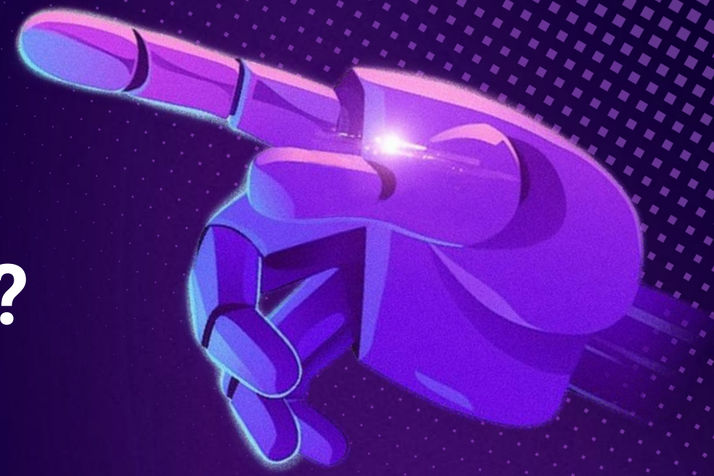
2026

ETHICS AND ARTIFICIAL INTELLIGENCE

Presented by:
Paola Doždor,Ljubica Agić,Petra Mrše,Ana Patrlj,Ema Marin, Maria Vukman i Antonija Kovač 4.c

- • Can a machine be moral?
- Does AI distinguish between good and evil—or does it simply execute commands?
- • Is intelligence sufficient for morality?



WHAT IS ETHICS?

- "The world is not perfect. Neither are we, either as individuals or as a group" (Boddington, 2017).
- **Ethics = the philosophical discipline concerning good and evil / moral philosophy**
- It deals with:
- 1. **Well-being and harm**
- 2. **Justice and fairness**
- It is particularly important in the development of artificial intelligence.
- **Normative ethics** -> guidelines for making correct decisions.

Classical Ethical Theories

Virtue Ethics

WHAT KIND OF PERSON SHOULD I BE?

Focus on character

• Aristotle – The Golden Mean

Consequentialism

- The consequences of a behavior are the ultimate basis for judging the correctness of that action (a posteriori).
- • Jeremy Bentham – Utilitarianism
- • John Stuart Mill

Deontological Ethics

- A moral theory according to which the rightness of an action lies in the act itself, rather than in its consequences (a priori).
- Immanuel Kant – The Categorical Imperative

AI through ethical theories

Utilitarianism:

AI chooses the option that is used by the most people

Focus because on the consequences of decisions

What happens when rules and consequences conflict?

If AI follows rules at all costs → bad outcomes may occur
If AI only looks at consequences → it may violate important moral rules

Example
Autonomous car:
should one follow the rule (not stray from the lane)
or avoid an accident (even if it means breaking the rule)?

Can AI have morality?

What is morality?

- the ability to distinguish between good and evil
- making decisions based on values
- a sense of responsibility for actions

Key elements of morality:

- awareness (understanding oneself and others)
- intention (conscious decision-making)
- responsibility (bearing the consequences)

AI systems

- analyze data and patterns
- have no consciousness or emotions
- are not responsible

- example: a self-driving car must make a split-second decision: turn and endanger a passenger or continue straight and hit a pedestrian

CONCLUSION:

- AI does not think and does not possess morality, but simulates thinking and morality

The problem of responsibility

If artificial intelligence causes harm (e.g. wrong medical decision, traffic accident)
Who is responsible?

- developers
- companies
- users
- AI system

Problem:

- AI can cause serious demage (making wrong decisions)

Key fact:

- makes decisions, but has no consciousness, intention, or moral responsibility

Philosophical dilemma:

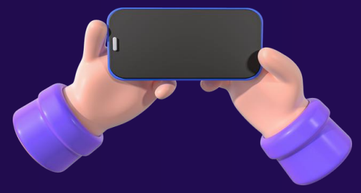
- **can there be action without a responsible subject?**

Consequence:

- there is a void of responsibility

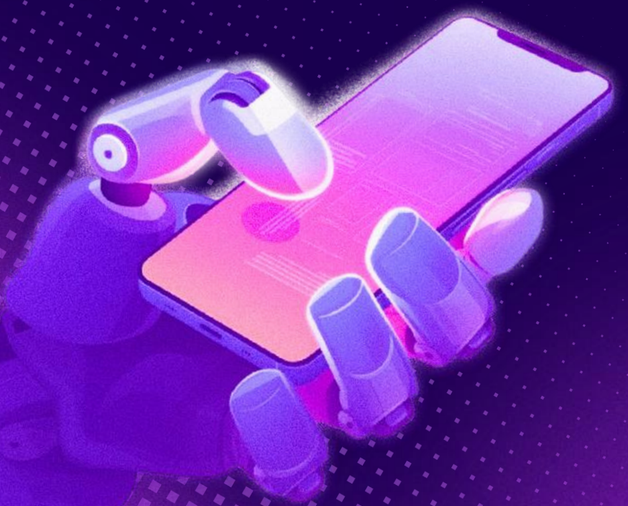
Freedom and autonomy

- Free will = the foundation of ethics
- Humans make conscious decisions
- AI acts according to algorithms
- It has no consciousness or moral intent
- Who bears moral responsibility for AI?



Stuart Russel

Stuart Russell is one of the most famous experts on artificial intelligence and a professor at Berkeley
• His main concern is how to prevent AI from becoming dangerous to humans



Ključne ideje:

AI must be aligned with human values
→ Problem: how to teach a machine what “real” human values are?
Criticizes the idea that AI has fixed goals
→ If AI gets the goal wrong, it can do great harm (e.g. achieve the goal in a way that harms humans)
Proposes a model where:
• AI does not know exactly what it wants (is not “sure”)
• learns from human behavior
• always allows human control

“We don't just need powerful AI, we need safe AI."

Yuval Noah Harari

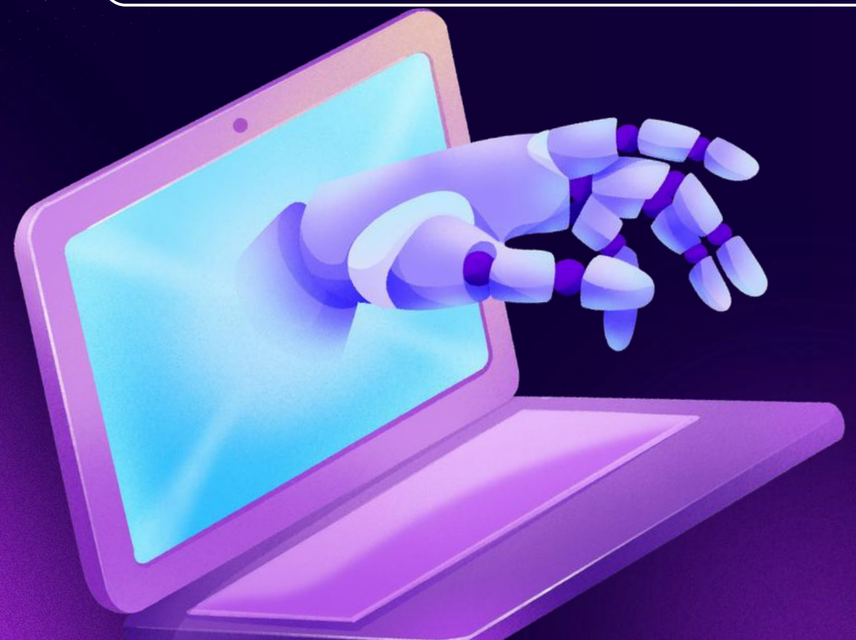
The impact of artificial intelligence on man and society

AI can change the way people live, work and make decisions

the power of data and information

AI can analyze human behavior and habits

loss of independent thinking



Key issues:

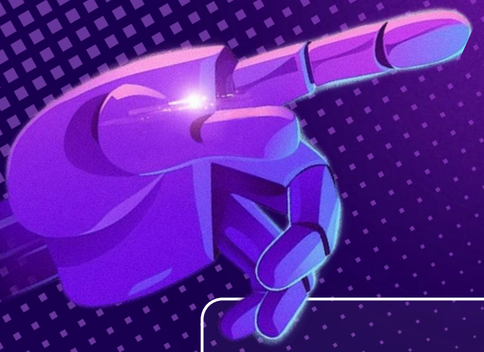
- **Loss of privacy**
- **Manipulation of people**
- **Unemployment and inequality**
- **Free will (“if AI can predict our decisions are we really free?”)**

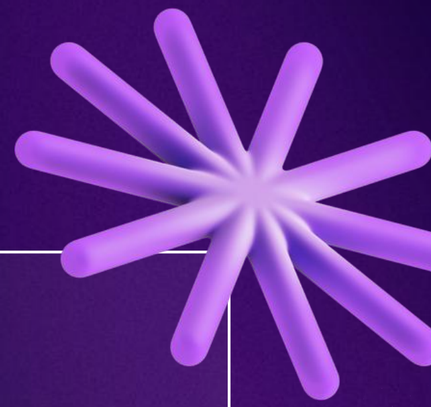
Truth and manipulation

- AI can create text, images and videos
- The emergence of deepfakes and disinformation
- It is difficult to distinguish real from artificial
- The impact on opinion, media and society
- Loss of trust in information
- What is truth in the era of artificial intelligence?
- Who bears responsibility for manipulation?

AI and the future of man

- AI can now learn, analyze data, make decisions, and create content, which was once exclusively human
- therefore, the line between man and machine is increasingly blurred
- The increasing reliance on AI is affecting our everyday lives
- people are increasingly leaving decisions to algorithms, from simple to more important life choices
- this can make life easier, but it can also reduce our independence and critical thinking





The limits of technology

If we can do something, does that mean we can?

Examples:

artificial intelligence in decision making

mass surveillance and privacy

information manipulation

Issues:

- Technology is advancing faster than ethics
- Potential for misuse (or possible abuses)
- Diminishing human control

Who sets the boundaries?

- Society
- Laws (or legislation)
- Ethical principles


THE END