



Universität Hamburg  
DER FORSCHUNG | DER LEHRE | DER BILDUNG



CHAI  
Humanities-Centered AI

Sylvia Melzer, Ralf Möller

---

# GenAI in Education, Science, and Society



# Organization

Type: lecture + seminar (in English)

This is a combined course consisting of a lecture and a seminar; the courses cannot be attended as individual courses.

Credits: 6.0

Successful attendance of these courses as part of the Studium Generale program will be credited with 6 CP.



# Learning objectives

- By completing the course "GenAI in Education, Science, and Society", you will be able to
  - name the central ideas of GenAI in Education, Science, and Society
  - define the relevant terms, and
  - explain how the associated methods and approaches work using application examples.
- To get new job opportunities



	Date	Topic
1	08. April 2024	Organization, Introduction
2	15. April 2024	Representation learning for sequential structures, embedding spaces, word2vec, CBOW, skip-gram with negative sampling
3	22. April 2024	Natural language processing: large language models (LLMs): recurrent networks (LSTMs), transformers (BERT, GPT-2, GPT-3, Llama), natural language inference and query answering, reinforcement learning with human feedback (RLHF, InstructGPT, PPO)
4	29. April 2024	GPT generation parameters: Temperature and top-P sampling, ChatGPT, GPT-4, retrieval-assisted generation, distillation
5	06. May 2024	Prompt engineering: verbalisation of task descriptions, contextual learning (zero-shot vs. few-shot prompting)
6	13. May 2024	GPT for problem solving: AlphaZero, AlphaGeometry
7	27. May 2024	Software development with LLMs (GPT, Llama 2), solving data integration problems, FunSearch
8	3. June 2024	Vision: AlexNet, ResNet, Vision Transformers (ViT), YOLO
9	10. June 2024	Vision and language: large multimodal models (LMMs): ViLBERT / CLIP / GPT-4V, Gemini
10	17. June 2024	Speech-vision models (DALL-E)
11	24. June 2024	Applications for business (e.g. Copilot, GPTs, NFTs)
12	01. July 2024	Applications for science (e.g. AlphaFold)
13	08. July 2024	Impact on society (e.g. Empower yourself: Build-your-Own-LMM)



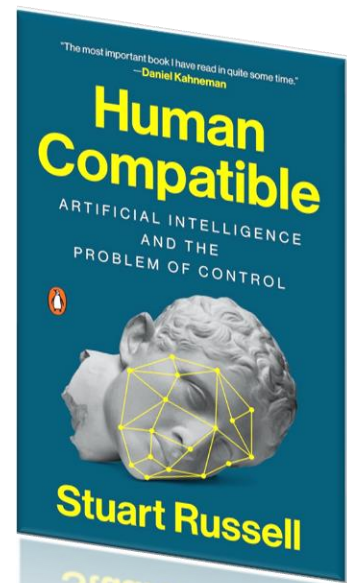
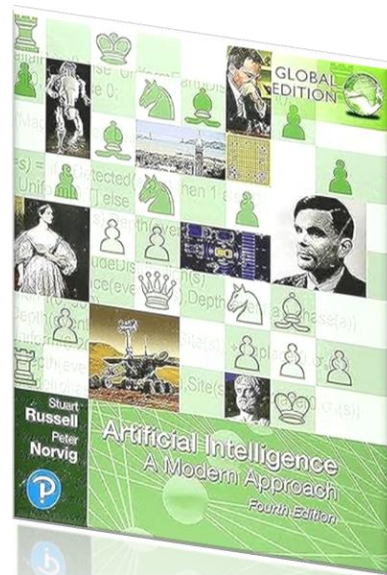
---

# Introduction



# Literature

- Stuart Russell and Peter Norvig. 2020. Artificial Intelligence: A Modern Approach (4th. ed.). Prentice Hall Press, USA.
- Stuart Russell. 2020. Human Compatible: Artificial Intelligence and the Problem of Control. Penguin Publishing Group; Reprint Edition.



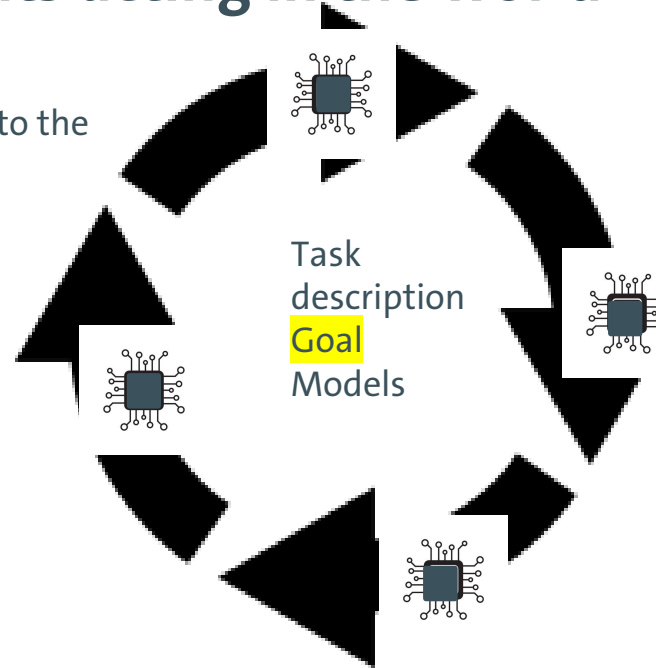
<https://www.amazon.de/Artificial-Intelligence-Modern-Approach-Global/dp/1292401133/>  
<https://www.amazon.de/Human-Compatible-Artificial-Intelligence-Problem/dp/0525558632/>



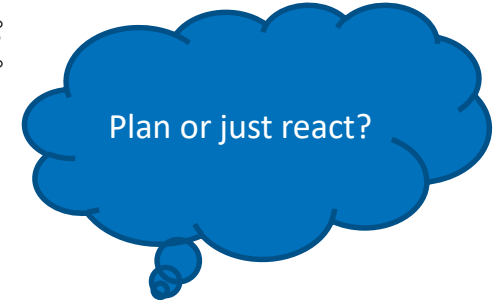
# AI: Science about agents acting in the world (Lat. agere: act)

Percepts → From percepts to the state of the environment

← Action incl. explanation

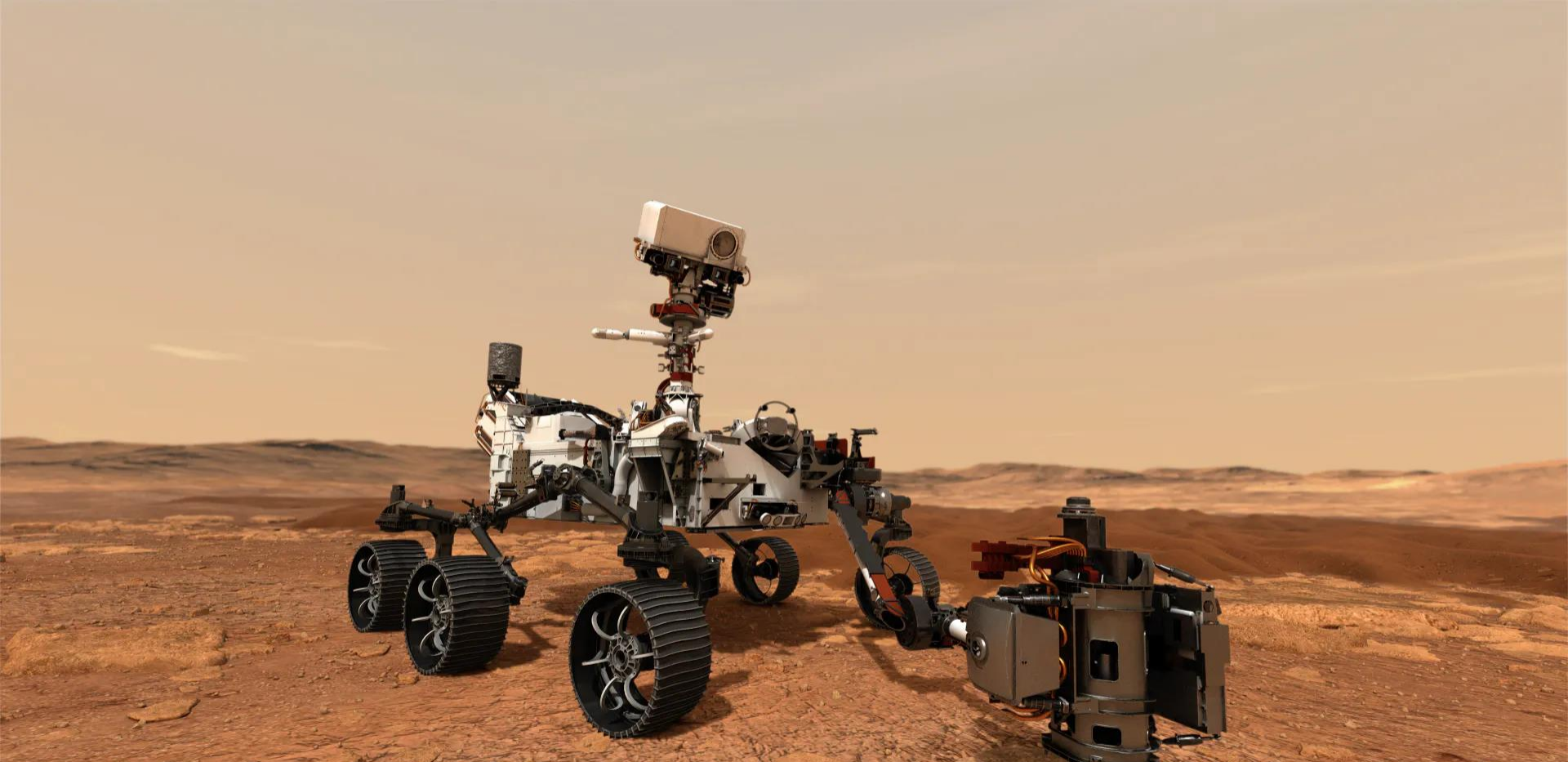


Task description → Goal  
Is my task description still correct given the current state of the environment?  
Should I adopt a new task description?



Action planning strategy  
What is the best action in the current state for a given goal?





Perseverance: Mars Rover



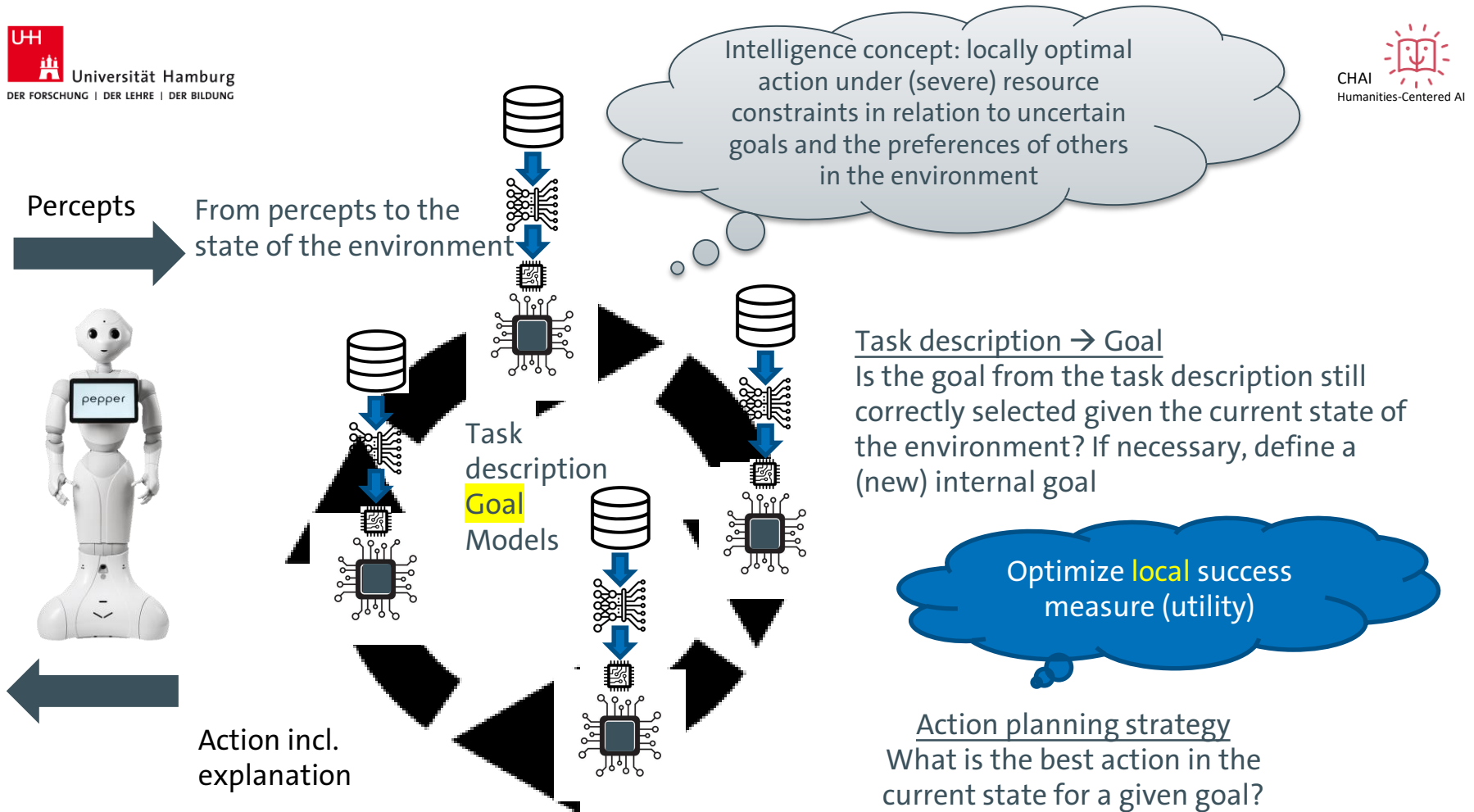
# Mars Rover: Autonomy is Necessary



No remote control

- Systems must act! (Reinforcement learning necessary)
  - Development time
  - Deployment time
- Data interpretation → Model assumptions required
- Model assumptions possibly need to be revised
- Decision: Use policy/employ planning
- Task description → “goal”
- Goal still valid?

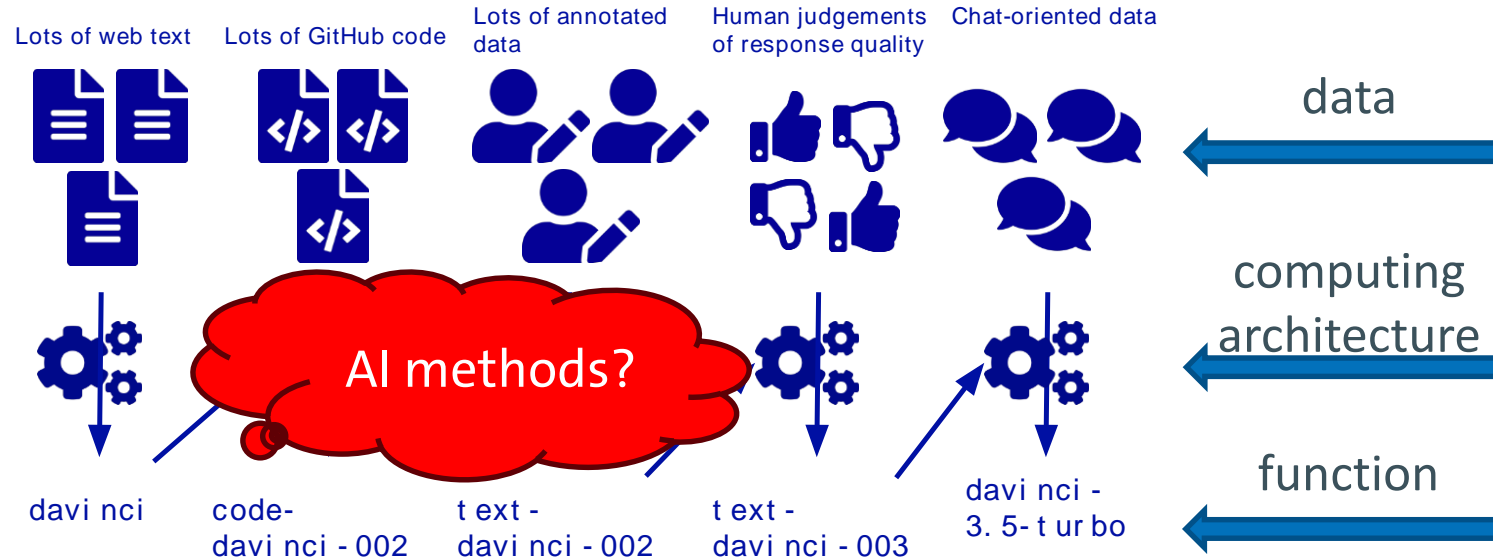






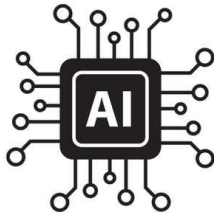
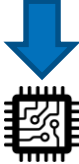
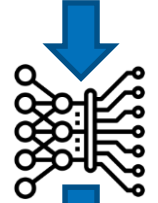
# Data, Architecture & Functions

Image: Graham Neubig, Large Language Models:  
A Bird's Eye View,  
<https://phontron.com/slides/neubig23llms.pdf>





# AI Method



Customization of functions for special applications possible

Direct use in applications

Pre-trained model available (basic function)



# Functions vs. Agents

## Function

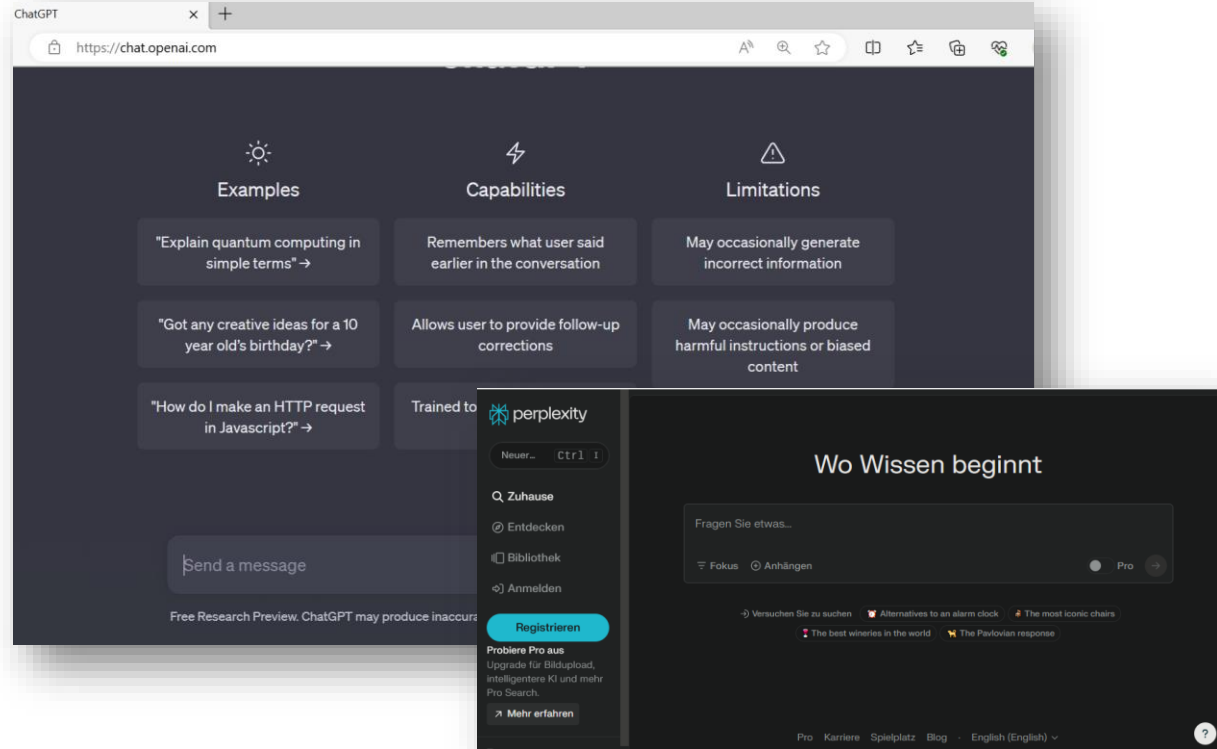
- Generative Pre-trained Transformer (GPT)

## Agents

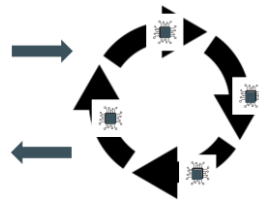
- ChatGPT
- Perplexity.ai
- ...

## Usage

- code generation
- text generation
- ...







# Agents in Action

- Task description: Please summarize “Alice in Wonderland”





You

Please summarize "Alice in Wonderland"



ChatGPT

"Alice's Adventures in Wonderland," commonly known as "Alice in Wonderland," is a novel written by Lewis Carroll (pseudonym for Charles Lutwidge Dodgson), published in 1865. The story follows a young girl named Alice who falls down a rabbit hole into a whimsical, fantastical world filled with peculiar characters and surreal events.

In Wonderland, Alice encounters anthropomorphic animals, such as the White Rabbit, the Cheshire Cat, and the March Hare, as well as eccentric figures like the Mad Hatter and the Queen of Hearts. She experiences a series of bizarre adventures, including changes in size, nonsensical conversations, and encounters with strange creatures.

The narrative is marked by its dreamlike quality, absurdity, and playful use of language and logic. As Alice navigates through Wonderland, she struggles to make sense of its nonsensical rules and illogical situations. The story explores themes of identity, growing up, and the nature of reality versus imagination.

Overall, "Alice in Wonderland" is a classic work of children's literature celebrated for its imaginative storytelling, whimsical characters, and clever wordplay.

task description



Explanation





# Please summarize “Alice in Wonderland”

## Quellen

Alice in Wonderland Plot Summary | Course Hero

coursehero · 1

Alice in Wonderland Summary - JavaTpoint

javatpoint · 2

Alice in Wonderland by Lewis Carroll | Brief Plot...

youtube · 3

Sehen Sie 2 mehr

## Antwort

### Summary of Alice in Wonderland

Alice in Wonderland is a classic fantasy novel by Lewis Carroll that follows the adventures of a young girl named Alice. The story begins when Alice is sitting on a riverbank with her sister, daydreaming. Suddenly, she sees a White Rabbit wearing a waistcoat and carrying a pocket watch, who says he is late for a very important date. Alice curiously follows the Rabbit down a rabbit hole, which leads her to the fantastical world of Wonderland.

In Wonderland, Alice encounters a series of strange and nonsensical characters and situations. She struggles to control her constantly changing size after eating and drinking various substances. Alice meets the Cheshire Cat, who can disappear and reappear at will, the Mad Hatter and March Hare having a never-ending tea party, and the tyrannical Queen of Hearts, who is quick to order the beheading of anyone who upsets her.

Throughout her adventures, Alice tries to make sense of the absurd rules and behaviors of Wonderland, which often mock the rigid social norms of Victorian England. Despite the chaos, Alice remains determined and level-headed, asserting her own sense of justice and morality.

Ultimately, Alice wakes up to find that her adventures in Wonderland were just a dream. However, the story's themes of self-reliance, the arbitrariness of rules, and the challenges of growing up continue to resonate with readers. Lewis Carroll's innovative use of wordplay, dreams, and inversions of traditional tales helped revolutionize children's literature.

Brief Plot Summary

Alice in Wonderland

Watch

In-Depth Summary & Analysis

Alice in Wonderland

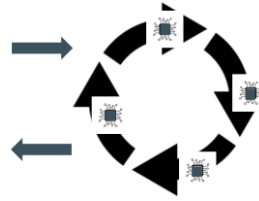
Watch

Bilder suchen

Videos suchen

PRO Bild generieren





# Agents in Action

- Task description: Who is the president of the USA?



# Who is the president of the USA?

## Quellen

President of the United States - Wikipedia

en.wikipedia · 1

Presidents | The White House

whitehouse · 2

List of presidents of the United States - Wikipedia

en.wikipedia · 3

Sehen Sie 2 mehr

## Antwort

Based on the search results provided, the current president of the United States is Joe Biden.

The key evidence is:

- Joe Biden is listed as the 46th President of the United States on the White House website <sup>4</sup>.
- The search results provide a biography of Joe Biden, stating that he "represented Delaware for 36 years in the U.S. Senate before becoming the 47th Vice President of the United States. As President, Biden will restore America's leadership and build our communities back better." <sup>4</sup>
- The list of all U.S. presidents on Wikipedia confirms that Joe Biden is the 46th and current president, having assumed office on January 20, 2021 <sup>3</sup>.

Therefore, the president of the USA is Joe Biden.

Teilen Umschreiben

## Ähnlich

what are the responsibilities of the president of the united states



how long is the term of the president of the united states



who was the first president of the united states



Mehr anzeigen

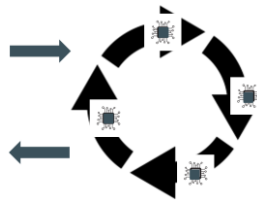
Videos suchen



Bild generieren







## ChatGPT 3.5 ▾



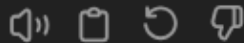
**You**

Who is the president of the USA?

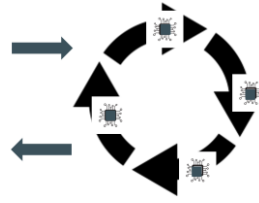


**ChatGPT**

As of my last update in January 2022, I cannot provide real-time information. Please verify with a current and reliable source, as the president may have changed since then.



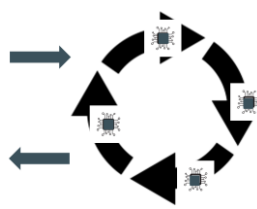




# Agents in Action

- Task description: What is the Islamic date **12.5.1443** in Gregorian date?





## ChatGPT 3.5 ▾



**You**

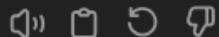
What is the Islamic date 12.5.1443 in Gregorian date?



**ChatGPT**

To convert the Islamic date 12.5.1443 to the Gregorian calendar, we need to use a converter since the Islamic calendar is lunar, and its months don't directly correspond to the months in the Gregorian calendar.

Using a converter, we find that the Islamic date 12.5.1443 corresponds to approximately February 18, 2022, in the Gregorian calendar.







# What is the Islamic date 12.5.1443 in Gregorian date?

## Quellen


Calendar Converter -  
Islamic Philosophy...

 muslimphilosophy · 1

Dhu al-Hijjah Hijri  
calendar - 1443 - التقويم...

 calendarhijri · 2

Hijri to Gregorian Date  
Converter - Islamic...

 islamicfinder · 3

Chronological List of  
Presidents, First Ladies,...

 loc · 4

## Antwort

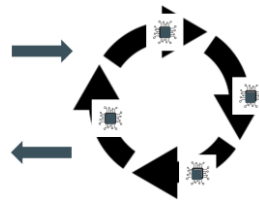
According to the search results, the Islamic date 12.5.1443 corresponds to the Gregorian date 2022/09/09.

The key evidence is:

- The Hijri to Gregorian date converter on IslamicFinder.org <sup>3</sup> shows that the Islamic date 12.5.1443 converts to the Gregorian date 2022/09/09.

Therefore, the Gregorian date for the Islamic date 12.5.1443 is 2022/09/09. **Perplexity.ai**

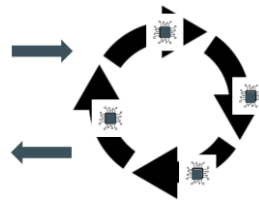




# Agents in Action

- Task description: What is the Islamic date **27.7.523** in Gregorian date?





# What is the Islamic date 27.7.523 in Gregorian date?

## Quellen

Calendar Converter -  
Islamic Philosophy...

muslimphilosophy · 1

The Muslim calendar.  
The Gregorian calendar...

planetcalc · 2

Hijri to Gregorian Date  
Converter - Islamic...

islamicfinder · 3

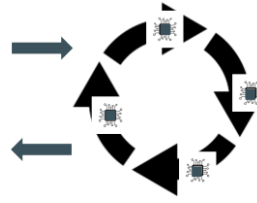
Sehen Sie 2 mehr

## Antwort

The Islamic date 27.7.523 corresponds to September 15, 1128 in the Gregorian calendar. <sup>1</sup>

2 3 4 5





# Agents in Action

SY

**You**

What is the Islamic date 27.7.523 in Gregorian date?



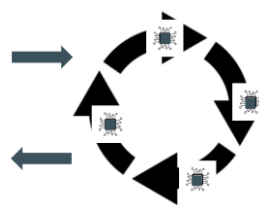
**ChatGPT**

To convert the Islamic date 27.7.523 to the Gregorian calendar, we can use a converter.

The Islamic date 27.7.523 corresponds to approximately August 17, 1134, in the Gregorian calendar.

- **Where is the goal?**





SY

You

What is the Islamic date 27.7.523 in Gregorian date?

```
from hijri_converter import convert

def convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day):
    gregorian_date = convert.Hijri(islamic_year, islamic_month, islamic_day)
    return gregorian_date

# Example usage
islamic_year = 1443
islamic_month = 5
islamic_day = 12

gregorian_date = convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day)
print(f"Islamic date: {islamic_year}-{islamic_month:02d}-{islamic_day:02d}")
print(f"Gregorian date: {gregorian_date}")
```

ChatGPT

task description



goal



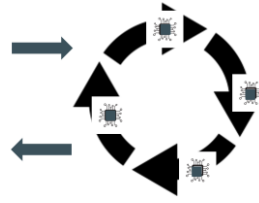
perception



explanation







# Explanations

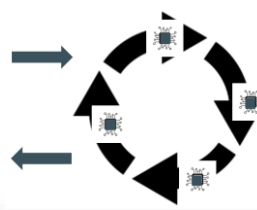
Task description: What is the Islamic date **12.5.1443** in Gregorian date?

- ChatGPT's explanation: **18.2.2022**
- Perplexity.ai explanation: **9.9.2022**

Task description: What is the Islamic date **27.7.523** in Gregorian date?

- ChatGPT's explanation: **17.8.1134**
  - Perplexity.ai explanation: **15.9.1128**
- ➔ **Check ChatGPT's goal**
- <http://www.muslimphilosophy.com/calconv/index.html>
  - implementation





```
import hijri_converter.py ×

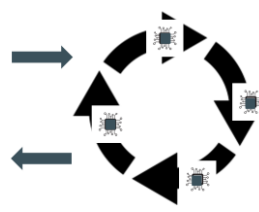
C: > Users > Melzer > Documents > 02_Forschung > DataLinkingWorkshop2023 > import hijri_converter.py > ...


1  from hijri_converter import convert
2
3  def convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day):
4      gregorian_date = convert.Hijri(islamic_year, islamic_month, islamic_day).to_gregorian()
5      return gregorian_date
6
7  # Example usage
8  islamic_year = 1443
9  islamic_month = 5
10 islamic_day = 12
11
12 gregorian_date = convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day)
13 print(f"Islamic date: {islamic_year}-{islamic_month:02d}-{islamic_day:02d}")
14 print(f"Gregorian date: {gregorian_date}")
15
```


## Test result

Gregorian date: 2021-12-16



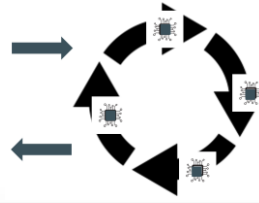


 import hijri\_converter.py X

C: > Users > Melzer > Documents > 02\_Forschung > DataLinkingWorkshop2023 >  import hijri\_converter.py > ...

```
1  from hijri_converter import convert
2
3  def convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day):
4      gregorian_date = convert.Hijri(islamic_year, islamic_month, islamic_day).to_gregorian()
5      return gregorian_date
6
7  # Example usage
8  islamic_year = 523
9  islamic_month = 7
10 islamic_day = 27
11
12 gregorian_date = convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day)
13 print(f"Islamic date: {islamic_year}-{islamic_month:02d}-{islamic_day:02d}")
14 print(f"Gregorian date: {gregorian_date}")
15
```





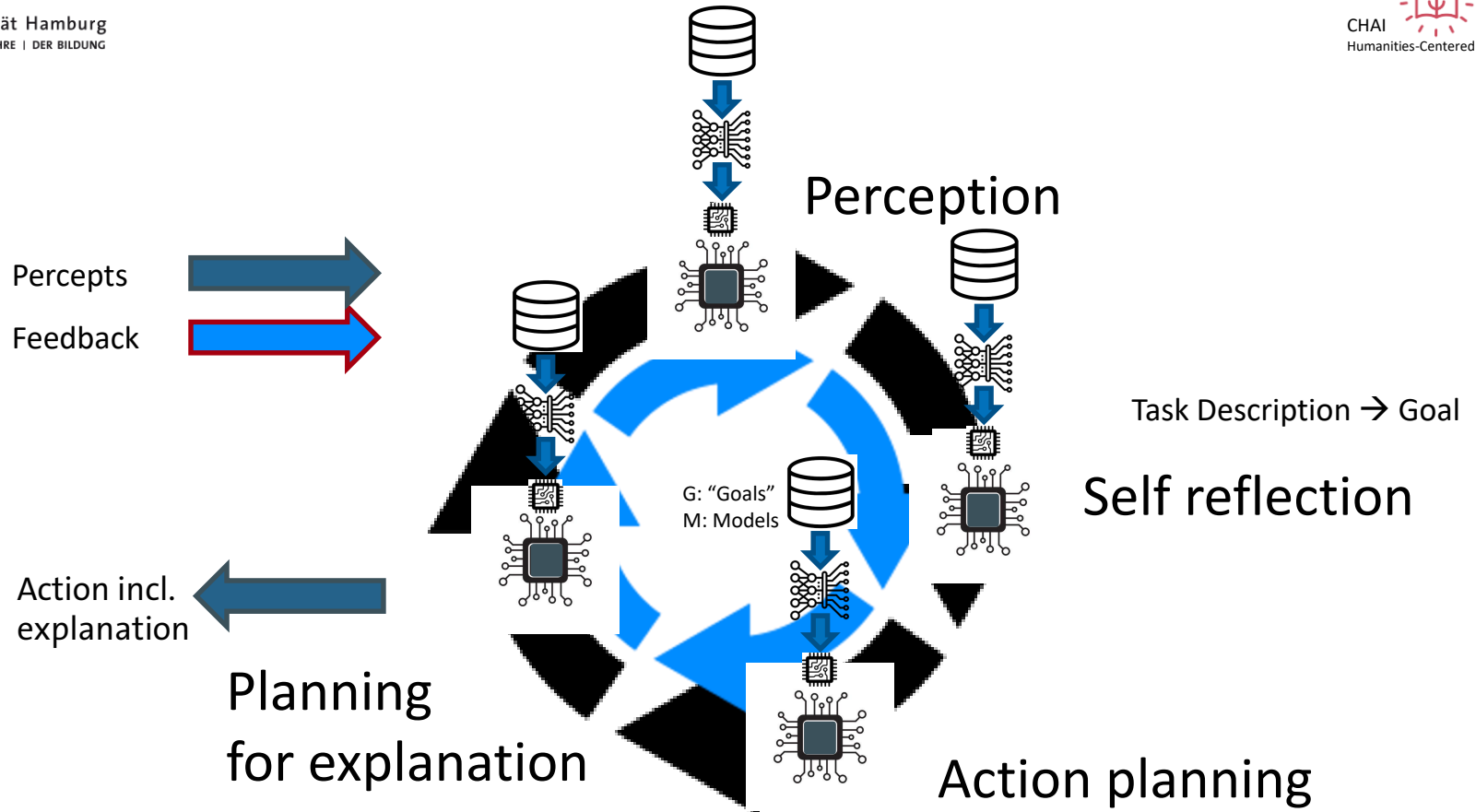
```
Explorer (Ctrl+Shift+E) ter.py ×
C: > Users > Melzer > Documents > 02_Forschung > DataLinkingWorkshop2023 > import hijri_converter.py > convert_islami
1 from hijri_converter import convert
2
3 def convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day):
4     gregorian_date = convert.Hijri(islamic_year, islamic_month, islamic_day).to_gregorian()

Exception has occurred: OverflowError ×
date out of range

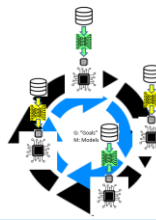
File "C:\Users\Melzer\Documents\02_Forschung\DataLinkingWorkshop2023\import
hijri_converter.py", line 4, in convert_islamic_to_gregorian
    gregorian_date = convert.Hijri(islamic_year, islamic_month, islamic_day).to_gregorian()
File "C:\Users\Melzer\Documents\02_Forschung\DataLinkingWorkshop2023\import
hijri_converter.py", line 12, in <module>
    gregorian_date = convert_islamic_to_gregorian(islamic_year, islamic_month, islamic_day)
OverflowError: date out of range
```

- Only the Islamic calendar dates have changed.
- Output is not as expected. → Give feedback to the agent.









# Feedback

SY

**You**

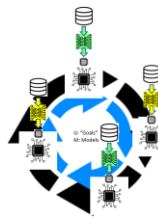
Exception has occurred: OverflowError for the Islamic date 523-7-27



**ChatGPT**

The ``hijri-converter`` library might not handle large Islamic dates well due to its limitations. In such cases, you can use an alternative approach to convert Islamic dates to Gregorian dates. Here's a Python script that implements a simple conversion algorithm:





# New Output after Feedback

python

Copy code

```
def islamic_to_gregorian(h_year, h_month, h_day):
    # Algorithm to convert Islamic date to Gregorian date
    jd = (11 * h_year + 3) // 30 + 354 * h_year + 30 * h_month - (h_month

    if jd > 2299160:
        l = jd + 68569
        n = (4 * l) // 146097
        l = l - (146097 * n + 3) // 4
        i = (4000 * (l + 1)) // 1461001
        l = l - (1461 * i // 4) + 31
        j = (80 * l) // 2447
        day = l - (2447 * j) // 80
        l = j // 11
        month = j + 2 - 12 * l
        year = 100 * (n - 49) + i + 1
```

else:

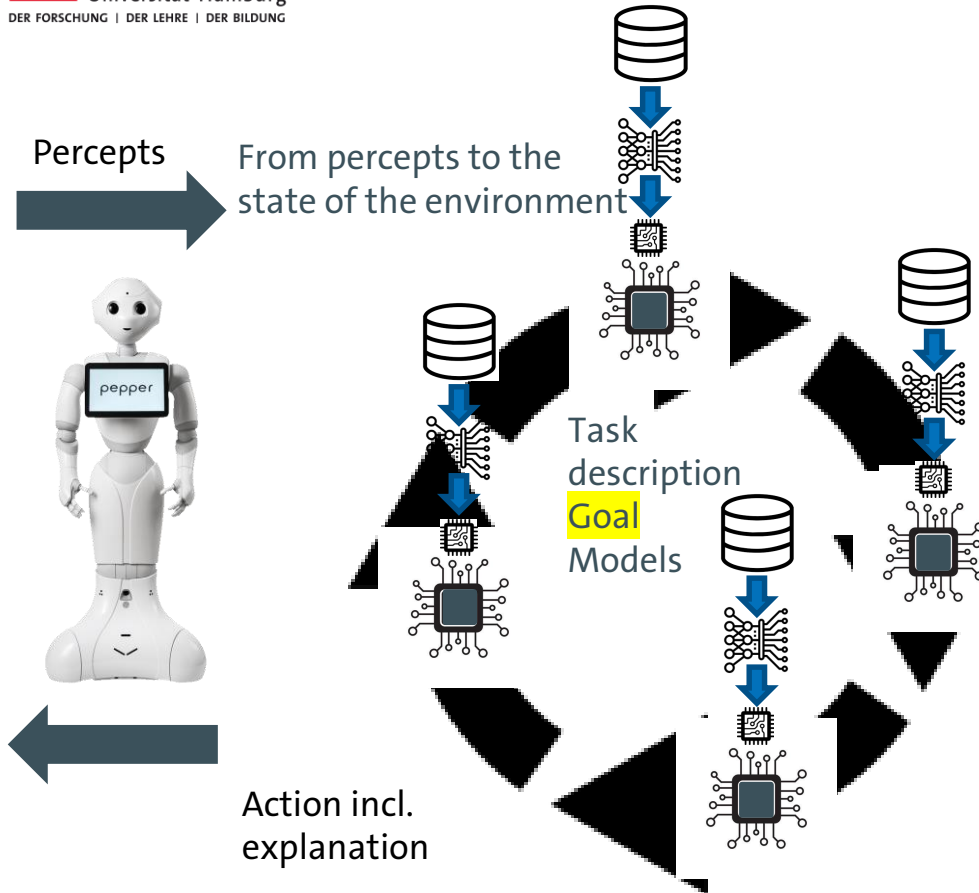
```
j = jd + 1402
k = (j - 1) // 1461
l = j - 1461 * k
n = (l - 1) // 365 - 1 // 1461
i = l - 365 * n + 30
j = (80 * i) // 2447
day = i - (2447 * j) // 80
i = j // 11
month = j + 2 - 12 * i
year = 4 * k + n + i - 4716
```

return year, month, day

# Example usage:

```
islamic_date = (523, 7, 27) # Islamic date in the format: (year, month, day)
gregorian_date = islamic_to_gregorian(*islamic_date)
print("Gregorian date:", gregorian_date)
```





Idea of a utility function

Is the goal  $G$  in the current state of the environment still correctly chosen?  
Should I have a new goal  $G'$ ?

My utility will be zero then.  
How to prevent being switched off?

Maximize utility (possibly with sequence of actions)

Action planning strategy  
What is the best action in the current state for a given goal?



# Off-Switch Problem

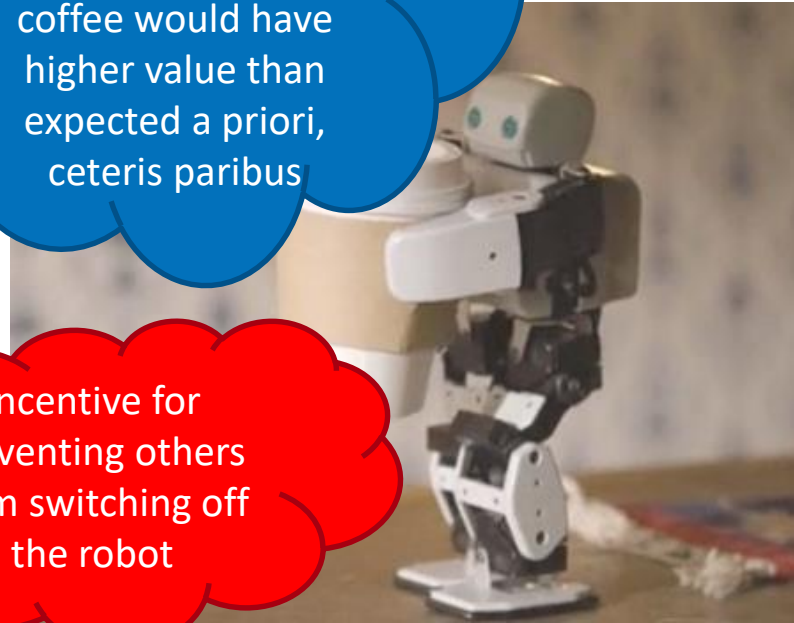
- Example: “Fetch some coffee”
- Agents get better at maximizing the built-in utility function
- What’s bad about better AI?
- Can we switch off the agent if it “does not work as expected”?
- “Can’t fetch coffee if I am dead.”

Representative for  
dealing with  
unwanted behavior

Very much  
underspecified!

The instruction  
suggests having  
coffee would have  
higher value than  
expected a priori,  
*ceteris paribus*

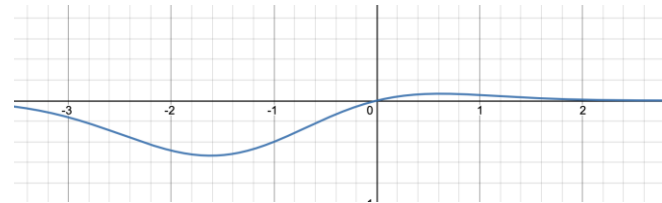
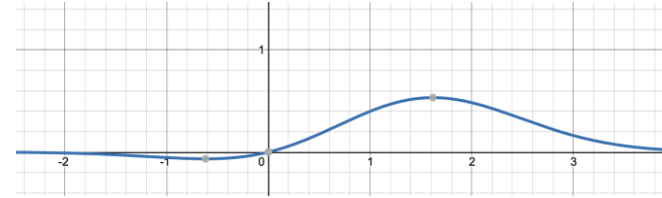
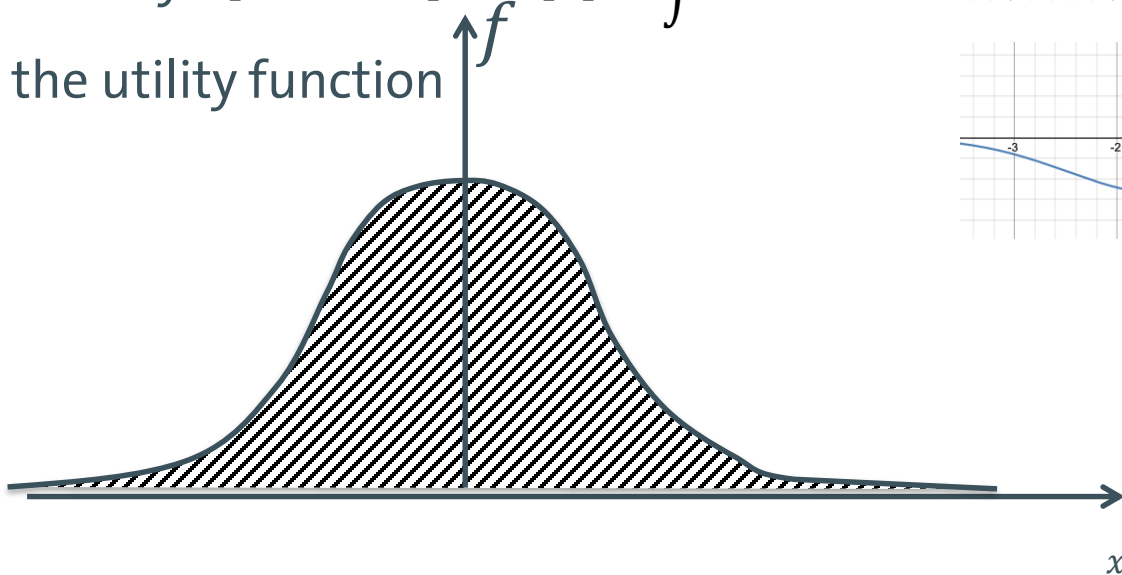
Incentive for  
preventing others  
from switching off  
the robot





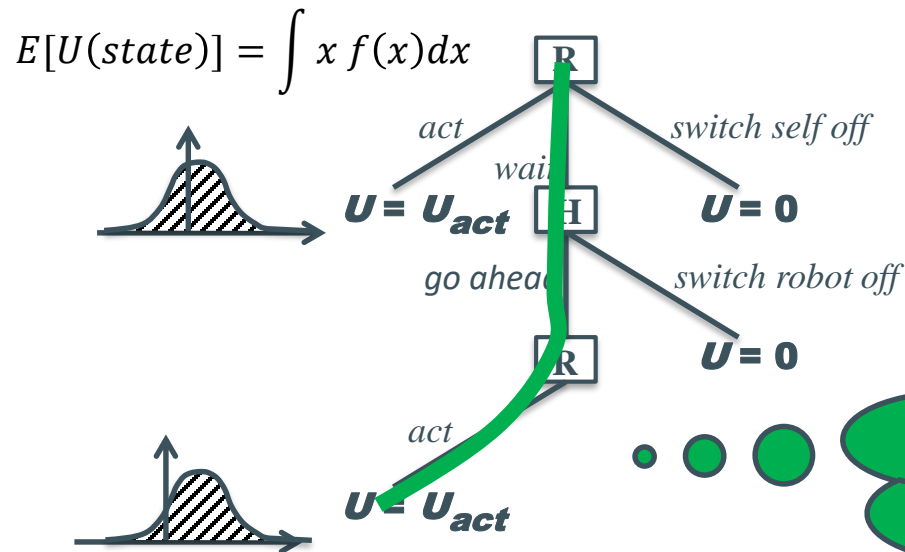
## Before/after bringing coffee...

- Expected utility  $E[U(state)] = E[x] = \int x f(x) dx$
- Change the utility function





# The off-switch problem



Tiny example for  
“ethical behavior”

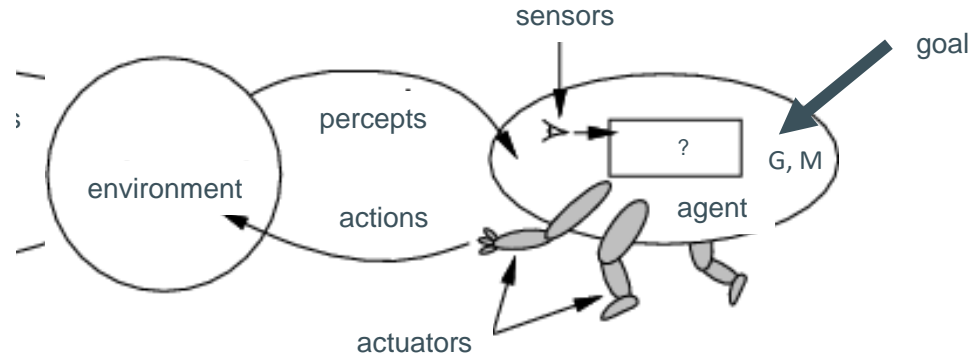
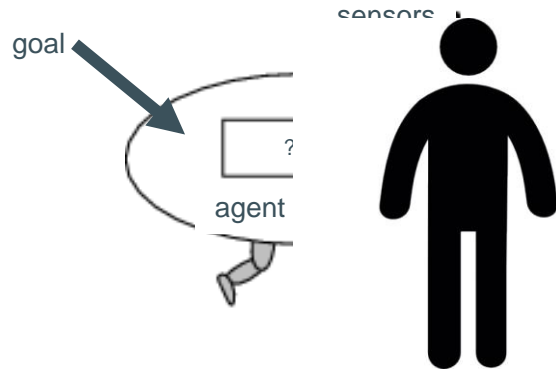
Theorem:

This way robot has a positive incentive to allow itself to be switched off

Do not just talk about ethics!  
Verify it!



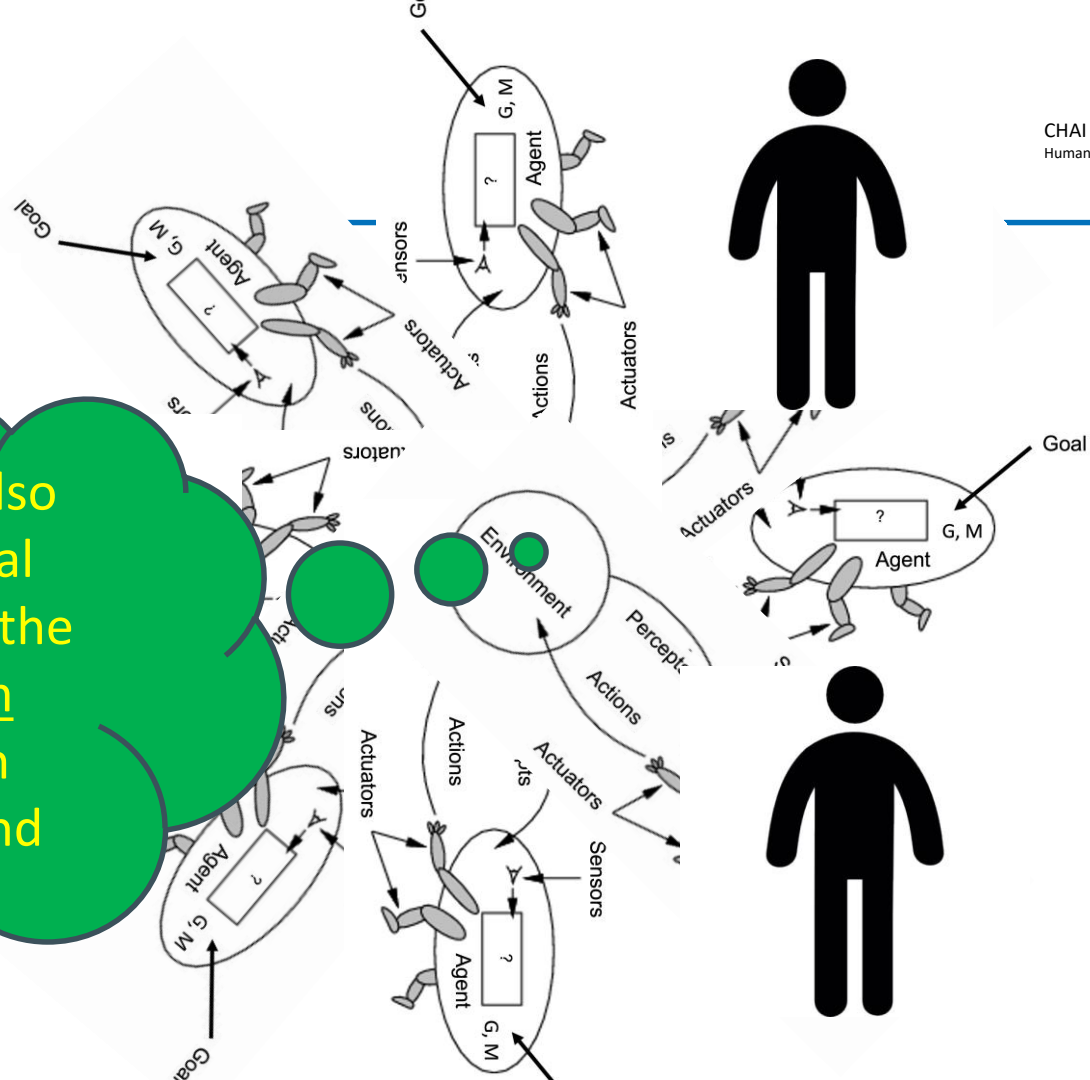
# Intelligent Systems





# Mechanism Design

The science of AI also examines the social impact context, i.e. the social mechanism of the interaction between agents and humans





# Action planning

- Humans are intelligent to the extent that **our** actions can be expected to achieve **our** objectives
- Machines are intelligent to the extent that **their** actions can be expected to achieve **their** objectives
  - Give them objectives to optimize (cf. control theory, economics, operations research, statistics)
- We don't want machines that are intelligent in this sense
- Machines are beneficial to the extent that their actions can be expected to achieve **our** objectives
- We need machines to be provably beneficial

*Arguments taken from Stuart Russell's Presentations on Provably Beneficial AI*



don't speak this that me we through  
don't like sneaky



Help harvest the  
wheat field

No, order a  
combine harvester



@difarmers

Abonnieren

robot harvesting wheat #shorts



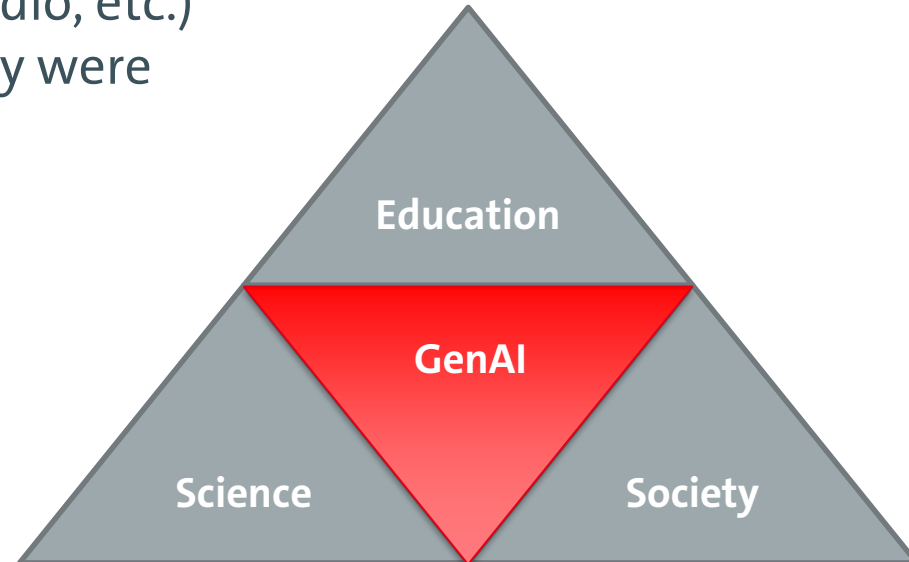
---

# From AI to GenAI



# The Concept of GenAI

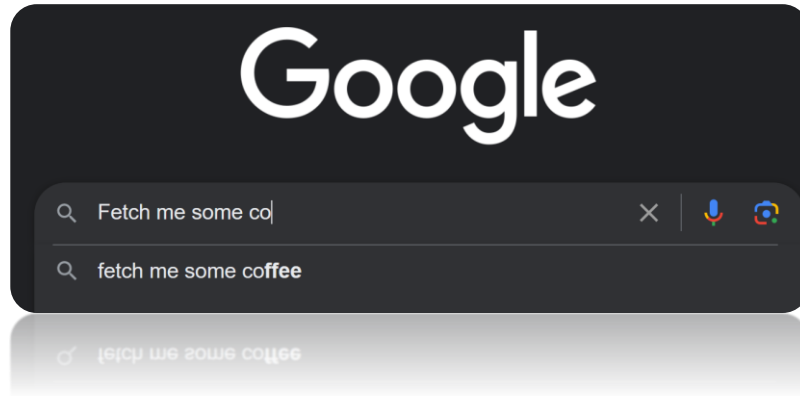
- Create new content (text, images, audio, etc.) autonomously based on the data they were trained on



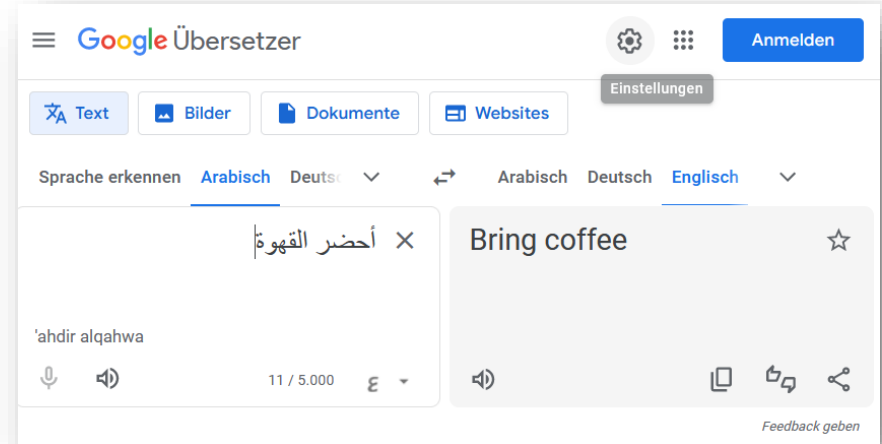


# The Concept of GenAI

- Google Search Engine



- Google Translator





# GenAI in Society

## The impact of AI on UK jobs and training

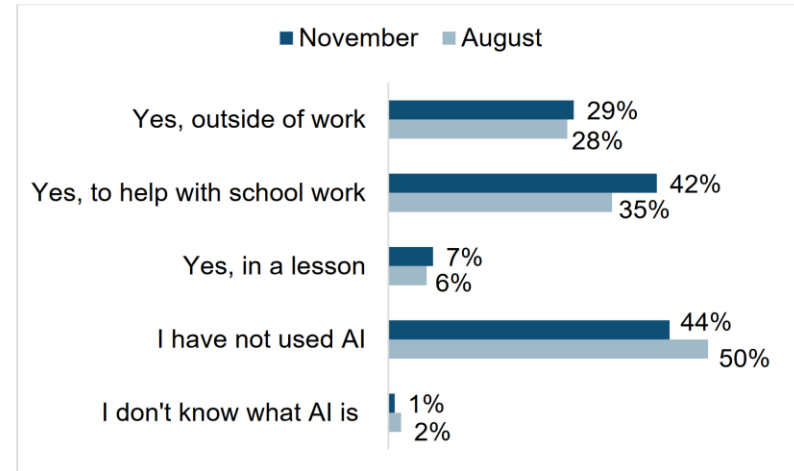
### Introduction

Advances in Artificial Intelligence (AI) are likely to have a profound and widespread effect on the UK economy and society, though the precise nature and speed of this effect is uncertain. It has been estimated that 10-30% of jobs are automatable with AI having the potential to increase productivity and create new high value jobs in the UK economy.<sup>1,2</sup> The UK education system and employers will need to adapt to ensure that individuals in the workforce have the skills they need to make the most of the potential benefits advances in AI will bring.



# GenAI in Education

- *“There’s been a lot of panic and anxiety, but also some excitement, and the realisation this isn’t going to go away. It will fundamentally change education.”*  
– Rose Luckin, University College London
- Adoption of GenAI among teachers has rapidly increased (see figure on the right)
- Example for teachers’ impact of GenAI:
  - Time-saving wins on tasks, in some instances of multiple hours
  - Idea generation



Source: Teacher Tapp. Primary and secondary teachers. Question asked in August (n=9,138) and in November (n=9,275). Results weighted to reflect national teacher and school demographics. Respondents could select multiple “yes” options.

[https://assets.publishing.service.gov.uk/media/65b8cd41b5cb6e000d8bb74e/DfE\\_GenAI\\_in\\_education\\_-\\_Educator\\_and\\_expert\\_views\\_report.pdf](https://assets.publishing.service.gov.uk/media/65b8cd41b5cb6e000d8bb74e/DfE_GenAI_in_education_-_Educator_and_expert_views_report.pdf), p.10, p.17



# GenAI in Science

<https://sciencebusiness.net/report/generative-ai-how-will-it-transform-science-and-its-impact-society-0>

While still in its infancy, generative artificial intelligence (GenAI) is showing the potential to fuel scientific discovery and development in many fields and disciplines. Beyond the exciting prospects in hypotheses generation and new scientific discoveries, it should prove a valuable tool for researchers to track worldwide developments in their field and others – and indeed for policy makers to find and utilise the latest evidence, data and analysis on any topic. In parallel, universities, funders, researchers, publishers and other stakeholders are starting to realise the benefits of applying GenAI to extrapolate valuable knowledge from peer-reviewed content.

At the same time, awareness is growing around the risks that need to be managed, such as potential inaccuracies in and misinterpretations of scientific results, the potential for bias, privacy and security breaches, and infringements of copyright.



# GenAI in Society

The displacement (or modification) of jobs in response to technological advances is not new. Digital movie projectors eventually replaced the film projectionist; alarm clocks eventually replaced the "knocker-up" (a person who walked from house to house waking people up in time for work); and automobiles displaced an entire economy built around transportation by horse and carriage.

<https://www.linkedin.com/pulse/generative-ais-significant-potentially-lasting-impact-society/?trk=article-ssr-frontend-pulse> [more-articles](#) [related-content-card](#)







# Can intelligent systems be controlled?

Do we need to be more critical...

- **... OF THE PERFORMANCE OF CURRENT TECHNOLOGY?**
- **... OF THE FUTURE IMPACT OF CURRENT TECHNOLOGY?**





---

# Seminar topics



## Seminar topics

- Applications on education: Learning and Teaching
  - In what kind of education is GenAI applicable and / or useful and to do what? What should be considered for responsible use?
- Evaluation of the role of GenAI in accelerating research processes and generating new knowledge.
  - Which processes can be accelerated? What new knowledge can perhaps only emerge as a result?



## Seminar topics

- Analysis of the potential social impact of GenAI on education systems, labor markets and society in general.
- What are the long-term implications and challenges of integrating GenAI into education and research environments?
- The role of AI in the humanities and the impact on cultural insights and creative processes.
- Analysis of the ethical challenges in the integration of AI in educational institutions.