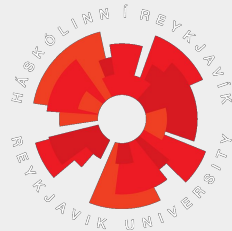


AI, Ethics, and Society

14.09.2026



Who am I?

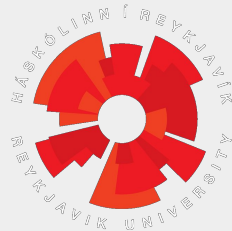
Leonard Matthias Eberding

- Ph.D. student at CADIA researching Empirical Reasoning Systems
- Based in Berlin, Germany

Background:

I came to AI from biology, not from computer science:

- B.Sc. in Biomimetics
- M.Sc. in Navigation and Field Robotics



AI, Ethics, and Society

On AI:

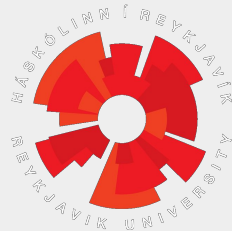
AI systems are making decisions affecting millions of people. They embody the values of the people who built it, not necessarily ethical ones.

On Ethics:

Ethics is not a rule, but the foundation that everything else is built on.
Laws codify it, products embody it, and society is its result.

On Society:

A society is as strong as its weakest link. Discrimination against the vulnerable isn't just immoral, it is self-detrimental.



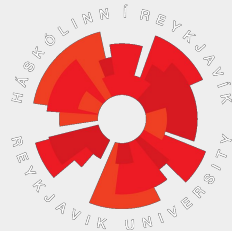
About this course

I am open to ideas and suggestions from you on where we go with this course.

Some hard constraints I have:

1. Moral Theories
2. Principles of AI and Ethics
3. Ethics-by-Design and participatory Ethics
4. Next-Gen AI

Everything else we can move around. So if you have any ideas or want to know about something in particular, I am happy to dig into them.



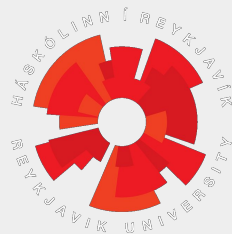
About this course

Last year, for example we did:

1. AI in practice, what can go wrong, what can go right?
2. EU AI Act

But we can also look more detailed at, for example:

1. Digital Colonialism through AI
2. Data maintenance
3. Etc.



About this course

Grading:

Assignments + Final exam. No midterms

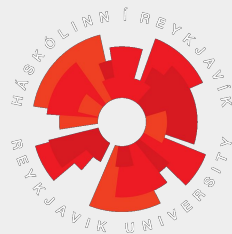
Assignments:

After each lecture you will have to write a reflection on the lecture and ask 2 questions on canvas about the last lecture. 1% of the grade each.

Two in-class assignments, approximately 6 and 14 %, respectively.

Final Exam:

Written 3 hour exam. Roughly 65-70 %



About this course

Guest Teachers:

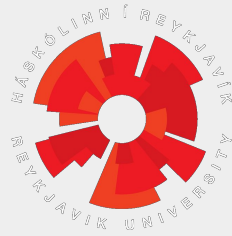
You already met Berglind, there will be probably one more guest teacher. TBD.

Lecture format:

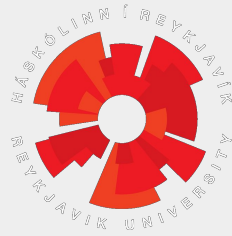
All future sessions (except guest teachers, if applicable), will be recorded and put online.

Alternating between online and in-person. For the next two weeks online, then two weeks in person. I'll let you know when and how exactly.

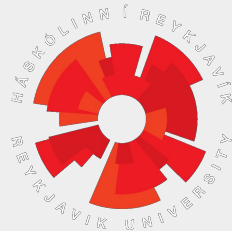
Any questions so far?



Or any ideas what you want to talk about?



Let's get going with **Moral Theories** then!

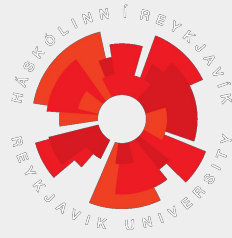


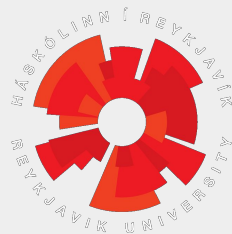
Moral Theories I

14.09.2026

Concepts - Morality

What is Morality?





Concepts - Morality

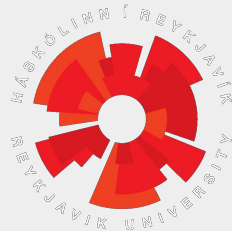
What is Morality?

Morality refers to a system of principles or rules that guide human behaviour.

Particularly in matters of right and wrong, fairness, justice, and harm.

However, defining it precisely isn't easy because morality can mean different things in different contexts.

E.g.: Cultural, legal, philosophical, and emotional



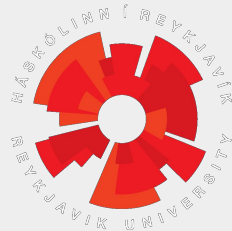
Concepts - Morality

Why is defining morality difficult?

Different philosophical traditions and cultures offer competing definitions.

Some focus on social norms, other on rational justification, others on universal truths.

This creates tension between understanding what people believe is moral and what actually is moral.



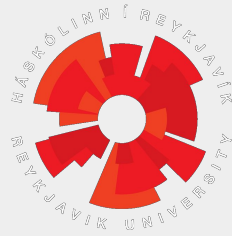
Concepts - Morality

Descriptive Morality:

A sociological or anthropological approach.

It describes whatever rules, values, or practices are accepted by a group or society.

This includes customs, taboos, and shared moral intuitions. Even if they differ dramatically across cultures.



Concepts - Morality

Normative Morality:

A philosophical approach.

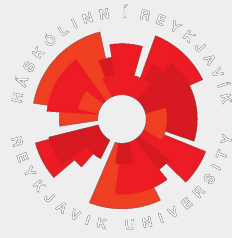
It is not just what people happen to believe (i.e., descriptive), but what they ought to believe.

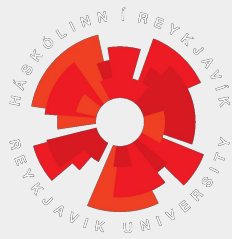
It refers to codes of conduct that are justified.

This justification includes reasoning, ethical theory, or principles regardless of social consensus.

Concepts - Ethics

What is Ethics?





Concepts - Ethics

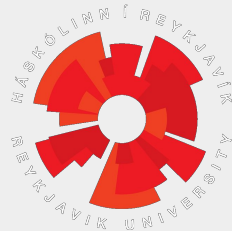
What is Ethics?

Ethics is the philosophical study of morality.

It examines how we ought to act, what values should hold, and how to justify our moral decisions.

Ethics involves critical reasoning about moral claims.

This includes not just what people believe, but whether those beliefs are justified.

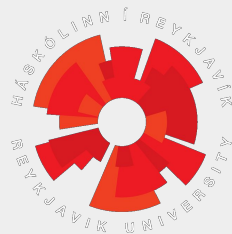


Concepts - Ethics

What isn't Ethics?

Ethics is not the same as:

- Feelings,
- Religion,
- Following the law,
- Following culturally accepted norms,
- Science.



Concepts - Ethics

Ethics vs. Morality:

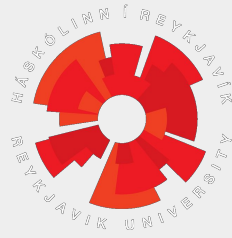
While morality refers to shared norms, intuitions, or cultural beliefs, ethics is the process of analyzing, questioning, and reasoning about those norms.

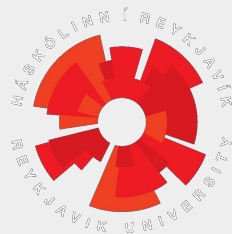
Ethics asks:

- Are these norms justified?
- What makes an action right or wrong?

Concepts - Moral Theory

What are Moral Theories?





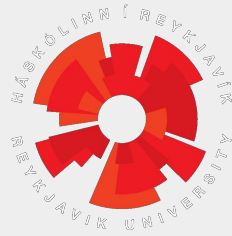
Concepts - Moral Theory

What are Moral Theories?

Moral theories aim to provide systematic answers to fundamental ethical questions:

- What makes actions right or wrong?
- What makes people good or bad?

They help us evaluate actions, people, institutions, and principles using consistent criteria.



Concepts - Moral Theory

Purpose of Moral Theories:

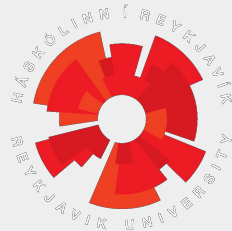
Provides a framework for ethical reasoning.

Helps us judge whether specific actions are morally acceptable, what obligations we have, and which values or principles should guide our decisions.

Role in Ethical Disagreement:

Moral theories help clarify the origins of disagreements.

Often, people don't just disagree about facts, they disagree about which moral principles matter most or how they should be balanced.



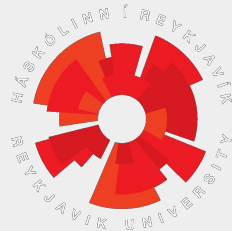
Concepts - Moral Theory

Why Moral Theories matter?

Without a theory, moral debate becomes inconsistent and purely emotional.

Moral theories offer a justificatory structure.

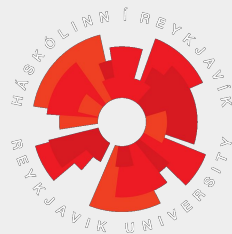
They show us how to move from intuition to reasoned judgement.



Basic Principles

Often, you will find a reference to the four basic principles of healthcare when evaluating the merits and difficulties of medical procedures. They are:

1. Autonomy
2. Justice
3. Beneficence
4. Non-Maleficence



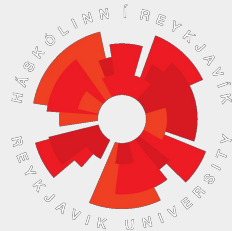
Basic Principles

Autonomy

Autonomy of thought, intention, and action when making decisions regarding procedures.

Therefore, the decision process must be free of coercion or coaxing.

Therefore, all risks and benefits, as well as likelihood of success, must be understood by the patient.

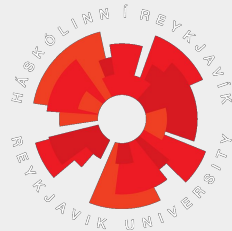


Basic Principles

Justice

Includes four main areas:

1. Fair distribution of scarce resources,
2. Competing needs,
3. Rights and obligations,
4. And potential conflicts with established legislation.



Basic Principles

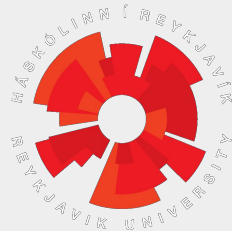
Beneficence

Intent of doing good for those involved.

Demands that providers maintain skills and knowledge, continually update training, consider individual circumstances, and strive for net benefit.

Non-Maleficence

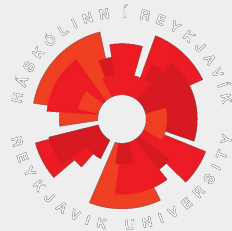
Do No Harm!



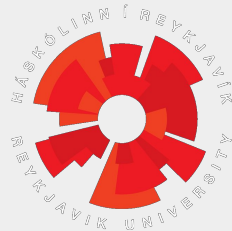
Basic Principles

We will talk more about basic principles of ethical decision making later.

This was just a short primer.



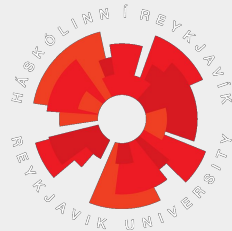
Questions so far?



Moral Theories

Today, we will look at the following moral theories:

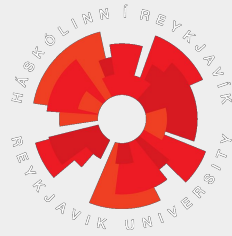
1. Moral Relativism (short intro)
2. Moral Objectivism
 - a. Consequentialism
 - i. Utilitarianism
 - ii. Virtue Ethics
 - b. Deontology
 - i. Categorical imperative



Moral Theories - Moral Relativism

What is morally right or wrong depends on the prevailing view in the society or culture we happen to be dealing with.

There is no objectively moral thing, action, or decision.

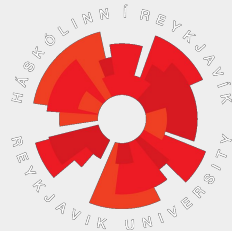


Moral Theories - Moral Objectivism

What is right or wrong doesn't depend on what anyone thinks is right or wrong.

Moral facts exist like “physical facts” independent of what anyone thinks they are.

They simply need to be discovered.



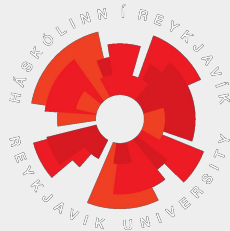
Moral Objectivism - Consequentialism

Whether an action is right or wrong depends on the action's consequences.

In any situation, the morally right thing to do is whatever will have the best consequences.

Sometimes called teleological theories.

It isn't very informative without a combination with a theory about what the best consequences are.

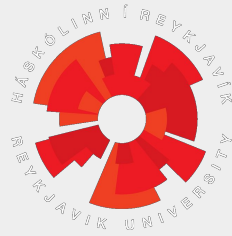


Consequentialism - Utilitarianism

Theory of best consequences.

Put forward by Jeremy Bentham (1748-1832) and improved by John Stuart Mill (1806-1873)

Most influential variety of consequentialism.



Consequentialism - Utilitarianism

Basis of Utilitarianism

Ask what has intrinsic value and assess the consequences of an action in terms of intrinsically valuable things

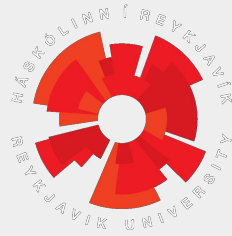
Instrumental and Intrinsic Value

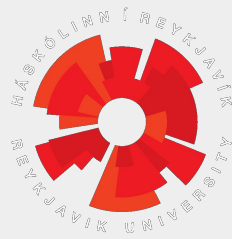
Instrumental Value: A thing has only instrumental value if it is only valuable for what it may get you (e.g., money)

Intrinsic Value: A thing has intrinsic value if you value it for itself (i.e., you would value it even if it brought you nothing else)

Consequentialism - Utilitarianism

What has Intrinsic Value?





Consequentialism - Utilitarianism

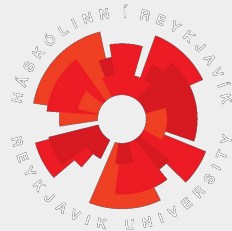
What has Intrinsic Value?

For utilitarians:

Only happiness has intrinsic value!

Therefore, the most ethical action is the one that produces the greatest overall happiness (or well-being) for the greatest number of people while minimizing pain or unhappiness for everyone affected by the choice.

Outcomes are prioritized over intentions.



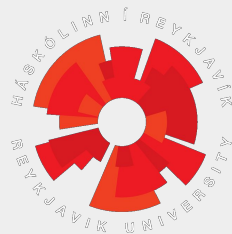
Consequentialism - Virtue Ethics

Actions according to (traditional) virtues lead to happiness.

Most widely discussed is Aristotle's account.

He believed humans had a specific function: To lead a life of true flourishing as a human.

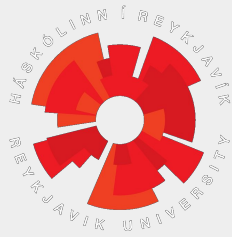
This requires abiding by the rules of rationality and acting with the traditional virtues.



Moral Objectivism - Deontology

- 'Duty-Based' Ethics
- Deny that what ultimately matters is an action's consequences.
- What matters is the kind of action it is - Doing our duty.
- Many kinds of deontological theory.
- The 'Golden Rule' is one of them: Do unto others as you'd have them do unto you.

Immanuel Kant (1724-1804) is the most influential deontologist.

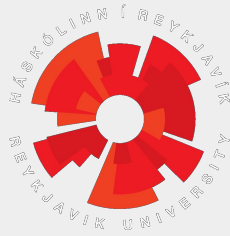


Moral Objectivism - Deontology

Rejecting consequentialism:

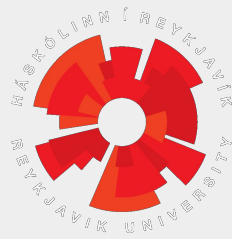
“A good will is good not because of what it effects or accomplishes.”

Even if, by bad luck, a good person never accomplishes anything much, the good will would “like a jewel, still shine by its own light as something which has its full value in itself.”



Moral Objectivism - Deontology

The Categorical Imperative?



Moral Objectivism - Deontology

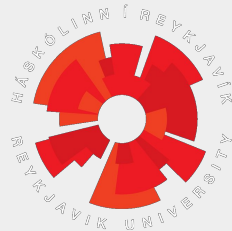
The Categorical Imperative:

First version:

Act only according to that maxim [i.e., rule] whereby you can at the same time will that it become a universal law.

Second version:

Act in such a way that you treat humanity, whether in your own person or in the person of another, always at the same time as an end and never simply as a means. → Important to treat people as autonomous agents.

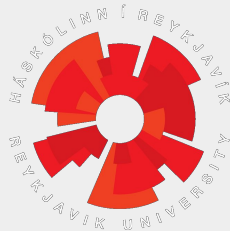


Moral Objectivism - Deontology

The Categorical Imperative:

Third Version (Not important here, just for completeness)

Every rational being must so act as if [they] were through his maxim always a lawmaking member in the universal kingdom of ends.



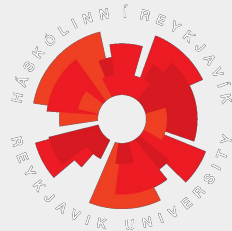
Moral Objectivism - Deontology

The Categorical Imperative:

Influential in arguments for human rights.

Informed consent.

Value the autonomy of individuals.



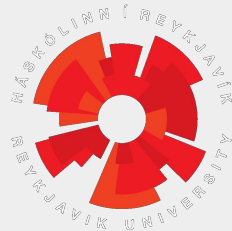
Moral Objectivism

Problems:

Deontology:

???

Consequentialism:



Moral Objectivism

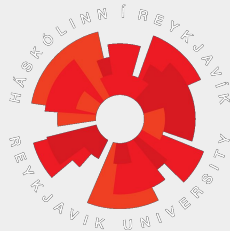
Problems:

Deontology:

What if doing your duty has repugnant consequences?

Consequentialism:

???



Moral Objectivism

Problems:

Deontology:

What if doing your duty has repugnant consequences?

Consequentialism:

What if you have to do something that seems wrong to produce the best consequences?