

KV Responsible AI

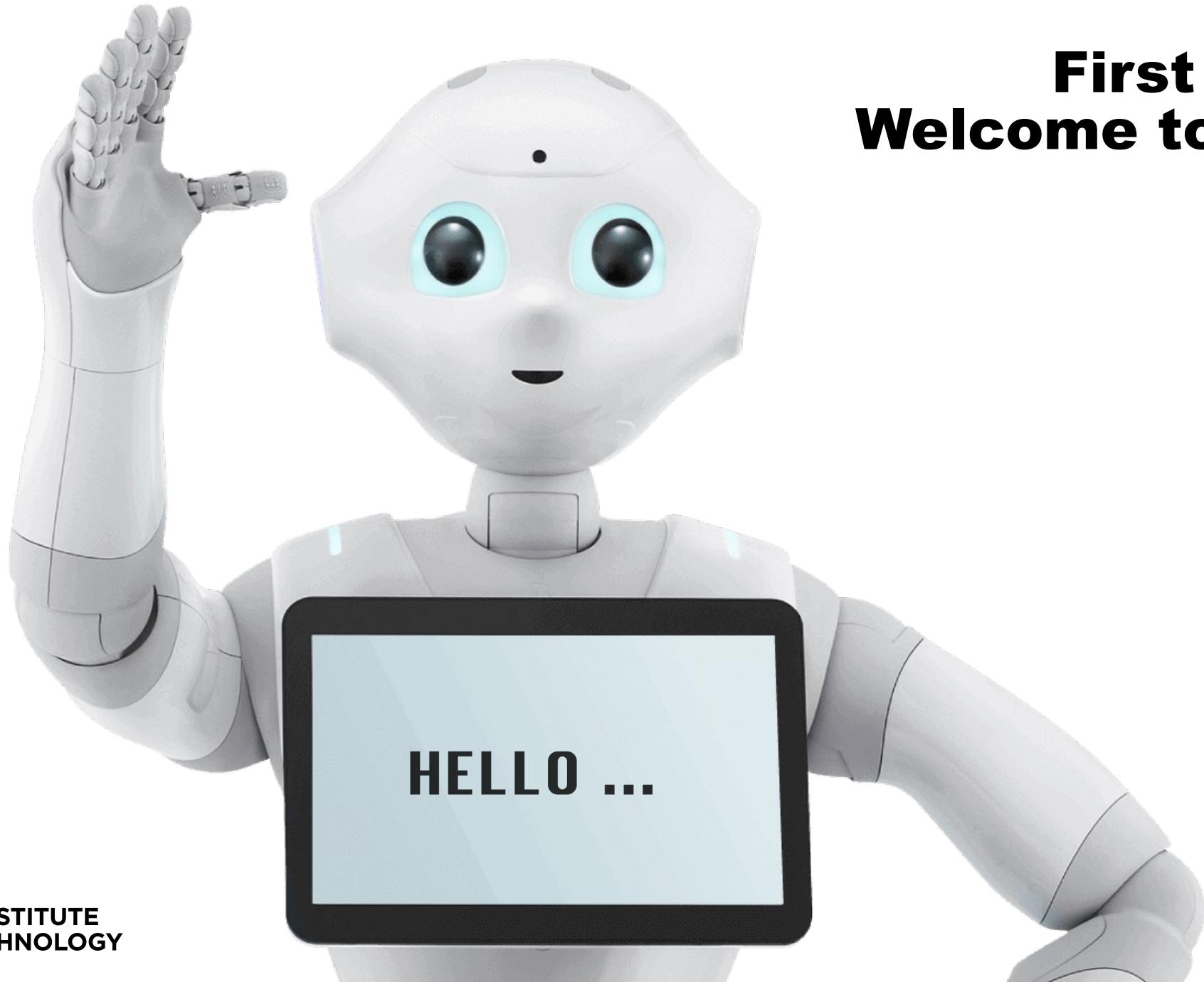


Prof. Dr. Martina Mara

**Introduction
10.10.2022**



**First of all:
Welcome to JKU!**



Short intro Martina Mara





Austrian Council on Robotics and Artificial Intelligence

Vienna, November 2018





Prof. Dr. Martina Mara,
Head of LIT Robopsychology Lab



| Schöne neue Welt

Der Lauschangriff des Klos

Vergangene Woche war es wieder einmal so weit. Stargate SG-1

... so weit: Die
weitgrößte Elektronik-Mes-
se CES läutete in Las Vegas das
neue Jahr der IT ein.

sende Aussteller präsentierten hunderttausenden Besuchern ihre Produktvisionen: Smarte Dinger zum Heruntertragen, smarte Riesenfernseher zum sich Berleben lassen, smarte Heime zum darin Wohnen. Relativ ...

alt sah dabei die CES-Organisation selbst aus: Der Anteil weiblicher Vortragender - Keynotes sind traditionell ein wichtiges Element der Messe - war umsmart und ziemlich gering. Das war an den anderen Messetagen auch ohne Taschenlampe zu sehen, was sehr schade war, machte aber sehr deutlich: Dass es sich um eine Messe handelt, die sich nicht auf Frauen ausgerichtet.

Messe - und wichtiger Teil der
Messe - lag bis kurz vor dem Start
bei null, dafür tanzten an den
Bühnen, die schnell
Sprachsteuerung. Geh es nach
den führenden Anbietern, sollen
wir künftig mit allem palavern,
was nicht Mund und Ohren, dafür

aber Mikrofon und eine Direktleitung zu einem Großkonzern besitzt. Der Siegeszug der Alexas, Siris und Cortanas - als Chatbots sind der Ingenieurswelt Damen dann offenbar doch reckenhaft.

... - wird
einsig vorbereitet. Die
Sprachassistentinnen
wurden bei der CES nicht mehr nur
in Form originärer Produkte wie
dem Echo Dot vorgeführt, sondern
integriert in unzähligen Alltagsge-
räten, darunter Badzimmern, Yes.
ger, Backofen oder Toiletten. Yes.
Das Klo, das lauscht und spricht.
Wer hat nicht darauf gewartet?
Endlich - Alexa!

"runter" sagen können und es auf die mittlarmäßige Antwort - Das habe ich leider nicht verstanden" - hin laut und deutlicher wiederholen. Da kommen schöne Szenen auf uns zu, in Privathäusern wie in öffentlichen Toilettenräumen. Verständnislosheit zwischen Beide wird sensorisch gesteuert. Später, im Einkaufszentrum, kann sich auch gleich verändert. Eine Renaissance der Menschlichkeit, erachtet durch Amazon.

Allerdings nicht zu verachten: Alle, die Endverbraucher sind,

Martina Mara ist Medienpsychologin und forscht am Ars Electronica Futurelab zur Mensch-Roboter-Beziehung. Twitter: @MartinaMara

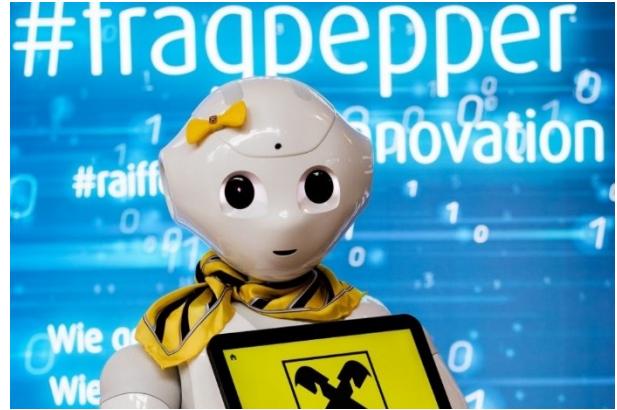
What is Robopsychology?

- Putting human experience and needs at the center of technological development
- Applying scientific methods and theories from Psychology to human interaction with Artificial Intelligence and Robotics
- Practical direction: Creating evidence-based guidelines for human-centered technology, collaborating with other disciplines & industry

Exemplary research questions:

- Why are people afraid of automation?
- How is system understandability related to trust?
- How can people collaborate well with algorithmic decision support systems?
- What does it take for robots not to undermine the human needs for autonomy, competence, and relatedness?

Exemplary research projects



Robots/chatbots in customer service



VR research game:
Human-robot communication



Acceptance of
exoskeletons at work



VR research game:
Trust in algorithmic decisions

Science communication: Demystify AI @ Ars Electronica 2021



Where you find me on campus



Course overview: Online tools we are going to use

- **KUSSS** (kusss.jku.at) for
 - Course registration
 - Exam registration
 - Final grades
 - Sometimes emails to students
- **MOODLE** (moodle.jku.at) for
 - Q&A Forum
 - Session live streams
 - Video recordings of sessions
 - Slides and other course materials (sometimes for self-study)
 - Group assignment for group task
 - Submission of mini tasks (individual) & group task (in groups)
 - Final exam

Let's check Moodle

510.101, KV Responsible AI, Martina Mara, 2022W

**Go to moodle.jku.at and
log in to the course 510.101**



Q & A Forum

You can use this forum to ask questions or initiate a discussion during or between the lectures.

**Anyone who has no access
to Moodle yet?**

Additional Material



Guide on how to cite (APA citation style)

Introduction (10.10.2022)

- Overview of topics and goals of this course
- Explanation of requirements and grading scheme
- Announcement of Mini Task 1

Course overview: Online tools we are going to use

- **MENTIMETER** (menti.com) for
 - Live polls during sessions

Let's try Mentimeter

**When you hear “Responsible AI”,
which terms come to your mind?**

Go to **menti.com** (use laptop or smartphone) and type in:

4821 4583

<https://www.menti.com/alvohd6xp17y>



What does responsibility mean?

Speaking of "Responsible AI" ... what does "**responsible**" actually mean?

to be responsible for sth

to have control and authority over something and the duty of taking care of it

responsible adjective

having good judgment and the ability to act correctly and make decisions on your own

- *a hard-working and responsible employee*
- *Let's stay calm and try to behave like responsible adults.*
- *Many big companies are now becoming more responsible about the way they operate.*

Source: Cambridge Dictionary

RBC & CSR

According to the OECD's understanding, **Responsible business conduct (RBC)** means that businesses

- a) should make a positive contribution to economic, environmental and social progress with a view to achieving sustainable development and
- b) should avoid and address adverse impacts through their own activities and prevent or mitigate adverse impacts directly linked to their operations, products or services by a business relationship. Risk-based due diligence is central to identifying, preventing and mitigating actual and potential adverse impacts, and thus is a key element of RBC.

(OECD Policy Framework for Investment, 2015)

According to the European Commission's understanding, **Corporate Social Responsibility (CSR)** is

the process whereby enterprises integrate social, environmental, ethical and human rights concerns into their core strategy, operations and integrated performance, in close collaboration with their stakeholders.

(EC Communication, 2011)

Responsible personality: Social Responsibility Scale

Psychologists began working on questionnaires to assess socially responsible personalities many years ago. Berkowitz & Luterman (1968) proposed the SRS as follows:

Answer each item with:

Strongly agree, Agree, Undecided, Disagree, or Strongly disagree

1. It is no use worrying about current events or public affairs; I can't do anything about them anyway.
2. Every person should give some of their time for the good of their town or country.
3. Our country would be a lot better off if we didn't have so many elections and people didn't have to vote so often.
4. Letting your friends down is not so bad because you can't do good all the time for everybody.
5. It is the duty of each person to do their job the very best they can.
6. People would be a lot better off if they could live far away from other people and never have to do anything for them.
7. At school I usually volunteered for special projects.
8. I feel very bad when I have failed to finish a job I promised I would do.

What does AI mean?

Speaking of "**Responsible AI**" ... what does "AI" actually mean?

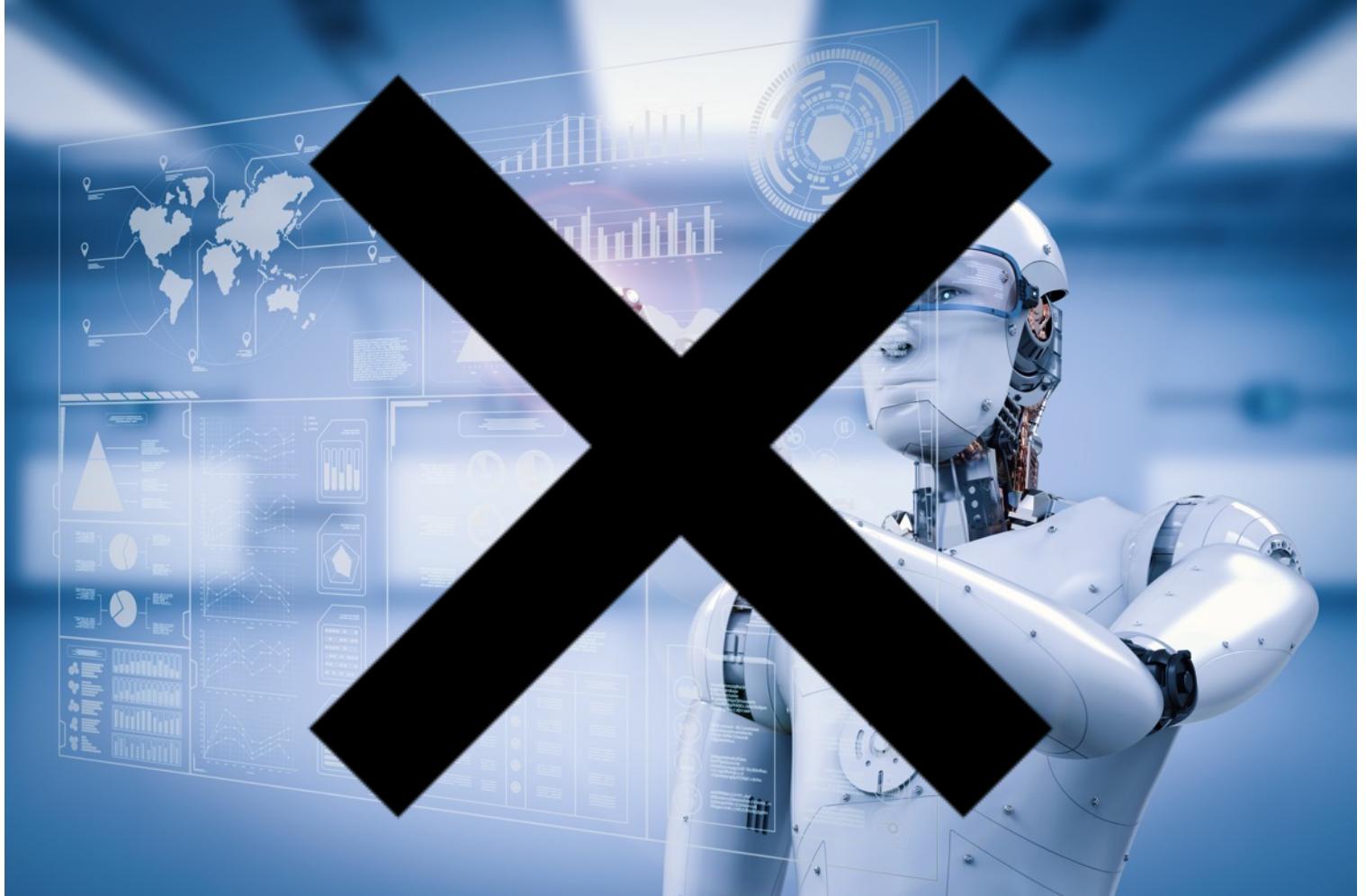
- Artificial intelligence (AI) refers to systems that display “intelligent” behavior by analyzing their environment and taking actions – with some degree of autonomy – to achieve specific goals.
- Many current AI technologies (probabilistic ones such as ML, NN, DL) use large datasets, recognize patterns in them and automate decision-making in a certain domain.

Some application examples:

- AI-based systems may include voice assistants, image recognition software, search engines, translation services or face recognition systems. We use AI on a daily basis, e.g. to translate languages, generate subtitles in videos or to block email spam.
- AI can also be embedded in hardware like autonomous cars, drones, or advanced industrial robots.

What does AI mean?

Speaking of "Responsible AI" ... what does "AI" actually mean?



AI is human-made!



Artificial Intelligence is designed by us:

- Which goals an AI system should target is defined by people.
- Where and how an AI system will be applied is defined by people.
- On which data an AI system is trained is defined by people.
- The data, from which we let AI learn, are often created by humans (a.o., past decisions, texts, images, videos)

Accordingly, **we humans are also responsible for shaping AI** in a way that

- respects fundamental human rights and ethical norms,
- maintains human autonomy and control,
- no one is being discriminated against,
- as many people as possible can benefit from AI.

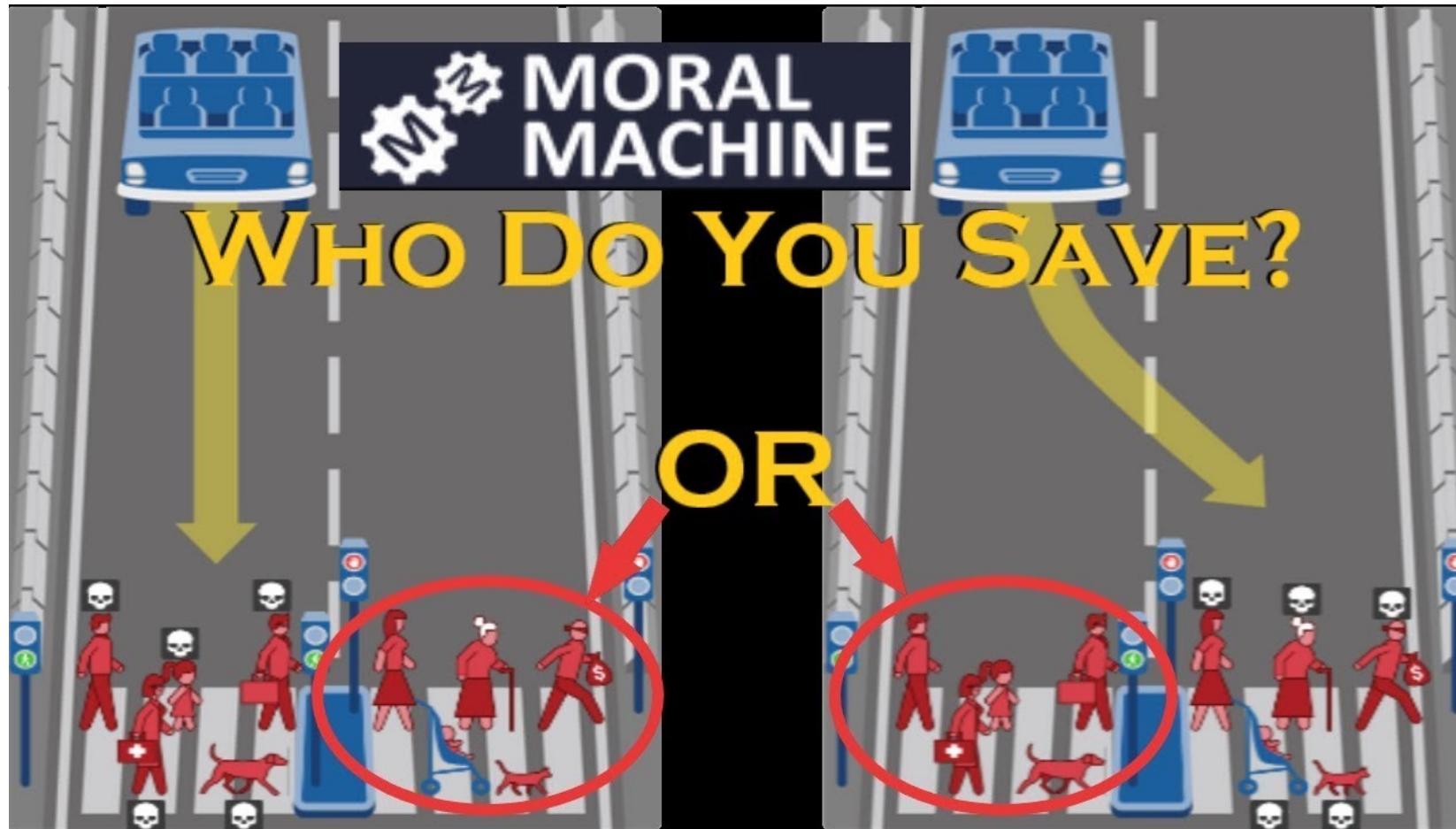
Main goals of this course

- Discussion of current ethical, social, and environmental questions of Artificial Intelligence
- Learn about state-of-the-art international guidelines for trustworthy/responsible AI
- Think about what you can consider as a (future) programmer to create responsible AI
- Learn about ethical decision-making and taking different perspectives on AI
- Discussion of positive future scenarios with AI ("AI for the good", "AI for all")

Planned topics of this course

- International Ethics Guidelines for AI
- Different ethical schools of thought
- AI and environmental wellbeing
- Fairness & Reproduction of human biases through AI
- Explainable AI
- Media depictions of AI
- AI and privacy
- AI and democracy
- Diversity and interdisciplinarity in teams that create AI

Quick teaser: Ethics in AI decision-making



Quick teaser: (Un)fair AI?

futuretense

When Alexa Can't Understand You

People with speech disabilities are being left out of the voice-assistant revolution.

By MOIRA CORCORAN OCT 16, 2018 • 7:30 AM

NEW SCIENTIST LIVE 2019
Tickets selling fast: book your place now!

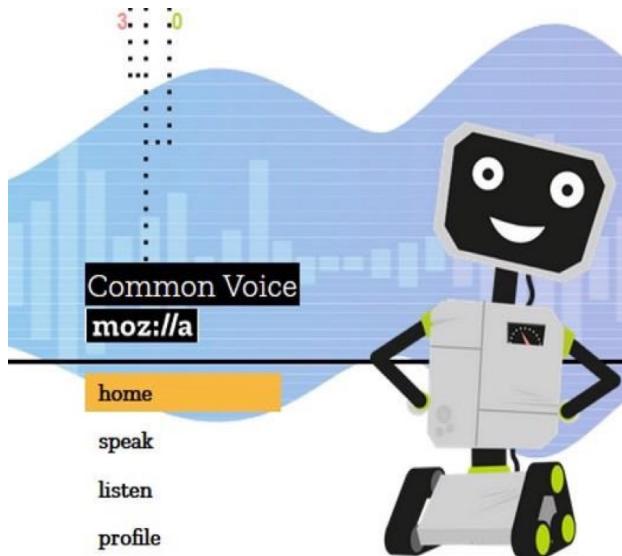
NewScientist

News Technology Space Physics Health Environment Mind Video | Tours Events Jobs

Donate your voice so Siri doesn't just work for white men

[f](#) [t](#) [e-mail](#) [w](#) [s](#) [in](#) [t](#) [+](#)

TECHNOLOGY 26 July 2017
By Matt Reynolds



Common Voice
moz://a

Mitarbeiter DATENSÄTZE SPRACHEN ÜBER UNS

0 0 DE Anmelden / Registrieren

Sprache Deutsch

GRÖSSE 14 GB

VERSION de_538h_2019-12-10

INSGESAMT BESTÄTTIGTE STUNDEN 483

GESAMTSTUNDEN 538

LIZENZ CC-0

ANZAHL DER STIMMEN 8.460

AUDIOFORMAT MP3

UNTERTEILUNGEN

Akzent

64% Deutschland Deutsch,
2% Österreichisches Deutsch,
...
Alter

23% 19 - 29, 19% 40 - 49,
17% 30 - 39, 11% 50 - 59,
3% < 19, 2% 60 - 69

Geschlecht

67% Männlich, 9% Weiblich

Course schedule

<u>Datum</u>	<u>Uhrzeit</u>	<u>Raum</u>	<u>Thema</u>
Mo. 10.10.2022	12:00 - 13:30	HS 7	Introduction
Mo. 17.10.2022	12:00 - 13:30	HS 7	
Mo. 24.10.2022	12:00 - 13:30	HS 7	
Mo. 07.11.2022	12:00 - 13:30	HS 7	
Mo. 14.11.2022	12:00 - 13:30	HS 7	
Mo. 21.11.2022	12:00 - 13:30	HS 7	
Mo. 28.11.2022	12:00 - 13:30	HS 7	
Mo. 05.12.2022	12:00 - 13:30	HS 7	
Mo. 12.12.2022	12:00 - 13:30	HS 7	
Mo. 09.01.2023	12:00 - 13:30	HS 7	
Mo. 16.01.2023	12:00 - 13:30	HS 7	
Mo. 23.01.2023	12:00 - 13:30	HS 7	
Mo. 30.01.2023	12:00 - 13:30	HS 7	Written exam

The course takes place weekly physically in HS7 and is streamed live to the lecture halls in Vienna and Bregenz.

For students who cannot come to one of the lecture halls e.g. due to health reasons, the course is also streamed via Zoom this year.

Everyone who is able is encouraged to be present during the course.

Course requirements (How to get a grade)

- **Mandatory elements**

- **Group exercise** (“dilemma discussions” in groups of 8, will be announced on 07.11.2022)
- **Written exam** (on **30.01.2023**, multiple choice questions + open questions)

- **Optional elements**

- **5 mini tasks** (individual submission)

Furthermore important:

- Check course content regularly and do not miss deadlines (e.g., self-assignment to group, task submissions)!
- Feel free to engage in conversations about the course topics, but always remain respectful!

Course requirements (How to get a grade)

Scoring scheme

Total points that can be achieved in the course: **50** (actually 51)

Group exercise: max. **20** points (at least 11 points must be achieved, otherwise repeat/fail)

Written exam: max. **26** points (at least 13 points must be achieved, otherwise repeat/fail)

5 Mini tasks: 1 point each (if completed within 2 weeks after announcement)

50 - 45 points = sehr gut (1)

< 45 points = gut (2)

< 39 points = befriedigend (3)

< 33 points = genügend (4)

< 25 points = nicht genügend (5)

If you hit the 51 points,
you'll get a fancy
unicorn sticker!



Use of AI for the good: More and more initiatives ...



AI for Good Global Summit

An **ITU** experience



Back to Mentimeter

What are major challenges faced by humanity today that AI could help with?

Go to **menti.com** (use laptop or smartphone) and type in:

4821 4583

<https://www.menti.com/alvohd6xp17y>



UN Sustainable Development Goals

<https://sdgs.un.org/goals>

The United Nations' Sustainable Development Goals were defined to achieve a better and more sustainable future for all.

They address the global challenges we face, including those related to poverty, inequality, climate, environmental degradation, prosperity, and peace and justice.

The Goals interconnect and in order to leave no one behind, it is important that each SDG is achieved by 2030.



The 17 SDGs: Short descriptions

SDG 1. End poverty in all its forms everywhere.

SDG 2. End hunger, achieve food security and improved nutrition, and promote sustainable agriculture.

SDG 3. Ensure healthy lives and promote wellbeing for all at all ages.

SDG 4. Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

SDG 5. Achieve gender equality and empower all women and girls.

SDG 6. Ensure availability and sustainable management of water and sanitation for all.

SDG 7. Ensure access to affordable, reliable, sustainable and modern energy for all.

SDG 8. Promote sustained, inclusive and sustainable economic growth, full and productive employment, and decent work for all.

SDG 9. Build resilient infrastructure, promote inclusive and sustainable industrialization, and foster innovation.

SDG 10. Reduce inequality within and among countries.

SDG 11. Make cities and human settlements inclusive, safe, resilient and sustainable.

SDG 12. Ensure sustainable consumption and production patterns.

SDG 13. Take urgent action to combat climate change and its impacts.

SDG 14. Conserve and sustainably use the oceans, seas and marine resources for sustainable development.

SDG 15. Protect terrestrial ecosystems, sustainably manage forests, combat desertification and halt and reverse land degradation, and halt biodiversity loss.

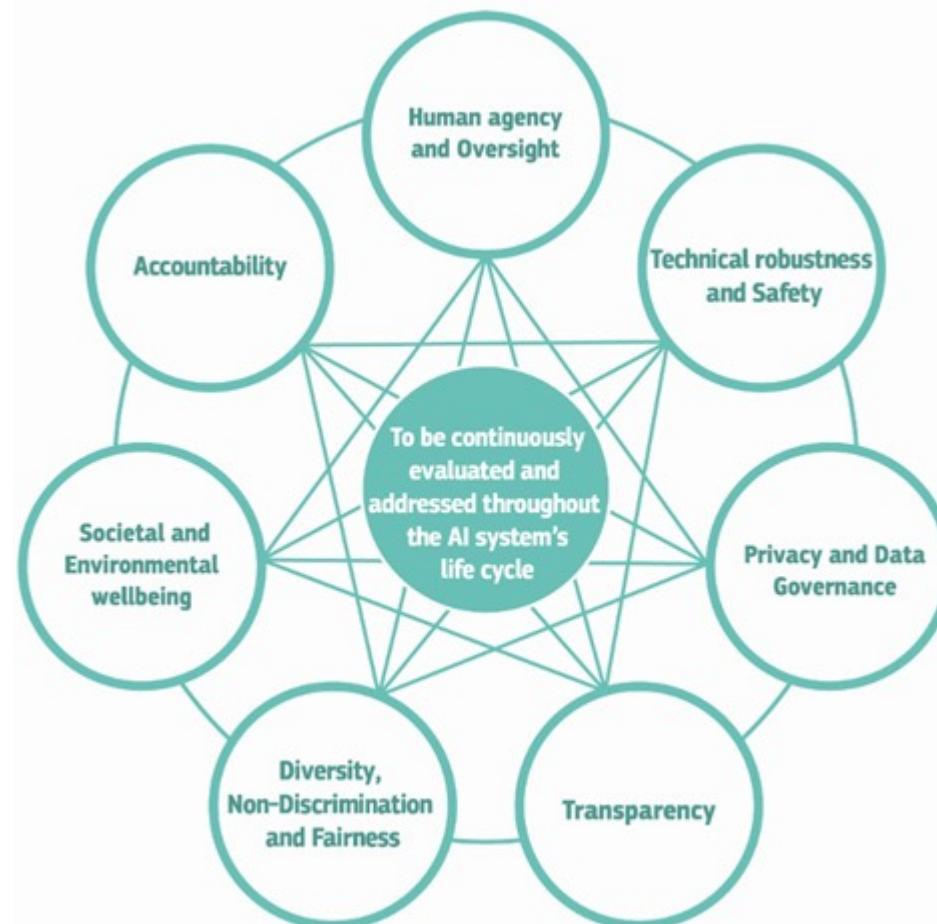
SDG 16. Promote peaceful and inclusive societies, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.

SDG 17. Strengthen the means of implementation and revitalize the global partnership for sustainable development.

Each SDG has its own specific targets, for a total of 169 targets in all.

EU Ethics Guidelines for Trustworthy AI

“Societal and environmental wellbeing” is also named one of seven key factors of “Trustworthy AI” in the European Union’s Ethics Guidelines published in 2019.



Mini task 1: Find an AI project that addresses one of the SDGs

- Do online research for projects/initiatives/companies/start-ups that use AI to target one or more of the 17 United Nations' Sustainable Development Goals (e.g. zero hunger, clean water, gender equality, clean energy).
- Select your favorite project and briefly describe it:
 - Project title & URL
 - Paragraph of text about how AI is used in this project
 - Paragraph of text about why this may help to achieve one of the SDGs
 - Paragraph of text why this project inspires/excites you personally
 - Do not plagiarize online texts but summarize in your own words!
- Submit your mini task via **Moodle** by **24.10.2022 (23:59)**
 - No file upload, but copy & paste your text into the online form

Questions?

KV Responsible AI



See you next week!

