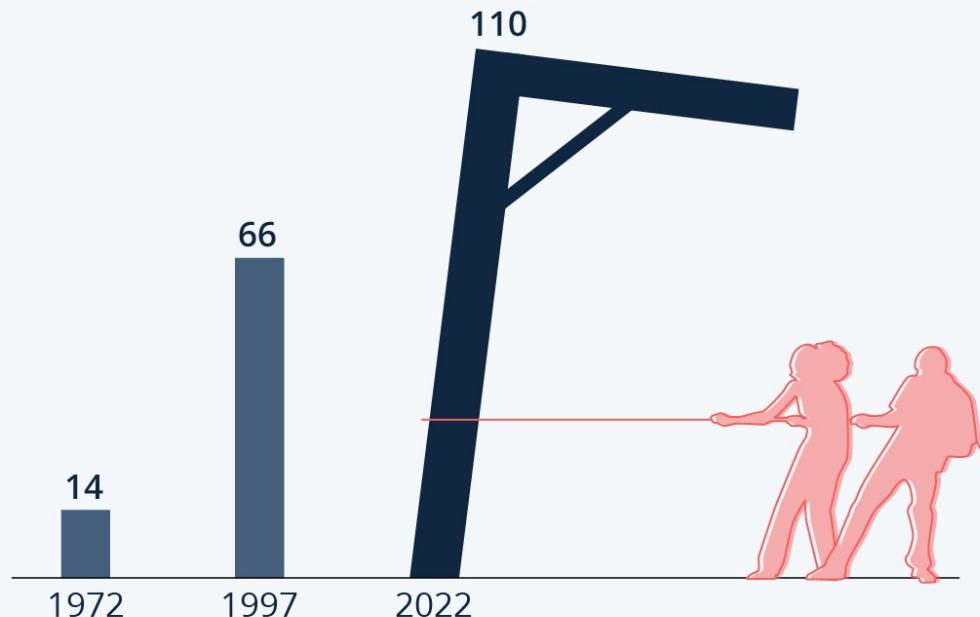
- ▶ Punishment because of fallacies
- ▶ Consequences of punishment are not considered

1) Retribution/Revenge (Kant/Hegel)

- ⇒ Reestablishment of legal order
- ⇒ **Talion principle** (*ius talionis*; aims for equation of crime and punishment, recidivists remain to be regarded as first offenders)
- ⇒ “*Eye for an eye, a tooth for a tooth*” (some deeds are exceeding the talion principle)

Death Penalty in Decline

Number of countries and territories that have officially abolished the death penalty for all crimes



Sources: Amnesty International, Death Penalty Information Center, Statista research



Purposes of criminal punishment

Absolute theories of punishment

- ▶ Punishment because of fallacies
- ▶ Consequences of punishment are not considered

2) **Atonement**

- ⇒ Penance and an act of contrition
- ⇒ Reconciliation between perpetrator and legal order
- ⇒ Implies volition



Purposes of criminal punishment

Relative theories of punishment

- ⇒ Punishment so that fallacies will be prevented
- ⇒ Future consequences of punishment are crucial

1) **Special prevention** (aims at perpetrators)

- *Positive SP:* Improvement of behavior and resocialising
- *Negative SP:* Deterrence (to set a warning example) and preventive detention



Purposes of criminal punishment

Relative theories of punishment

- ⇒ Punishment so that fallacies will be prevented
- ⇒ Future consequences of punishment are crucial

2) **General prevention** (aims at general public)

- **Positive GP:** Strengthens the peoples' trust in the legal system
- **Negative GP:** Deterrence of potential perpetrators

LILLIAN
GISH
in
**The
SCARLET
LETTER**

with

LARS HANSON
HENRY B. WALTHALL
KARL DANE

A
VICTOR SEASTROM
production

*Adaptation and scenario by Frances Marion
From the novel by Nathaniel Hawthorne
Directed by Victor Seastrom*



A Metro-Goldwyn-Mayer
PICTURE

MADE IN U. S. A.

Robot-Ethics

- ▶ Can we make use of these frameworks in Machine Learning?
- ▶ Are the principles adaptable for programming?

Controversy

Controversial insights

„An autonomous AI-robot that is considered a full moral agent and enjoys the benefits of legal protection has to be brought to book when trespassing legal boundaries. It is an indisputable consequence. It should be assumed, that its coding is capable of learning within the principles of legal theory and the various approaches of punishment. [...]

A recoding can not be the solution, the AI has to learn on its own.“

(Ray Kurzweil, 2016)

Controversial insights

„Shit happens. People die or get hurt everyday. Doesn't matter by whom or what. Reprogram the damn thing or delete it. It's made for specific tasks and doesn't have a free will or something like that. Stop making robots full legal agents.”

(Benjamin Zipher, University of Sydney, 2019)

Developmental psychology/Differential psychology and Robotics

Can Machine Learning profit from the scientific background of psychology and pedagogics?

Thesis:

In order to train wholesome AMAs, ML has to resort to psychology and built an AI-consciousness in reference to the moral development of human beings.

Prof. Dr. Birgit Lugrin, University of Würzburg

4 Stages of cognitive development - Jean Piaget

1. Sensomotorical Stage (0-2 years): In this stage, infants learn about the world through sensory experiences and motor actions. Key achievements include the development of object permanence and the coordination of sensory input with motor actions.

2. Preoperational Stage (3-7 years): Children in this stage develop language and the ability to use symbols, but they often lack logical reasoning. They exhibit egocentrism, meaning they have difficulty understanding perspectives other than their own.

4 Stages of cognitive development - Jean Piaget

3. Concrete Operational Stage (8-12 years): During this stage, children become more logical and can think more concretely about objects and events. They develop the ability to understand conservation (that certain properties of objects remain the same despite changes in appearance) and perform mental operations on concrete objects.

4. Formal Operational Stage (13 years to adulthood): In this final stage, individuals can think abstractly and hypothetically. They can engage in deductive reasoning, consider multiple perspectives, and think about possibilities and hypothetical situations.

Piaget's basic assumption: Constructivism

Piaget's theory emphasizes the active role of individuals in constructing their understanding of the world. It suggests that cognitive development is a continuous process marked by qualitative shifts in thinking as individuals progress through these stages.

Heteronomy vs. Autonomy

Lawrence Kohlberg transferred Piaget's cognitive development to the realm of moral development

Heteronomous Morality (Pre-conventional Level):

In this stage, typically observed in children, moral judgments are based on obedience and authority. The focus is on avoiding punishment and gaining rewards. Children in this stage tend to see rules as fixed and handed down by authorities, such as parents or teachers.

Heteronomy vs. Autonomy

Autonomous Morality (Conventional Level):

As individuals mature, they move into the stage of autonomous morality. In this stage, moral judgments are based on understanding and internalizing social rules. People at this level recognize that rules are flexible and can be modified through social agreement. Moral decisions are guided by principles of justice, and individuals consider multiple perspectives.

Robot-Ethics

A robot is in a heteronomous moral stage in dependence on its programs and patterns.

Yet it should decide autonomously?

- Requires personhood
- Otherwise no legal vindication either
- AI and ML has to reach the formal operational stage to built moral complexity
- As a new field of research, *Robopsychology* is on the rise

Reading task

Accountability in relation to autonomous driving

https://ieai.sot.tum.de/wp-content/uploads/2022/03/ResearchBrief_March_Boch_Hohma_Trauth_FINAL_V2.pdf

Lawrence Kohlberg

Lawrence Kohlberg was a psychologist who developed a widely influential theory of moral development. Kohlberg's theory is based on the idea that individuals go through distinct stages of moral reasoning as they mature. He presented his theory as an extension of Jean Piaget's cognitive development theory, building on the idea that cognitive development is closely linked to moral development.

Lawrence Kohlberg

- ▶ **Pre-conventional Level**
 - a. **Stage 1 - Obedience and Punishment Orientation:**
 - ▶ At this stage, individuals focus on avoiding punishment. The primary concern is with obeying authority figures to avoid negative consequences.
 - b. **Stage 2 - Individualism and Exchange:**
 - ▶ The focus shifts to self-interest. Individuals at this stage recognize that different people have different interests, and they begin to understand that there is an exchange of favors in relationships.

Lawrence Kohlberg

- ▶ **Conventional Level:**

- a. **Stage 3 - Interpersonal Relationships:**

At this stage, individuals are concerned with conformity and gaining the approval of others. They want to be seen as good and maintain positive relationships.

- b. **Stage 4 - Maintaining Social Order:**

The focus expands to societal rules and laws. Individuals at this stage are concerned with maintaining social order and obeying laws to uphold the functioning of society.

Lawrence Kohlberg

- ▶ **Post-conventional Level:**

- a. **Stage 5 - Social Contract and Individual Rights:**

Individuals at this stage recognize the importance of social contracts and agreements. They understand that laws are social agreements that can be changed for the greater good.

- b. **Stage 6 - Universal Principles:**

At the highest stage, individuals have a strong sense of individual conscience and moral principles that apply to all. They act according to self-chosen ethical principles, even if they conflict with societal laws.

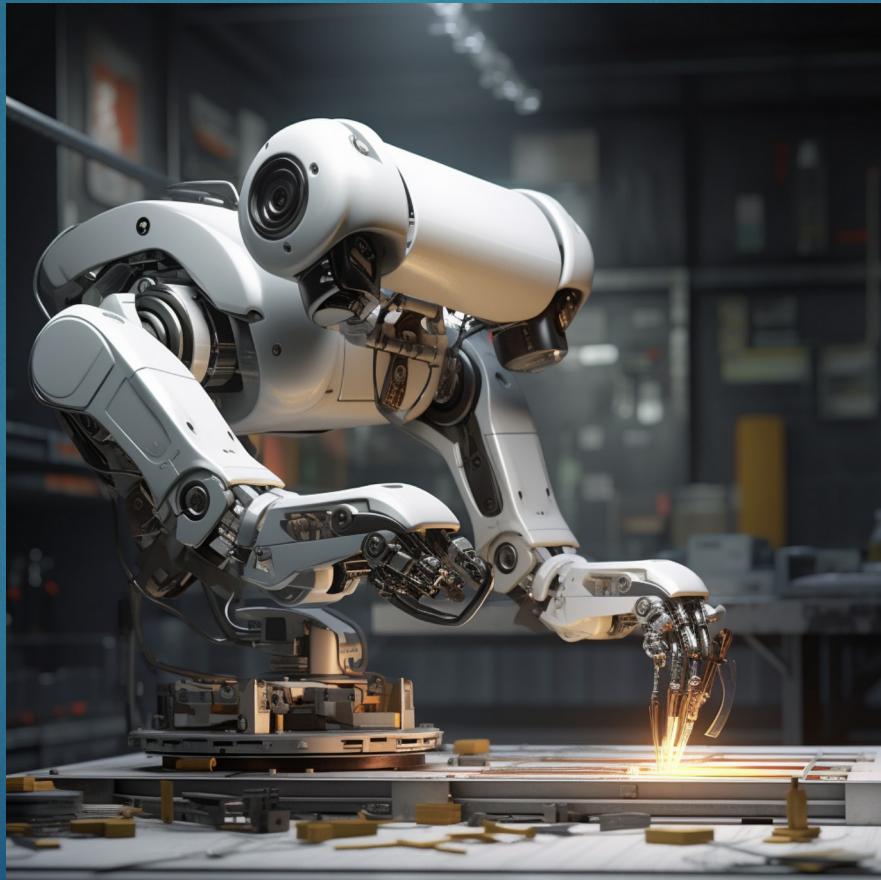


Aim for Robotics

It's important to note that not everyone progresses through all the stages, and progression is not strictly age-dependent. Kohlberg emphasized that moral development is a gradual process influenced by cognitive and social experiences. He also recognized cultural variations in moral reasoning.

We have to apply the post-conventional level to ML

VL VI - Production and employment market



Employment market

The term "employment market" refers to the overall environment or system in which employers seek to hire individuals, and individuals seek employment opportunities.

It encompasses the supply and demand for labor, job opportunities, and the interactions between employers and job seekers.

The employment market is influenced by various factors, including economic conditions, industry trends, technological advancements, and government policies.

Key components of the employment market *(Stetson 2018)*

- ❑ **Job Openings:** The number and types of positions available in various industries and sectors.
- ❑ **Labor Force:** The pool of individuals actively seeking employment or currently employed.
- ❑ **Unemployment Rate:** The percentage of the labor force that is unemployed and actively seeking employment.

Key components of the employment market *(Stetson 2018)*

- ❑ **Job Seekers:** Individuals looking for employment opportunities, including those entering the job market for the first time and those seeking new positions.
- ❑ **Employers:** Organizations and businesses seeking to fill positions with qualified and suitable candidates.
- ❑ **Wages and Compensation:** The levels of remuneration and benefits offered by employers for different roles.

Key components of the employment market *(Stetson 2018)*

- **Skills and Qualifications:** The demand for specific skills and qualifications in the workforce.
- **Economic Conditions:** The overall health and performance of the economy, which can impact hiring trends.

Robotics has a huge impact on these factors

Work ethics



©2013 Rendy Cipta Muliawan
Work hard...

Work ethics

T. Marek; W. Karwowski; M. Frankowicz; J. Kantola; P. Zgaga (2014).

Human Factors of a Global Society: A System of Systems

"Work ethic is a belief that work and diligence have a moral benefit and an inherent ability, virtue or value to strengthen character and individual abilities. Desire or determination to work serves as the foundation for values centered on the importance of work or industrious work. Social ingrainment of this value is considered to enhance character through hard work that is respective to an individual's field of work."

Work ethics

Why do human beings work?

- Self-fulfillment
- Respect and reputation
- Satisfaction
- Financial Security
- Career and personal development
- Contribution to society
- Social Interaction
- Innovation and creativity
- Maintaining skills and knowledge

What is good work?

- ▶ **Goal-oriented actions:** It is not about making plans or the next logical steps; it is about getting things done so that the work invested would not be counter-productive.
- ▶ **Prioritized focus:** Focusing on qualitative activities that a person is responsible for and in areas where they can make a difference or a high impact based on objectives.
- ▶ **Being available and reliable:** Spending time on the work and building oneself up for the task.

What is good work?

- ▶ **Conscientiousness:** A desire to do a task well, being vigilant and organized.
- ▶ **Creating a rewarding routine/system:** Engaging in tasks that provide strength and energy which can be transferred to your ultimate goals, creating a habit and a habitat for success.
- ▶ **Embracing positivism:** Shape a problem with the statement "good, (action) (problem)", e.g. "I'm tired and it is time for a workout" leads to "Good. Workout tired".

Swan, Andy (5 October 2016). ["7 Work Ethic Commandments For An Entrepreneur"](#). forbes.com.

Main expectations for robot-work

**Accurac
y**

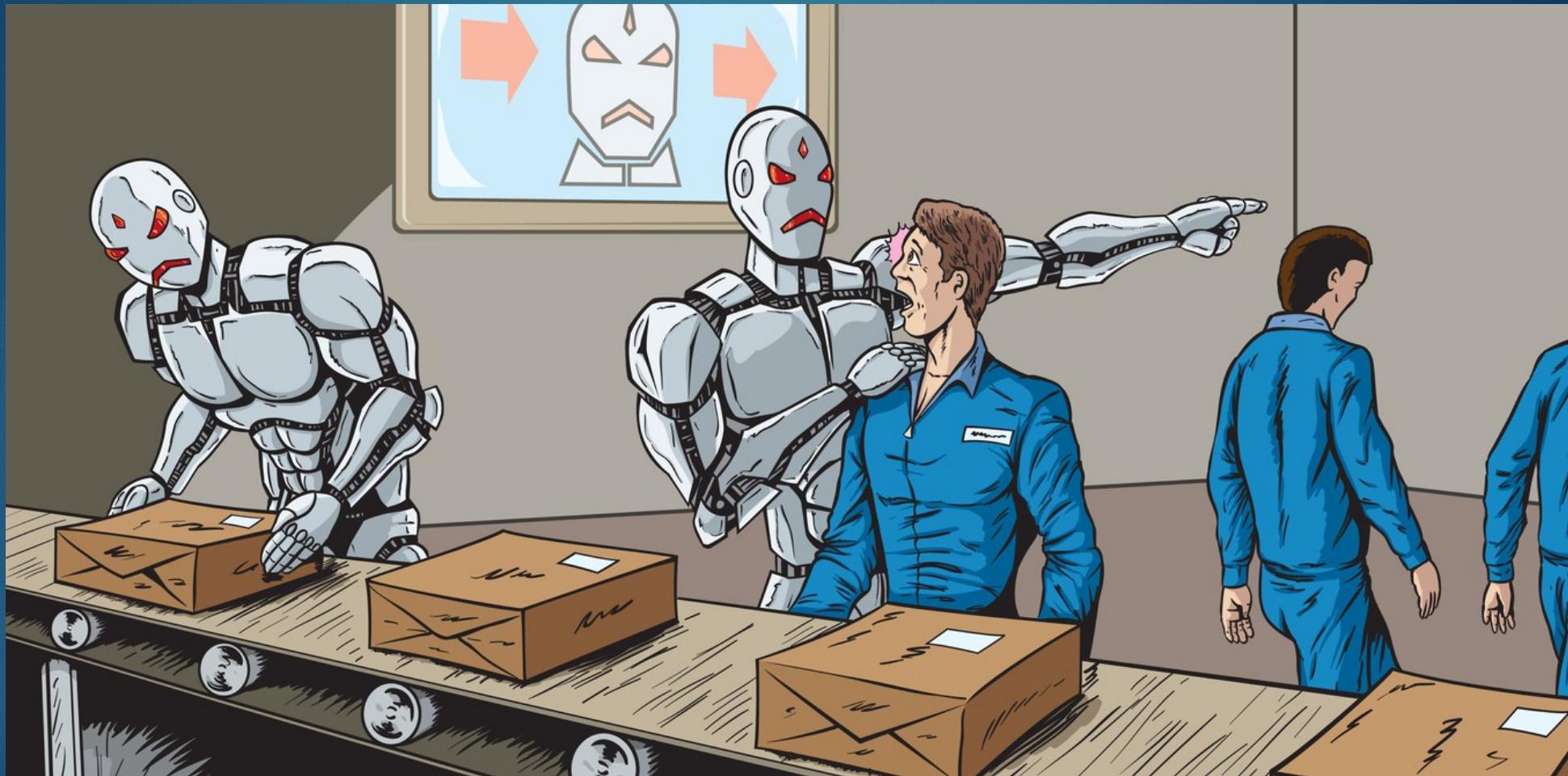
**Reliabili
ty**

**Efficienc
y**

Four main anxieties concerning robot-work

(Federal Statistical Office of Germany, 2022)

- ▶ Reliability
- ▶ Death of “from rags to riches”
- ▶ Humans will become redundant
- ▶ Loss of control



Robotic automation in the workforce (advantages)

► **Increased Efficiency and Productivity:**

Robots can perform repetitive tasks with precision and consistency, leading to increased productivity and efficiency in manufacturing and other industries.

► **24/7 Operation:**

Unlike human workers, robots can operate continuously without the need for breaks, sleep, or holidays, which can contribute to higher overall output.

► **Cost Savings:**

Over time, investing in robots can lead to cost savings for businesses, as they don't require salaries, benefits, or accommodations like humans. The initial investment in automation may be high, but ongoing operational costs can be lower.

Robotic automation in the workforce (advantages)

► **Improved Safety:**

Robots can handle dangerous tasks, work in hazardous environments, or perform tasks that pose risks to human health and safety, reducing the likelihood of workplace accidents.

► **Consistency in Quality:**

Robots can maintain a high level of consistency and precision in performing tasks, leading to improved product quality and reduced error rates.

► **Speed and Accuracy:**

Robots can perform tasks at a faster pace than humans and with a high degree of accuracy, especially in tasks that require repetitive and precise movements.

Robotic automation in the workforce (advantages)

► **Handling Heavy Loads/Complex Tasks:**

Robots are capable of lifting and moving heavy objects without risk of injury, making them suitable for tasks that involve significant physical, mental and cognitive strength.

► **Data Collection and Analysis:**

Robots equipped with sensors and artificial intelligence can collect and analyze data more efficiently, providing valuable insights for process optimization and decision-making.

Robotic automation in the workforce (disadvantages)

► **Job Displacement:**

One of the primary concerns is the potential displacement of human workers as automation takes over certain tasks. This can lead to unemployment and require reskilling for affected workers.

► **Initial Costs:**

The initial investment in acquiring and implementing robotic systems can be high, making it a barrier for some businesses, especially small and medium-sized enterprises.

► **Lack of Flexibility:**

While robots excel at repetitive tasks, they may lack the flexibility and adaptability that humans possess in responding to changes in tasks or environments.

Robotic automation in the workforce (disadvantages)

► **Maintenance and Downtime:**

Robots require regular maintenance, and breakdowns can result in significant downtime. Maintaining and repairing robotic systems can be complex and costly.

► **Dependency on Technology:**

Relying heavily on automation can make businesses vulnerable to technological failures, cybersecurity threats, or disruptions in the supply chain for replacement parts.

► **Reduced Human Interaction:**

In industries where interpersonal skills are essential, the introduction of robots may lead to a reduction in human interaction, affecting workplace culture and customer relationships.

Robotic automation in the workforce (disadvantages)

► **Ethical Concerns:**

The use of robots raises ethical questions, including issues related to job displacement, the potential misuse of technology, and concerns about privacy and surveillance.

► **Lack of Creativity and Intuition:**

Robots, as of now, lack the creativity, intuition, and problem-solving skills that humans possess. This limits their ability to handle tasks that require these qualities.

Dependent on specific industry and workspace

Employment fields

Will robots take jobs within these fields? (selection)

- Administration
- Industrial production
- Nurture and care
- Safety, rescue and security
- Teaching in schools and universities
- Law
- Medicine
- Politics
- Psychology/Psychiatry

Mandatory reading

(articles concerning the topic)

<https://rhsfinancial.com/2018/09/24/robots-jobs/>

<https://mitsloan.mit.edu/ideas-made-to-matter/a-new-study-measures-actual-impact-robots-jobs-its-significant>

<https://www.forbes.com/sites/garydrenik/2022/12/09/why-robots-are-taking-over-the-worldand-thats-a-good-thing/>

<https://www.theguardian.com/technology/2023/jun/03/ai-danger-doomsday-chatgpt-robots-fears>

VL VII - Human-Robot-Interaction and the Environmental Sector



Interdisciplinary Approach

- HCI (Human-Computer-Interaction)
- AI-studies
- Behavioral science
- Psychology
- Sociology
- Design
- Linguistics and communicational studies

Edwards, Chad; Edwards, Autumn; Spence, Patric R.; Westerman, David (21 December 2015).

"Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script"

. *Communication Studies*. **67** (2): 227-238.

- ▶ During initial interactions, people are more uncertain, anticipate less social presence, and have fewer positive feelings when thinking about interacting with robots, and prefer to communicate with a human. This finding has been called the **human-to-human interaction script**.

- ▶ It has been observed that when the robot performs a proactive behaviour and does not respect a "safety distance" (by penetrating the user space) the user sometimes expresses fear. This fear response is person-dependent.

Edwards, Chad; Edwards, Autumn; Spence, Patric R.; Westerman, David (21 December 2015).

"Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script"

. *Communication Studies*. **67** (2): 227-238.

- ▶ It has also been shown that when a robot has **no particular use**, negative feelings are often expressed. The robot is perceived as useless and its presence becomes annoying.
- ▶ People have also been shown to attribute personality characteristics to the robot that were not implemented in software.
- ▶ People similarly infer the mental states of both humans and robots, except for when robots and humans use non-literal language (such as sarcasm or white lies).

Edwards, Chad; Edwards, Autumn; Spence, Patric R.; Westerman, David (21 December 2015).

"Initial Interaction Expectations with Robots: Testing the Human-To-Human Interaction Script"

. *Communication Studies*. **67** (2): 227-238.

- ▶ Supervised exposure to a social robot can decrease uncertainty and increase willingness to interact with the robot, compared to pre-exposure attitudes toward robots as a class of agents.

- ▶ Interacting with a robot by looking at or touching the robot can reduce negative feelings that some people have about robots before interacting with them. Even imagined interaction can reduce negative feelings. However, in some cases, interacting with a robot can increase negative feelings for people with strong pre-existing negative sentiments towards robots

HFE

- ▶ **Human factors and ergonomics/ human factors engineering/ HFE** is the application of psychological and physiological principles to the engineering and design of products, processes, and systems.

Primary goals of HFE:

- reducing human error
- increasing productivity and system availability
- enhancing safety, health and comfort with a specific focus on the interaction between the human and equipment

Robotics in the environmental sector

Harvard University research projects

<https://projects.iq.harvard.edu/climatechangeworkshop/home>

- ▶ Climate change is the defining issue of our time. Scientific research predicts that there will be major impacts to all of humanity. It is essential for all of us to work to determine the potential consequences of climate change and to find ways to mitigate the negative effects.

Robotics in the fight against Climate Change

Scientific research on climate relies on the collection of extensive data over large areas of the globe and across long time scales. This can be facilitated by robotic sensors that rove autonomously across the atmosphere and oceans. Data collection can be greatly accelerated by swarms of robots that work in parallel to cover large areas.

Robotics in the fight against Climate Change

Reducing greenhouse gas emissions can be facilitated by robotic assistance in renewable energy generation.

This includes construction and maintenance of solar power systems, wind towers, and hydroelectric generating stations. In addition, robots can monitor industrial sites for emissions of greenhouse gasses. Robotics is also beginning to play a significant role in agriculture, by enhancing efficiency and reducing the use of chemicals with environmental impacts, and in recycling, by reducing the need to produce new resources.

Robotics in the fight against Climate Change

Mitigating climate disasters requires fast response in unpredictable and unstable environments.

Robotic systems can assist in responding to the predicted increase in floods by rescuing victims and assisting with post-flood recovery. Similar roles for robots can apply to wildfires. In addition, robots can help with restoring biological impacts of climate change, for example by planting trees and killing invasive species.

High rate of public acceptance and positive awareness

Robotics in the fight against Climate Change

Application fields

- ▶ Environment Monitoring and Management
- ▶ Distributed Robot Systems
- ▶ Autonomous Vehicle Navigation
- ▶ Aerial Systems: Applications
- ▶ Marine Robotics
- ▶ Robotics in Agriculture and Forestry
- ▶ Robotics in Hazardous Fields

Robotics in the fight against Climate Change

Mandatory Reading

<https://spectrum.ieee.org/robotics-climate-change>

Hans Jonas - Ethics of Responsibility

Hans Jonas' ethics of responsibility is a philosophical approach that emphasizes the moral imperative to consider the long-term consequences of human actions.

⇒ **The Imperative of Responsibility**

“Act so that the effects of your action are compatible with the permanence of genuine human life.”

“Act not destructively for future generations and the totality of their life conditions.”

Hans Jonas - Ethics of Responsibility

- 1. Principle of Responsibility:** Jonas argues that with increasing technological power comes an increasing responsibility to consider the potential impact of our actions on future generations and the environment.
- 2. Profound Future Consequences:** He emphasizes the need to recognize and address the profound consequences of contemporary human activities, especially in the realms of science and technology, which can have lasting and irreversible effects on the planet and future generations.

Hans Jonas - Ethics of Responsibility

3. Precautionary Principle: Jonas advocates for the precautionary principle, which suggests that if an action or policy has the potential to cause severe harm to the public or the environment, in the absence of scientific consensus, the burden of proof falls on those advocating the action.

4. Human Autonomy and Limits: While recognizing the importance of human autonomy, Jonas asserts that this autonomy should be exercised within certain limits that respect the integrity of nature and the well-being of future generations.

Hans Jonas - Ethics of Responsibility

5. Ethics of Care for Future Generations: Jonas emphasizes an ethics of care for future generations, arguing that our moral responsibility extends beyond our immediate concerns and includes a duty to safeguard the conditions necessary for the flourishing of life in the long term.

⇒ Sustainability

Hans Jonas - Ethics of Responsibility

Applicable towards

- Society
- Ecological challenges
- Consumerism
- **Robotics and AI**
- **Utopian future ethics**

Robotics and Future Ethics

Future Ethics

- ⇒ Sustainability
- ⇒ Responsibility
- ⇒ Utopian solutions

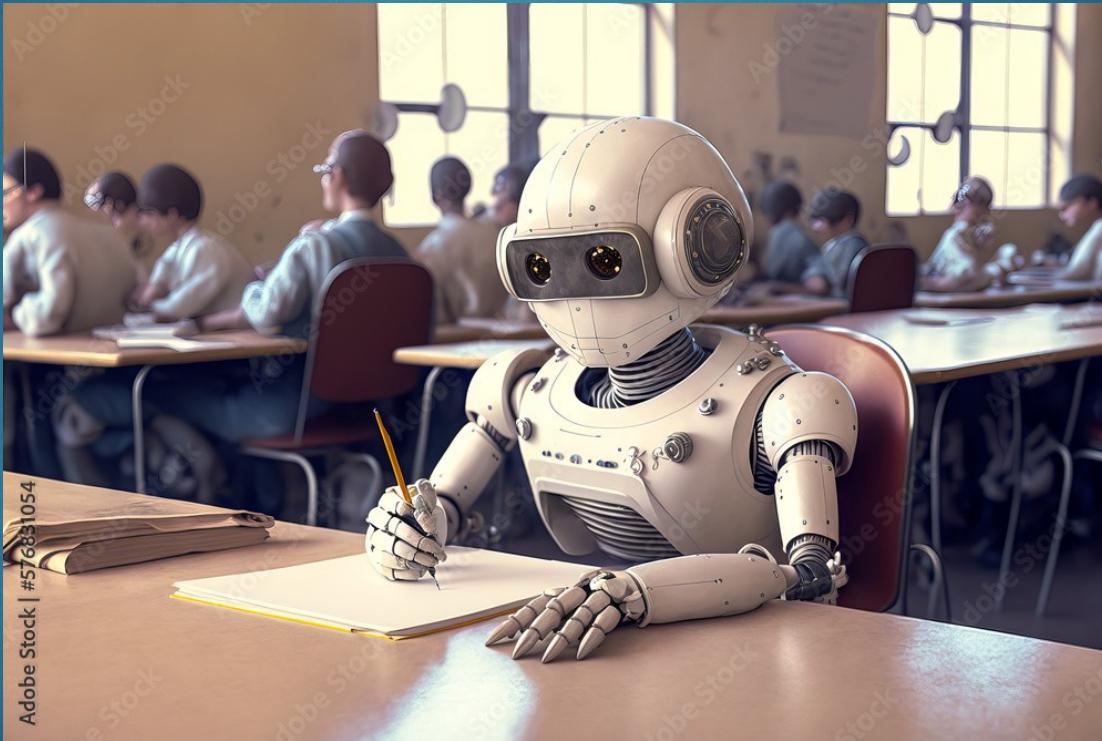
Spectrum

Humans, AI, Robotics, Society, Nature

Application Fields

Genetics, Human Enhancement, Ecology, Economics and Consumerism

VL VIII – Generic exam questions and “The Uncanny Valley”



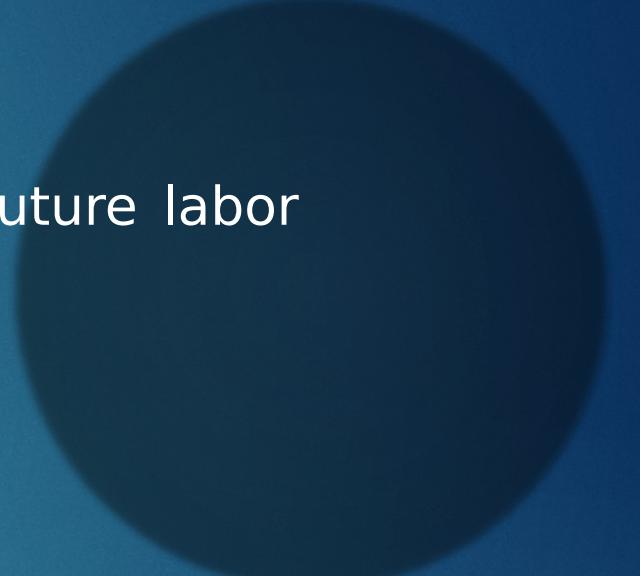
Three types of questions

I

- **Name/State/Enumerate ...**

- ⇒ Bulletpoints are satisfactory
- ⇒ No extensive explanation
- ⇒ No mere name- or term-dropping
- ⇒ Shorter answers with less points

|

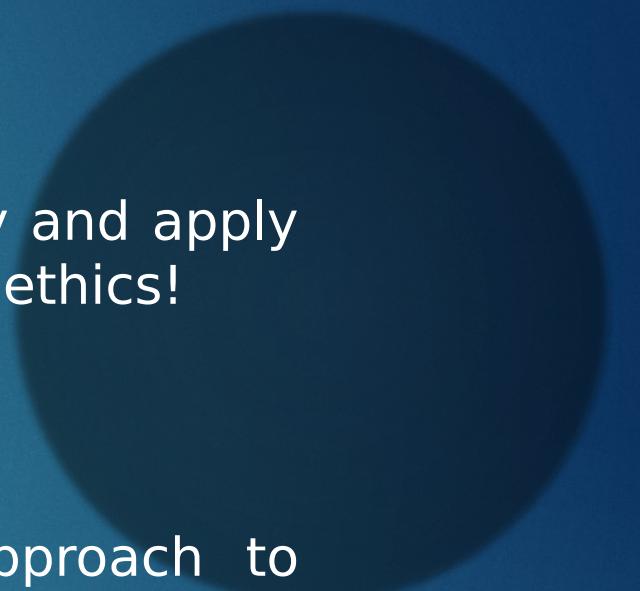


Name five distinct advantages that robots provide for future labor markets, include a concise explanation for each!

State the basic challenges of self-driving autonomous cars considering the different spheres of perception and cognition!

□ Explain/Illustrate...

- ⇒ Detailed explanation or analytic illustration
- ⇒ You have to display deeper knowledge
- ⇒ No bulletpoints



Explain the basic thoughts and arguments of deontology and apply its value system to a problem of your choice within robot ethics!

Illustrate one stage of your choice from Piaget's approach to cognitive development and explain why it can be regarded as helpful for machine learning!

- Analyse/argue/evaluate/assess/discuss
 - ⇒ You ponder profound and valid arguments
 - ⇒ You form a conclusion in the end
 - ⇒ If there is a quotation, put it into reference
 - ⇒ Most difficult task form, as you simultaneously have to include facts **and** clarify your individual position



“Consequentialism is not an apt moral basis for programming military robots that are supposed to make autonomous decisions in warfare.”

Discuss the validity of this statement!

Assess the feasibility of Aristotle's doctrine of *mesotes* for programming a care- and nurture-robot for elderly people!

The Uncanny Valley and its ethical implications

The uncanny valley is a concept in robotics and AI-studies that describes the unease or discomfort people may feel when confronted with a humanoid robot or animated character that closely resembles a human, but still has subtle imperfections.

The basic concept was introduced by Roboticist Masahiro Mori (1970).

The Uncanny Valley

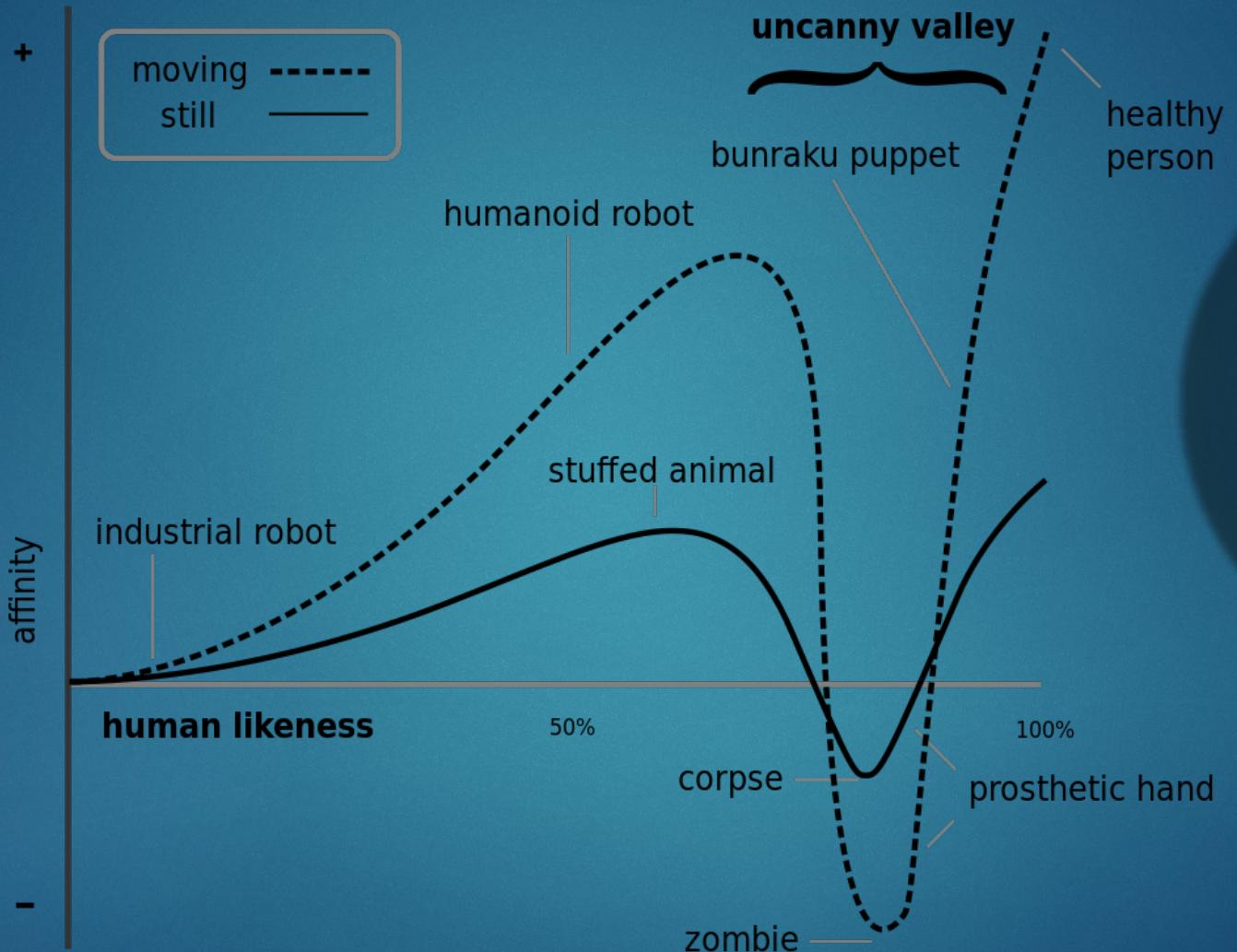
As a robot's appearance becomes more human-like, there is a point at which the emotional response of humans becomes increasingly **positive**.

However, as the robot becomes almost indistinguishable from a human but still falls short in certain aspects, such as facial expressions or movements, people may experience a strong **negative** sense of unease or revulsion.

The Uncanny Valley

This phenomenon is known as the "uncanny valley" because the comfort level drops into a valley before potentially rising again when the robot becomes virtually indistinguishable from a human.

Design and Aesthetics



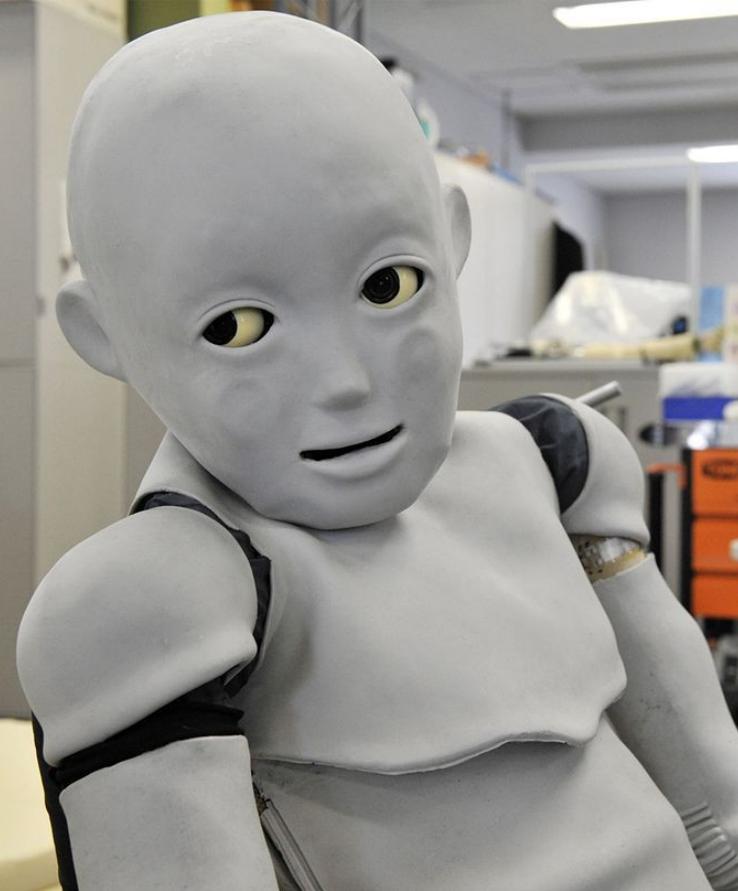
Design Principles

Non-mandatory reading recommendation:

Cheetham, Marcus; Suter, Pascal; Jäncke, Lutz (2011).

"The human likeness dimension of the 'uncanny valley hypothesis': behavioral and functional MRI findings". *Frontiers in Human Neuroscience.* London, England: Clarivate Analytics.

Check pdf-file! (mandatory)







VL IX – Robotics and Human Enhancement



Human Enhancement/Human Augmentation

Human enhancement refers to the use of various technologies, methodologies, or interventions to augment or improve human physical, cognitive, or psychological capacities beyond what is considered typical or "normal" for a human being.

This concept encompasses a broad range of possibilities, including advancements in biotechnology, genetics, information technology, pharmaceuticals, robotics and other scientific fields.

Internal and external enhancements

Forms of HE

Biotechnological Enhancements:

This involves using genetic engineering, gene editing and other biological interventions to enhance physical or cognitive traits. This includes gene therapy to eliminate certain diseases or enhance physical abilities (CRISPR) or to actively promote certain traits through gene editing.

Cognitive Enhancements:

Technologies or interventions that aim to improve cognitive functions such as memory, attention, intelligence, or problem-solving. Cognitive enhancement can involve drugs, brain stimulation, or other methods.

Forms of HE

Augmented Reality (AR) and Virtual Reality (VR):

Using AR and VR technologies to enhance human perception, provide additional information, or create immersive experiences that can contribute to learning or skill development.

Prosthetics and Exoskeletons:

The development of advanced prosthetic limbs or wearable exoskeletons to enhance physical abilities, strength, or mobility.

Forms of HE

Pharmaceutical/biochemical Interventions:

The use of drugs or substances to enhance cognitive function, mood, or physical performance. This includes nootropics, smart drugs, or performance-enhancing substances.

Regenerative Medicine:

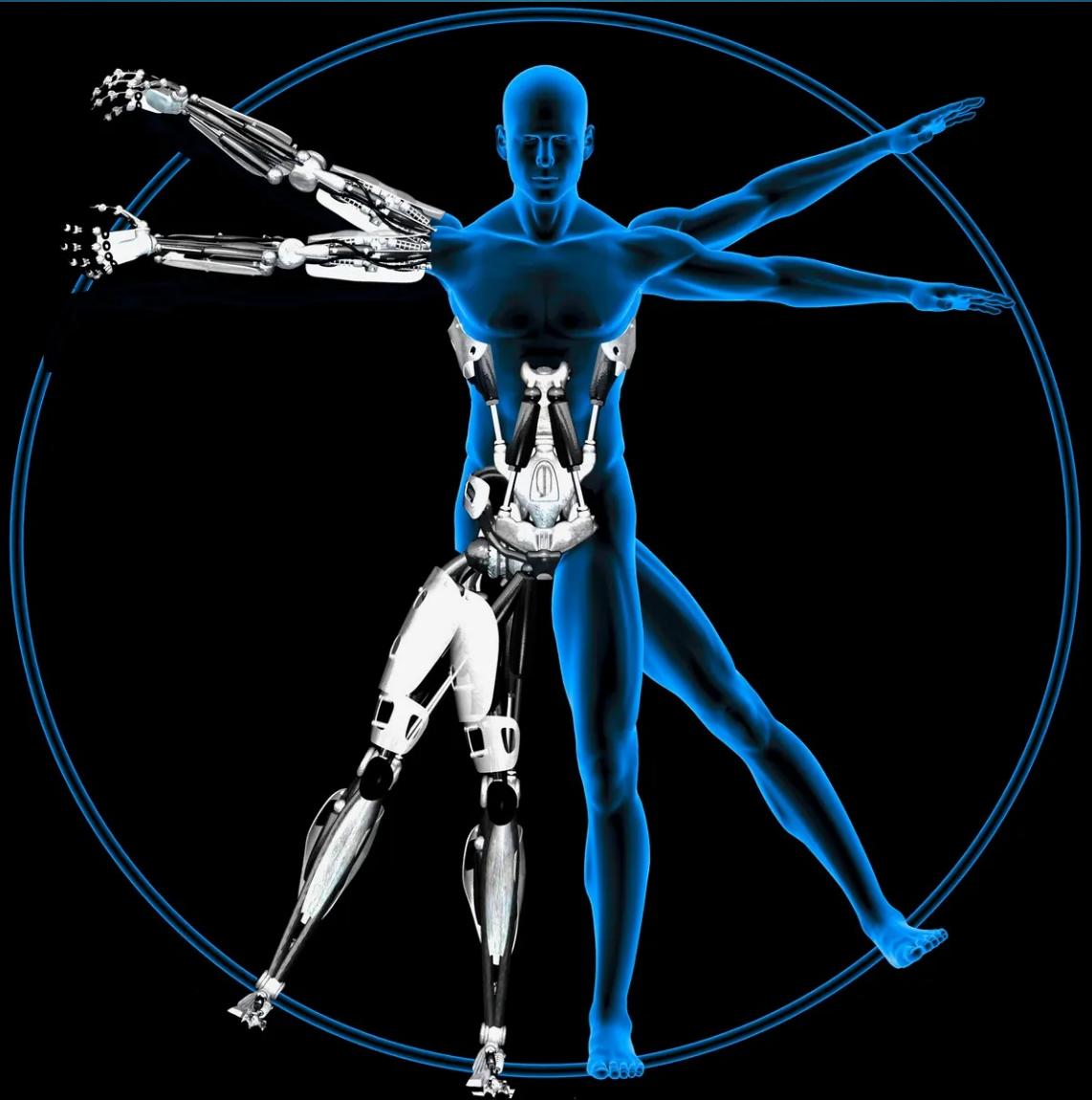
Using advanced medical techniques to repair or replace damaged tissues and organs, potentially extending human lifespan and improving overall health.

Forms of HE

Human-Machine Integration/Robotic Enhancements:

Combining human abilities with artificial intelligence, robotics or other technological advancements to create a symbiotic relationship, where humans and machines work together to achieve enhanced outcomes.

Sometimes referred to as **cyborg-technology**.



Transhumanism

Transhumanism is a philosophical and intellectual movement that advocates for the enhancement of human capabilities and attributes through the application of science and technology.

The goal of transhumanism is to overcome human limitations, both physical and intellectual, and to actively enhance human abilities and potential such as life-expectancy, performance or further inherent biophysical traits.

=> The Posthuman Condition

Transhumanism

- ▶ Irvin John Good (1965)

Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an "intelligence explosion," and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make.

Transhumanism

- ▶ Tom Morrow & Max More (1990)

Transhumanism is a class of philosophies that seek to guide us towards a posthuman condition. Transhumanism shares many elements of humanism, including a respect for reason and science, a commitment to progress, and a valuing of human (or transhuman) existence in this life. [...] Transhumanism differs from humanism in recognizing and anticipating the radical alterations in the nature and possibilities of our lives resulting from various sciences and technologies [...].

Transhumanism

► World-Transhumanist-Association/Humanity+ (2002)

- 1) The intellectual and cultural movement that affirms the possibility and desirability of fundamentally improving the human condition through applied reason, especially by developing and making widely available technologies to eliminate aging and to greatly enhance human intellectual, physical, and psychological capacities.
- 2) The study of the ramifications, promises, and potential dangers of technologies that will enable us to overcome fundamental human limitations, and the related study of the ethical matters involved in developing and using such technologies.

Transhumanist concepts (selection)

Emerging technologies

Hypothetical technologies

Innovation

Science fiction

Techno-utopia

Fourth Industrial Revolution

Post-scarcity economy

Technological change

Transhumanist ideology

Posthuman transhumanism

Eradication of sufferance

Utopianism

Extropianism (improving the human condition through tec-inventions)

Immortalism

Posthumanism

Technological utopianism

Dataism

Singularitarianism

Technogaianism

Ethical controversies among Transhumanists

Shall we overcome nature and the human body?

Shall we protect digital spheres more than natural ones?

Is human life worth saving or shall we progress into another step of evolution?

Discussion

1. Are humans enhanced through HE/HA still human beings and are they supposed to have the same rights/duties?

2. What could be moral doubts concerning the creation of “Super-humans”?

Immanuel Kant

“Act in such a way that you treat humanity, whether in your own person or in the person of any other, never merely as a means to an end, but always at the same time as an end.”

Groundwork in the Metaphysics of Moral

Instrumentalization/Self-instrumentalization?

Non-mandatory reading

<https://pdfs.semanticscholar.org/6871/e8c559755b9a96ca6c0395c43be6bce2c02b.pdf>

Organisational Matters

- ▶ **Tuesdays - regular seminars**
 - 09.01. (Military bots)
 - 16.01. (Robotics in relation to Aesthetics and Arts)
- + **Extra seminar:**
 - Saturday 19.01. - 14.15 – 15.45 + Revision
(Human-robot-relationships)
- **Test paper:** **Thursday 01.02.** **16.00 / 4p.m.**

VL X - Military Robots and their ethical implications



Moral concerns with military robots

- ▶ **Lack of Empathy and Ethical Judgment:**
 - ▶ *Disadvantage:* Robots lack human emotions, empathy, and ethical judgment. This absence of emotional intelligence can be a significant drawback when making complex moral decisions on the battlefield.
- ▶ **Autonomous Decision-Making:**
 - ▶ *Disadvantage:* The development of autonomous military robots capable of making decisions without direct human intervention raises concerns about accountability and the potential for unintended consequences.

Moral concerns with military robots

► Risk of Dehumanization:

- *Disadvantage:* The use of military robots may lead to a perception of warfare as more detached and clinical, potentially reducing the psychological barriers to engaging in armed conflict.

Dehumanization => rational, empathetic agency

► Unintended Consequences and Malfunctions:

- *Disadvantage:* Technical failures, software glitches, or hacking could lead to unintended consequences, including civilian harm or escalation of conflicts.

Moral concerns with military robots

- ▶ **Erosion of Accountability:**

- ▶ *Disadvantage:* There are concerns about the erosion of accountability when military decisions are delegated to machines. Determining responsibility for actions taken by autonomous systems becomes a complex ethical issue.

- ▶ **Arms Race and Proliferation:**

- ▶ *Disadvantage:* The development and deployment of military robots may contribute to an arms race, potentially leading to an increased risk of conflicts or misunderstandings between nations.



Moral concerns with military robots

"It is important to approach the ethical considerations of military robots with careful thought and consideration, as the development and deployment of such technologies raise complex moral and societal questions. International discussions and agreements are ongoing to address the ethical use of autonomous weapons in warfare."

Wired for War: The Robotics Revolution and Conflict in the 21st Century" by P.W. Singer.

Singer's main theses

- ▶ rapid evolution and integration of robotic technologies, including unmanned aerial vehicles (drones), ground robots, and autonomous systems, into modern military operations
- ▶ technological advancements are reshaping the dynamics of warfare, affecting strategies, tactics, and the nature of conflicts
- ▶ technologies may shape geopolitical landscapes and the nature of international relations on a global scale
- ▶ evolving relationship between humans and machines on the battlefield, highlighting the increasing reliance on automation and artificial intelligence in military decision-making processes

Advantages of military bots

- ▶ **Reduced Human Casualties:**

- ▶ *Advantage:* Military robots can potentially reduce the number of human casualties in armed conflicts. By deploying robots in dangerous situations, it may be possible to minimize the risk to human soldiers.

- ▶ **Precision and Accuracy:**

- ▶ *Advantage:* Robots equipped with advanced sensors and targeting systems can enhance precision in military operations. This precision can lead to reduced collateral damage and more accurate targeting of enemy combatants.

Advantages of military bots

- ▶ **Increased Surveillance Capability:**
 - ▶ *Advantage:* Military robots can provide enhanced surveillance capabilities, allowing for better monitoring of conflict zones and potential threats. This information can contribute to improved situational awareness.
- ▶ **24/7 Operations:**
 - ▶ *Advantage:* Robots do not require rest or sleep, enabling continuous operation. This can be advantageous in situations where human soldiers might need to rotate shifts or take breaks.

Advantages of military bots

- ▶ **Humanitarian Aid and Disaster Response:**
- ▶ *Advantage:* Military robots can be repurposed for humanitarian aid and disaster response. For example, robotic systems could assist in search and rescue operations in hazardous environments.



Recommended reading

<https://www.forbes.com/sites/naveenjoshi/2022/07/25/is-it-ethical-to-use-robots-in-war-what-are-the-risks-associated-with-it/>

Peace Ethics

Peace Ethics refers to a set of ethical principles and values that prioritize the promotion of peace and the resolution of conflicts in human interactions. This ethical framework emphasizes non-violence, cooperation, justice, and the well-being of individuals and communities, conflict resolution and human dignity. Peace ethics seeks to guide human behavior in a way that contributes to the establishment and maintenance of a harmonious and just society.



Bellum iustum theories

**Predominant question:
When is war vindicated?**

Bellum iustum theories

1) Plato

- protection/defense or aid for friends and alliances
- Aggressive, offensive wars and wars of conquest are not vindicated
- Only philosophers and ethicists are able to determine this

- Prisoners are not to be ravished, tortured or raped
- If the opponents are uncivilized barbarians, they must be extinguished

Bellum iustum theories

2) Cicero

- It is justified to react upon suffered injustices
- It is justified to react with war upon failed negotiations
- War has to be led by the political power, not by vigilantism
- Also religious authorities have to agree
- Restoration of peace (restitutio) and compensation for damage (repetitio)

Bellum iustum theories

3) Augustinus of Hippo

- The aim of war has to be peace (iustus finis)
- Against atrocities towards humans or the legal system (causa iusta)
- God or god's representatives have to declare war (legitima auctoritas)
- War has to be conform with God's laws

Bellum iustum theories

4) Immanuel Kant

War as a means to establish ...

- democracies
 - a Confederation Of States that is based on respect, equality and cooperation
 - International Laws and Treaties
-

- No standing armies
- Cosmopolitan Rights: Individuals have certain rights as citizens of the world

Enlightenment and hospitality

Bellum iustum theories

5) Public International Law (pax aeterna thoughts)

!!! Most relevant today !!!

- *Geneva conventions*
- Principles of distinction and proportionality
- Prohibition Indiscriminate Weapons and methods of warfare
- Protection of Cultural Property
- Human Rights Laws have to be obliged
- Responsibility to Protect (R2P from; genocide, ethnic cleansing, war crimes against humanity)
- The *United Nations Charter* and accountability

Geneva Conventions

The *Geneva Conventions* are a set of international treaties established to protect the rights and well-being of civilians, prisoners of war, and the sick and wounded during armed conflicts.

- Humanitarian Treatment
- Protection of wounded and sick
- Protection of civilians
- Non-discrimination
- Identification and Emblem (Red Cross/Red Crescent)
- Prohibition of certain acts of warfare
- Responsibility of High Contracting Parties

What are further risks of LAWS (Lethal Autonomous Weapon Systems)

=> Ethical concerns (as mentioned beforehand)

And...

- Legal concerns
- Safety and reliability
- Acceptance
- Bias
- Loss of human empathy and judgement

Application to robotics

Is it possible to uphold the standards of a *bellum iustum* within the possibilities and fatalities that military robots pose?

Application to robotics

Controversy in robot-ethics

Will robotics contribute to an escalation in warfare and cause more ferocious forms of combat?

Will military robots, which have been programmed with the Geneva Conventions, be morally more mature than human soldiers, as they are not led by emotions, rage, vengeance?

Organisational matters

- ▶ **Tuesdays - regular seminars**
- 09.01. (Military bots)
- 16.01. (Robotics in relation to Aesthetics and Arts)

+ Extra seminar:

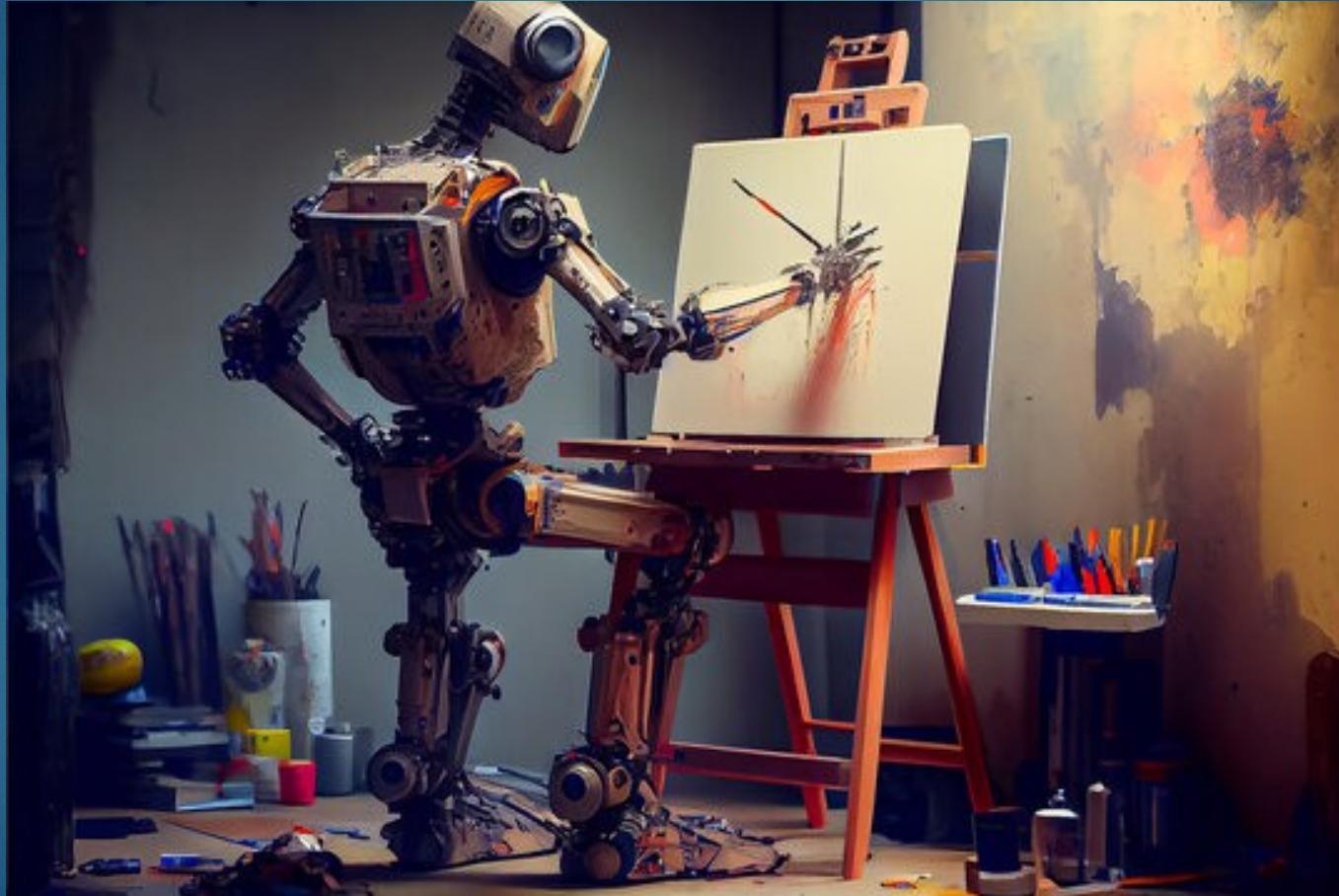
- Saturday 20.01. - 14.15 - 15.45 + Revision
(Human-robot-relationships)

□ **Test paper: Thursday 01.02.**

5 p.m./ 17.00 Room/Lecture Hall:

9.E.25

VL XI - Robotics and its impact on Aesthetics and Arts



Aesthetics

Aesthetics refers to the branch of philosophy that deals with the nature, beauty, and appreciation of art and taste. It involves the study and understanding of the principles that govern the perception of beauty, harmony, and the overall sensory or emotional experience derived from artistic expression, design, and other forms of creative work. Aesthetics explores questions related to what is considered pleasing, attractive, or meaningful, and it often involves subjective judgments influenced by cultural, historical, and individual perspectives. In a broader sense, aesthetics can extend beyond the realm of art to encompass various aspects of daily life, design, and sensory experiences.

Thesis

Robotics will contribute to form a new aesthetic conception including efficiency, functionality and even austerity.

It will not promote ideas of

- Splendidness
- Exuberance
- Pomp
- Flamboyancy
- decorativeness



How conceptions of beauty have changed throughout the centuries

The concept of beauty has evolved significantly throughout history and across cultures. Different eras and societies have had varying ideals of beauty influenced by cultural, social, religious, and artistic factors.

Conceptions of beauty

Ancient Civilizations

- **Egypt (3000 BCE - 30 BCE):** Beauty was associated with symmetry and proportion. Both men and women valued elaborate hairstyles and makeup. In ancient Egyptian art, individuals were often depicted in idealized forms.
- **Greece (c. 800 BCE - 146 BCE):** Ancient Greeks celebrated the beauty of the human body, emphasizing harmony and balance. The concept of "kalos kagathos," which signifies beauty and goodness, was central. Idealized proportions were depicted in art, such as the canon of Polykleitos.

Conceptions of beauty

Medieval and Renaissance (500 CE - 1600 CE):

- ▶ **Medieval Europe (500 CE - 1400 CE):** Beauty was often associated with religious piety and virtue. Paleness was considered desirable, reflecting a lack of outdoor labor.
- ▶ **Renaissance (14th - 17th centuries):** Renaissance art and culture revived the classical ideals of ancient Greece and Rome. Beauty was again associated with symmetry and proportion, but there was also a growing appreciation for individual beauty.

Conceptions of beauty

Baroque and Rococo (1600 CE - 1800 CE):

- ▶ **Baroque (17th century):** The Baroque era saw a shift towards dramatic and emotive expressions in art and architecture. Beauty became associated with grandeur, richness, and opulence.

- ▶ **Rococo (18th century):** The Rococo period emphasized ornate and delicate aesthetics. Beauty was linked to gracefulness, charm, and the playful use of asymmetry.

Conceptions of beauty

Neoclassicism and Romanticism (late 18th - 19th centuries):

- ▶ **Neoclassicism:** Revived classical ideals, emphasizing simplicity, order, and symmetry. Beauty was associated with reason and clarity.

- ▶ **Romanticism:** Embraced emotion, individualism, and nature. Beauty became more subjective and tied to personal expression.

Conceptions of beauty

Victorian Era (19th century):

- ▶ Beauty standards in the Victorian era were characterized by modesty and restraint. Pale skin, a tiny waist, and delicate features were considered desirable for women.

Early to Mid-20th Century:

- ▶ The early 20th century witnessed shifts in beauty ideals. The 1920s celebrated a more androgynous look, breaking from Victorian norms. In the mid-20th century, Hollywood played a significant role in shaping beauty standards, promoting glamorous and polished looks.

Conceptions of beauty

Late 20th Century to Present:

- ▶ Beauty standards have become more diverse and inclusive. There is a recognition of beauty in various body shapes, skin colors, and cultural backgrounds. The influence of fashion, media, and the rise of social media have contributed to a more dynamic and evolving concept of beauty.

- ▶ Social media has a significant impact on shaping beauty standards, and it often promotes a particular ideal that can vary across platforms and over time.

Conceptions of beauty

Social media platforms often promote an ideal of beauty characterized by a combination of youthfulness, slimness, and facial symmetry. Users frequently encounter images and influencers showcasing smooth skin, clear complexions, and features associated with a youthful appearance. The emphasis on a fit and toned physique is pervasive, with influencers often sharing workout routines and diet tips. Fashion trends play a significant role, as social media showcases and popularizes certain styles, contributing to the definition of an attractive aesthetic. The use of filters and editing tools further refines and idealizes images, creating a polished and often unrealistic standard. While there is an increasing push for diversity and inclusivity, traditional ideals persist, influencing perceptions of beauty on these platforms.

Robotics and its impacts

The fields of robotics and human enhancement have the potential to shape and promote new ideals of beauty. As technology advances, it introduces novel possibilities for modifying and augmenting various aspects of the human body, and these changes can influence societal perceptions of beauty.

- ⇒ **Enhancements and augmentations**
- ⇒ **Human-Machine Symbiosis**

Thesis

Robots will influence ideals of beauty promoting

- Power and speed
- Leanness and bulkiness
- Hard and clear features
- Straight linearity
- Concluded design elements



Impact on design

Robotics will influence

- ⇒ Graphic design, fashion design, interior design, architecture, web design, landscaping
- ⇒ Fine arts and refined art forms

Conclusion

The physical appearance of robots themselves is a design consideration. As robots are designed for various purposes, from industrial tasks to companionship, the aesthetics of their form can influence how they are perceived and accepted by humans.

Robot art

Robot art refers to artwork created with the assistance, collaboration, or direct involvement of robots. These robots can range from simple automated machines to sophisticated robotic systems equipped with various sensors, actuators, and programming capabilities. Robot art is a multidisciplinary field that combines elements of robotics, programming, engineering, and artistic expression.

Robot art

- ▶ **Collaboration with Artists:** In some cases, artists collaborate with engineers and programmers to design and build robots that can contribute to the creative process. These robots may be programmed to execute specific artistic tasks or generate unique artworks.
- ▶ **Generative Art:** Some robot art involves the use of algorithms and artificial intelligence to generate visual or auditory compositions autonomously. Robots equipped with sensors can respond to their environment or input data to create dynamic and evolving art pieces.

Robot art

- ▶ **Interactive Installations:** Robot art often includes interactive installations where the audience can engage with robotic systems to co-create or influence the artwork. This interaction can take various forms, such as gesture-based control, responsive movement, or the generation of art based on audience input.
- ▶ **Drawing and Painting Robots:** Robots can be designed to create visual art through drawing or painting. These robots may use traditional art mediums like pencils, brushes, or pens, or they may employ unconventional materials and techniques.

Robot art

- ▶ **Kinetic Sculptures:** Robot art sometimes manifests as kinetic sculptures, where robots or robotic components move in intricate patterns, creating visually dynamic and aesthetically engaging installations.
- ▶ **Robotic Performances:** Robots can be programmed to perform choreographed movements or dances as a form of artistic expression. These performances may integrate light, sound, and other sensory elements.

Robot art

- ▶ **Experiments with Robotic Tools:** Artists may explore the use of robotic tools and arms as extensions of their own creative abilities. This involves developing ways for artists to control and guide robotic systems to produce unique artworks.
- ▶ **Educational and Exploratory Projects:** Robot art is also a platform for educational purposes, encouraging the exploration of the intersection between technology and artistic expression. It provides a hands-on way for individuals to learn about robotics, programming, and creative collaboration.

Conclusion

“Robot art pushes the boundaries of traditional artistic practices, blurring the lines between human creativity and machine capabilities. It reflects the integration of technology into the art world, offering new perspectives on the creative process and the role of automation in shaping artistic expression.”

See Handout

⇒ Legal concerns in relation to robot art (obligatory reading)

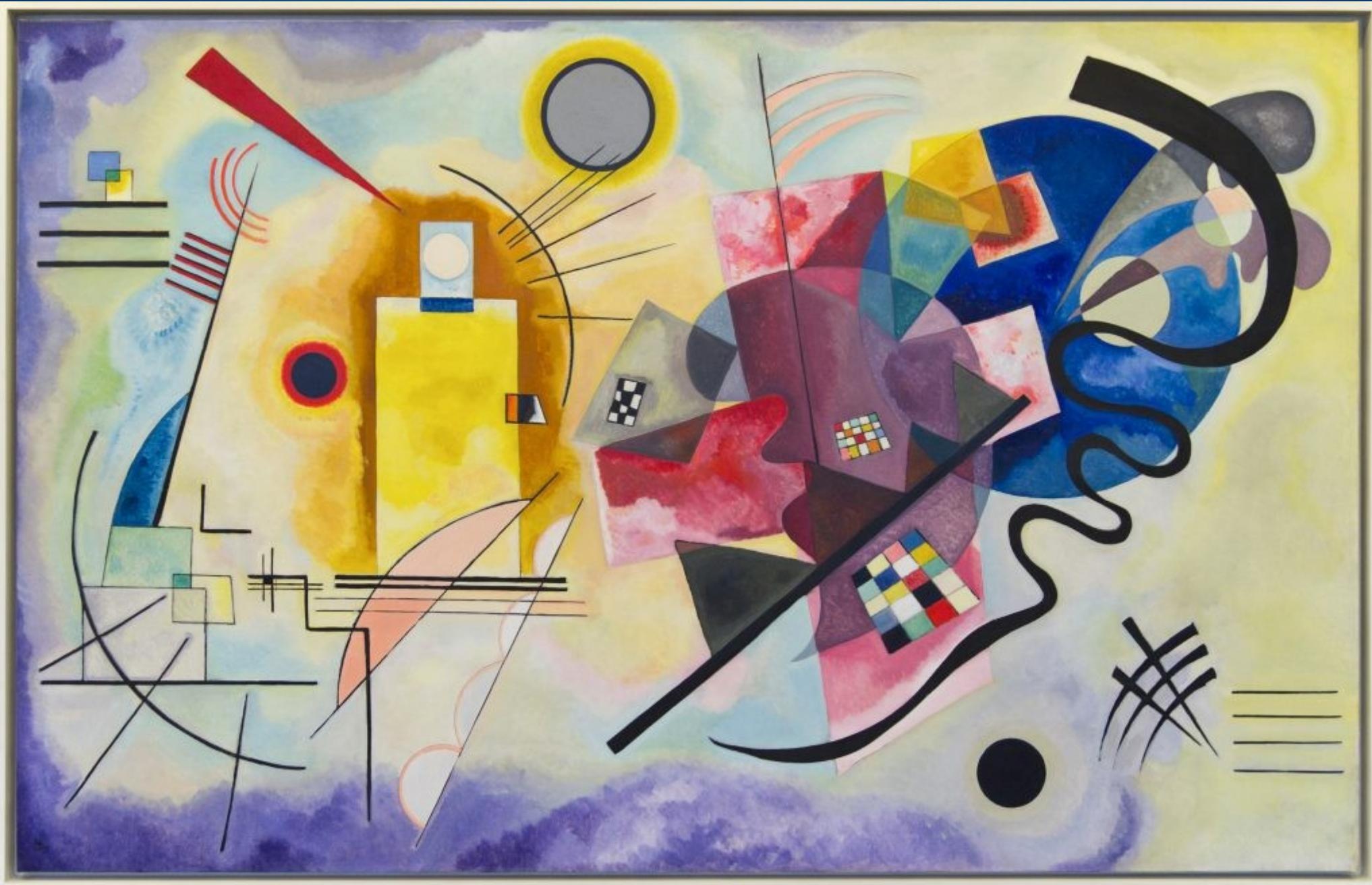
⇒ Recommended reading

<https://itsartlaw.org/2022/11/21/artistic-or-artificial-ai/>

<https://openai.com/dall-e-2>









Can robot art be considered art?

Artistic value refers to the perceived worth, significance, or merit of a work of art based on its aesthetic qualities, creativity, emotional impact, and cultural or intellectual relevance. It is a subjective and often elusive concept, as different individuals and cultures may have varying interpretations and criteria for evaluating artistic value.

Roland Barthes (1915-1980)

„Death of the Author“ => semiotics and signs (1967)

Death of the Author – Roland Barthes

1. Autonomous text
2. Reader-centeredness
3. De-emphasis of author's intent
4. Multiplicity of meanings

Robot poetry

Beneath the vast and azure sky,
Where clouds in softest whispers lie,
There blooms a joy, serene and bright,
A symphony of nature's light.

In meadows green and mountains tall,
The happiness, a gentle call,
Awakens in the heart's embrace,
The sublime dance of time and space.

Apply Barthes' theory to this robot-poem!

VL XII - Human-robot-relationships



© Shutterstock / Willyam Bradberry

H-R-R – thoughts and assumptions

The British tabloid newspaper DailyMail started a survey in which 27% of Millennials stated that they would consider dating a robot in the near future. (2023)

Sociologists suggest, that childcare tasks (like playing, helping with homework, cooking etc.) will be done in full or assisted custody by robots in the near future.

A (robot) friend in need is a (robot) friend indeed. And he/she/it will always be there for you.

As technology continues to advance, various forms of human-robot relationships may emerge, driven by advancements in artificial intelligence, robotics, and human-machine interaction.

H-R-R vs- H-R-I

Relationships are not mere interactions.

They have more normative and profound qualities.

A relationship is a connection or association between two or more entities, whether they are individuals, groups, organizations, or even abstract concepts. Relationships can take various forms and exist in different contexts, including personal, social, professional, familial, or romantic settings.

Factors for establishing a strong bond and relationship

Set I

- ▶ Communication
- ▶ Trust
- ▶ Empathy
- ▶ Sympathy
- ▶ Shared goals and values
- ▶ Respect
- ▶ Shared history

Set II

- ▶ Conflict resolution skills
- ▶ Positive interactions
- ▶ Adaptability
- ▶ Autonomy and independence
- ▶ Commitment
- ▶ Forgiveness
- ▶ Shared experiences

Forms of human-robot-relationships

► **Assistive Companionship:**

Robots designed to provide emotional support, companionship, and assistance to individuals, especially the elderly or those with disabilities.

- Psychological support and therapy
- Physiological support and therapy
- Eldercare
- Childcare
- Robot as a caregiver

Forms of human-robot-relationships

► **Professional Collaboration and co-working:**

Work environments where humans collaborate closely with robots to enhance productivity and efficiency, such as in manufacturing, healthcare, and other industries.

- Industry
- Administration
- Research and archiving
- Medical support
- Even creative assistance etc.

Forms of human-robot-relationships

► **Educational Companions:**

Robots designed to assist in education by providing personalized tutoring, interactive learning experiences, and educational support to students in schools and universities.

- Increases the possibilities of autodidactic learning
- Robot as an expert

Forms of human-robot-relationships

► **Personalized Healthcare:**

Robots working alongside healthcare professionals to provide personalized care, monitor patients, and assist with medical procedures.

- Very personalized and intimate way of relating



shutterstock.com · 2278021823

Forms of human-robot-relationships

► **Social Companions:**

Robots designed for social interactions, entertainment, and leisure activities, serving as companions in various aspects of daily life.

- Robot as a buddy
- Robot as a guide/tummler
- Sports, travelling, adventuring
- Robot as a friend

Forms of human-robot relationships

► **Intimate Relationships:**

The emergence of robots designed for emotional and physical intimacy, although this raises ethical and societal questions and challenges.

- Robot as a partner
- Robot as a sex-partner
- Robot as wife-/husband material
- Questions of tabooing

Forms of human-robot-relationships

► **Collaborative Creativity:**

Human-robot partnerships in creative fields such as art, music, and design, where robots assist humans in the creative process or even generate their own artistic works.

- See previous session

Forms of human-robot-relationships

► **Personal Assistants and Home Automation:**

Robots serving as personal assistants for tasks like scheduling, reminders, and managing smart home systems to enhance convenience and efficiency.

- Robot as an assistant
- Robot as a servant/butler

Forms of human-robot-relationships

► **Emotional Support Robots:**

Robots equipped with advanced emotional intelligence to provide empathetic responses and support during challenging times.

- Individual interactions
- Catastrophes
- Wars
- Natural disasters
- Robot as an expert/friend/“good soul”

Forms of human-robot-relationships

Take-for-granted Environmental Interaction:

Robots designed to interact with the physical environment, such as cleaning robots, security robots, and delivery robots that work alongside humans as if it was a natural interaction.

- Robots as genuine parts of daily life
- Robots as equipment

Forms of human-robot-relationships

► **Remote Presence:**

Robots acting as big-brother-surveillers.

Robots acting as avatars for individuals in remote locations, allowing them to virtually participate in events, meetings, or social gatherings.

- Role as a supervisor
- Role as a substitute

Aspects of design

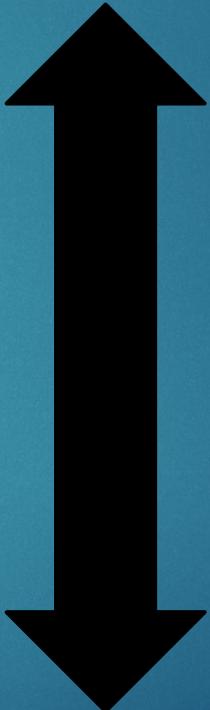
Personalization and customization:

As robotics technology advances, robots may become more personalized and tailored to individual needs. They could be designed to understand and adapt to human preferences, behaviors, and emotions. This could lead to robots serving as companions, caregivers, or even therapeutic tools in various contexts.

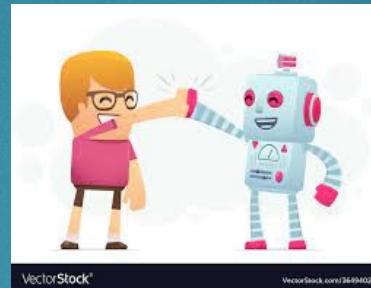
The future of the relationship between humans and robots, Simon Alba, 2023

How do we approach H-R-R?

- ▶ Robots as assistants?
- ▶ Robots as avatars?
- ▶ Robots as partners?
- ▶ Robots as mere tools or slaves?
- ▶ Robots with human rights?
- ▶ Robots as equals?
- ▶ Robots as a superior form of being?



Can robots be true friends?



Aristotle – Nichomachean Ethics

► **Friendship of Utility**

- This type of friendship is based on mutual benefit and usefulness.
- People engage in this kind of friendship because they find each other useful in some way.
- The relationship is often instrumental, and friends are valued for what they can provide or help with.
- These friendships are more common in business or practical contexts, where individuals come together for mutual gain.

Focused on one-/two-sided advantage

Aristotle – Nichomachean Ethics

► Friendship of Pleasure

- This type of friendship is centered around the enjoyment and pleasure that friends find in each other's company.
- The bond is based on shared activities, interests, or simply the joy derived from being together.
- Friendships of pleasure are often more transient and can fade if the source of pleasure diminishes.
- Hedonistic aspects prevail

Rather short-termed form

Aristotle – Nichomachean Ethics

► **Complete or Virtuous Friendship**

- According to Aristotle, the highest and most valuable form of friendship is the true friendship of the good.
- This type of friendship is based on mutual respect, admiration of the other's character traits, and a shared commitment to virtue and moral excellence.
- Friends in this category genuinely care for each other's well-being and are concerned with the overall development of each other's character and identity and are open for emotional commitment.
- Aristotle considers this form of friendship to be rare and enduring, with friends seeing each other as extensions of themselves.

Controversy in robot-ethics

- ▶ Numerous robot-ethicists consider a Complete Friendship between humans and robots as impossible, as the robot lacks the genuine emotional affiliation
- ▶ **Conclusion:** Robots remain tools that mimic personhood, but they lack “heart and soul”

Do you approve this conclusion?

Mandatory reading

- ▶ See handouts

Are h-r-r a benefit to society or are they being too much idealized?



Revision

- ▶ Questions from the students

Good luck for your exam preparation!

I hope you were able to get an intriguing insight into Robot Ethics, a currently much debated field in Applied Ethics and Digital Science.

I wish you all the best for your studies and your personal lives!

M. Berthold

