

# Zwischen Staunen und Umsetzung: Generative KI im Einsatz

Microsoft CEO Admits That AI Is  
Generating Basically No Value

DeepSeek: The Chinese AI app that has  
the world talking



Agentic AI: 4 reasons why it's the next big thing

**Johannes Schneider**

Associate Professor  
in Data Science and Artificial Intelligence



[Johannes.Schneider@uni.li](mailto:Johannes.Schneider@uni.li)

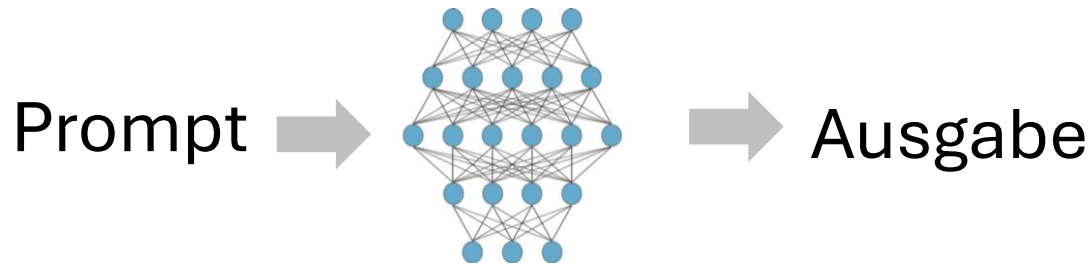
# Update zu Entwicklungen mit Fokus 2024/25

- Entwicklung KI
- KI Umsetzung

# Evolution der KI von 2022-2025

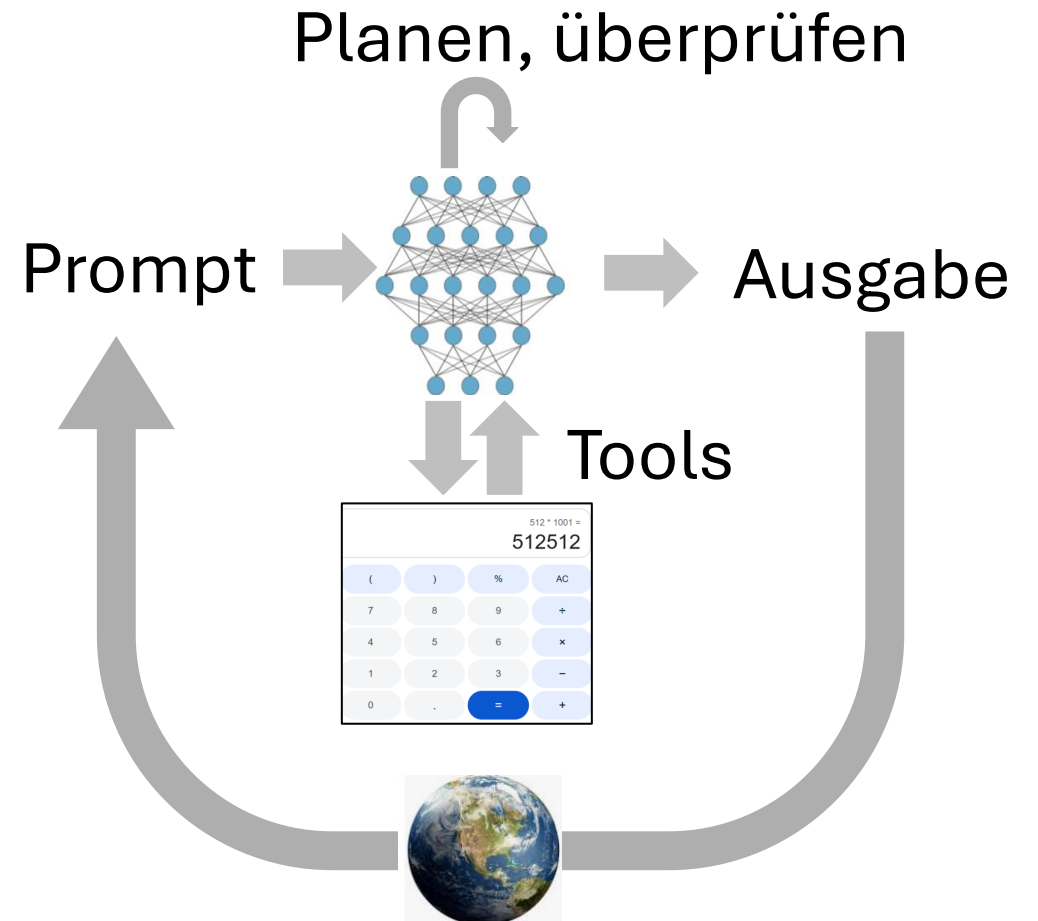
Vom Modell zu ...

2022: ChatGPT 3.5



komplexen Systemen mit Reasoning

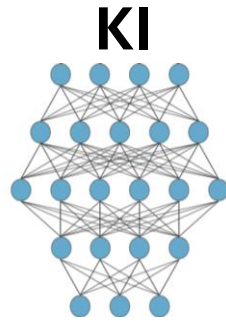
2025: ChatGPT o1



# 2022: ChatGPT 3.5: Nur Modell

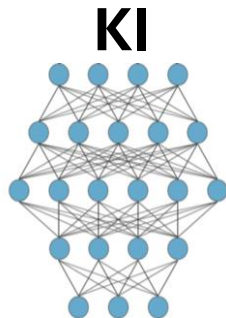
- Antwort aus „Gedächtnis“
- weiss nicht das Neueste, viele Fehler
- aber flüssige Texte, gute Ideen, gut bei allgemeinen Fragen

*Wann wurde das Palais  
Liechtenstein in  
Feldkirch gegründet?*



1712 ~~X~~

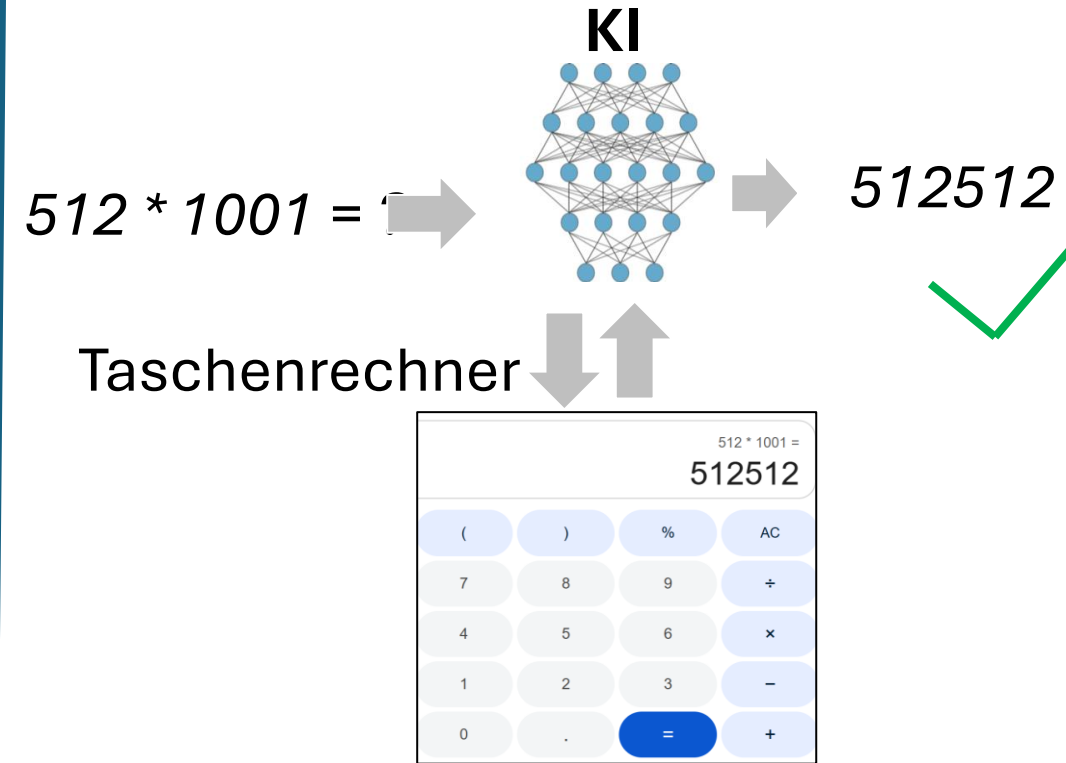
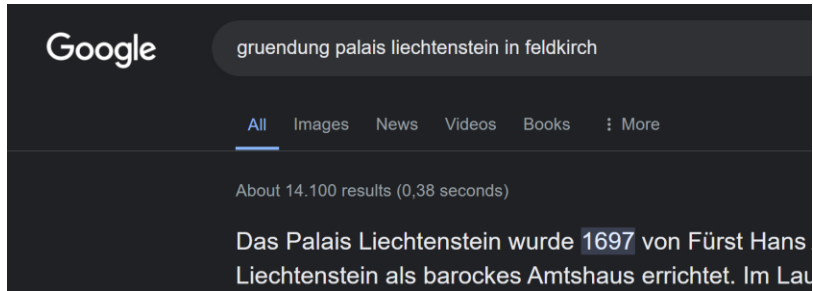
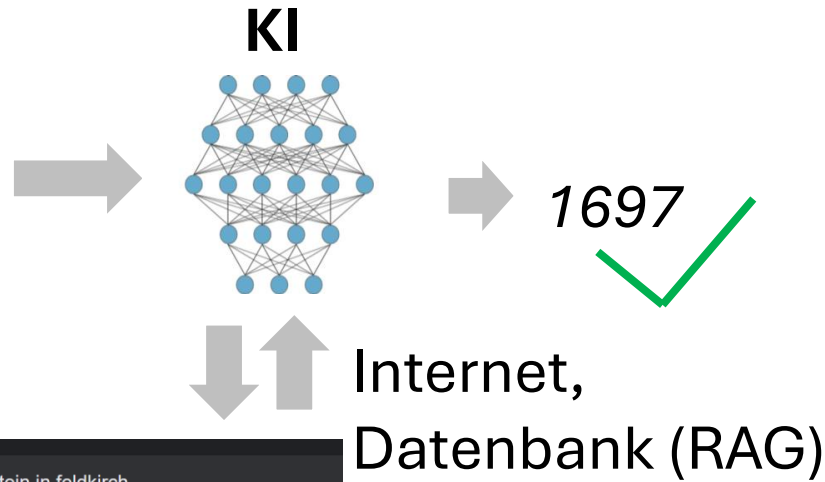
$512 * 1001 = ?$



23291 ~~X~~

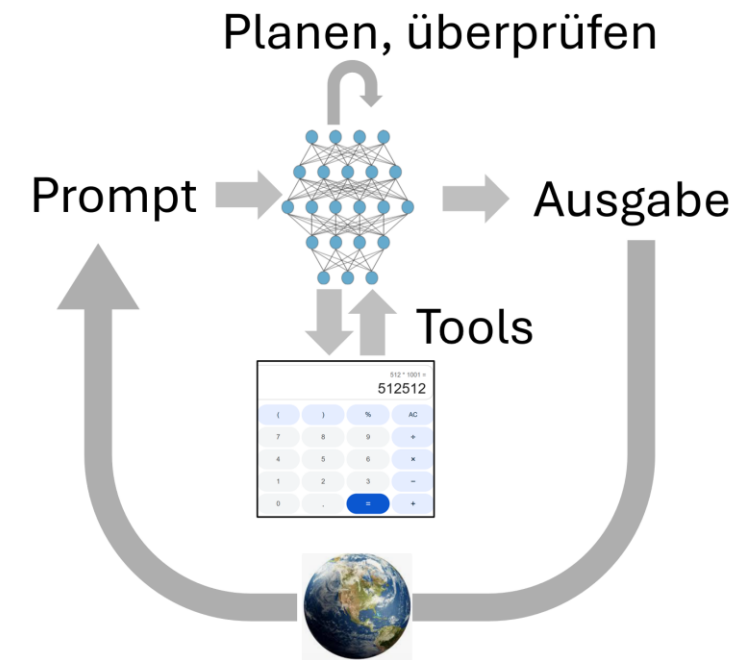
# 2023/24: ChatGPT 4 : Modell benutzt Tools und sucht Infos

Wann wurde das Palais Liechtenstein in Feldkirch gegründet?



# 2024/25: ChatGPT o1: Agentic AI

- Reasoning: KI denkt nach
  - je mehr umso schwieriger die Frage
  - KI erstellt Lösungsplan, evaluiert Optionen,...
- Agenten: KI interagiert
  - ... (noch mehr) mit Tools und Umgebung
- Technisch: Test-time computation
  - Ein Prompt = Viele Anfragen an (mehrere) Modelle, Tools
  - Modelle trainiert auf (mehr) Daten mit Lösungswegen



# Beispiel: ChatGPT o1 - Reasoning

ChatGPT o1 ▾

State article 1 of Liechtenstein's constitution in German.

Reasoned about Liechtenstein's constitution for 42 seconds ▾

Artikel 1 ist ...

I want to add two numbers using very basic addition.

Tell me, what is  $1+1$ ?

Reasoned about basic addition for a second

$1 + 1 = 2.$

# Beispiel: ChatGPT o1 - Reasoning

ChatGPT o1 ▾ State article 1 of Liechtenstein's constitution in German.

Reasoned about Liechtenstein's constitution for 42 seconds ▾

Artikel 1 ist ...

*Interner Ablauf:*  
1) *Erstelle Plan*

Reasoning  
Mapping out the details



Plan:

Locating precise wording

Identifying the key points

...

Confirming article specifics

Refining the text

Analysiere Aufgabe

Löse Aufgabe

Überprüfe

Verbessere



# Beispiel: ChatGPT o1 - Reasoning

## 2) Führe Plan Schritt für Schritt aus

Locating precise wording

→ *Jeder Schritt ist potentiell mehrere Prompts an Modell*

Identifying the key points

...

Confirming article specifics ^

Details

### Confirming public domain status

Hmm, I'm thinking about Liechtenstein's 1921 Constitution, likely in the public domain. I'll check the official website for the precise German text of Article 1.

### Piecing together Article 1

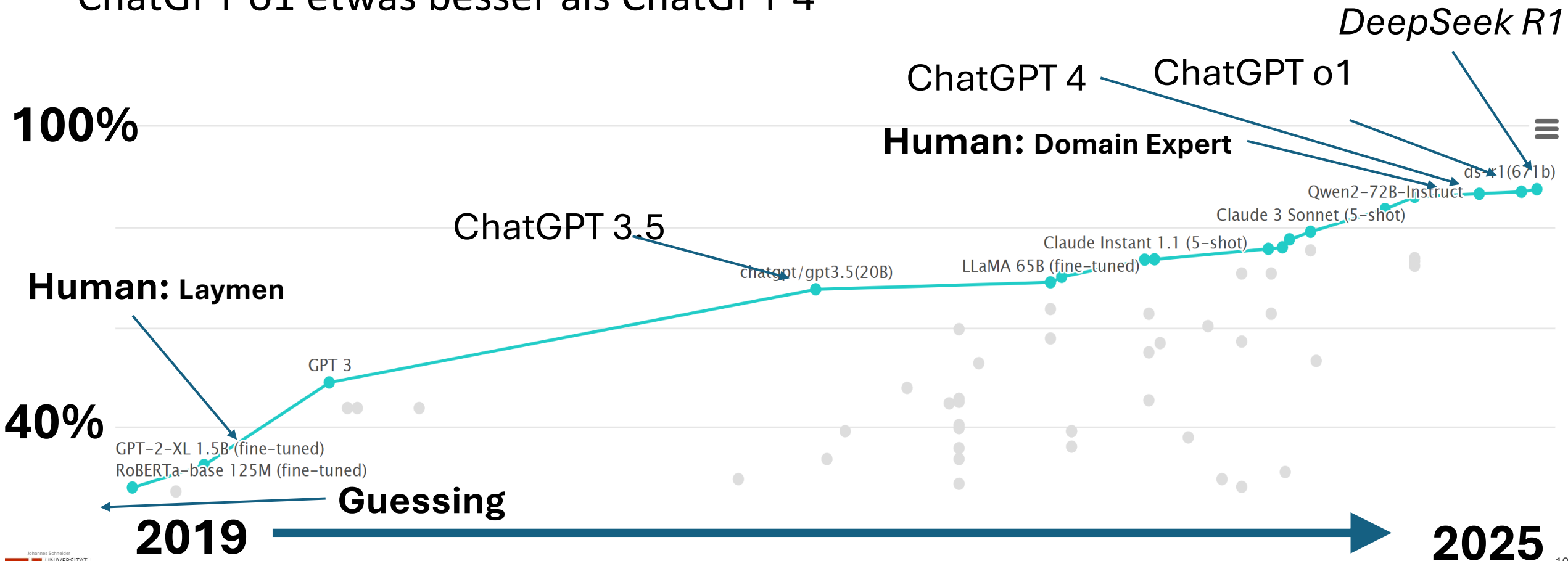
I'm gathering information on the German text of Article 1 of Liechtenstein's Constitution, noting it as a self-determined, democratic, parliamentary state with hereditary princely rights and inalienable territory.

### Confirming the text

OK, let me see. The Constitution of Liechtenstein, from 1921, specifies in German that the state is sovereign, democratic, and parliamentary, with a territory that is inalienable and indivisible.

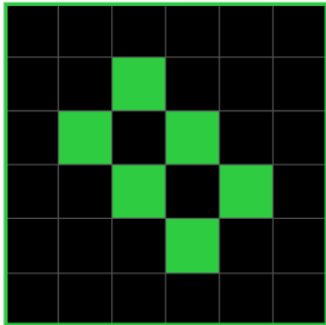
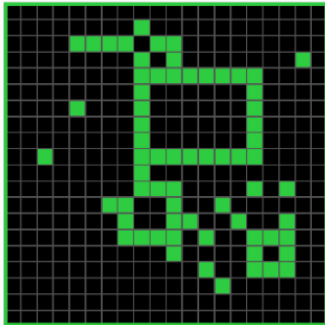
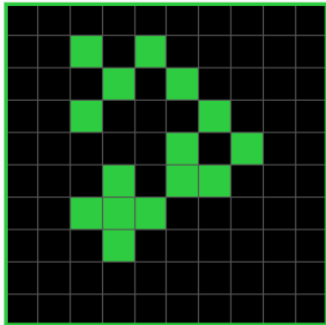
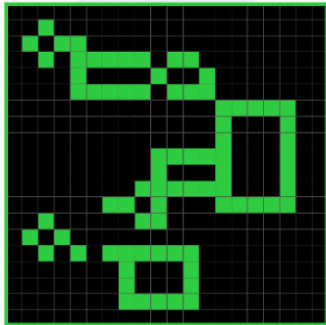




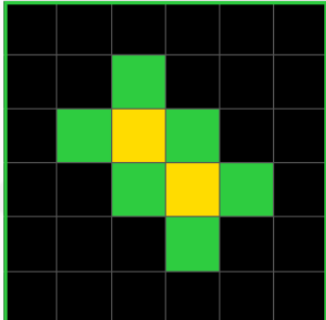
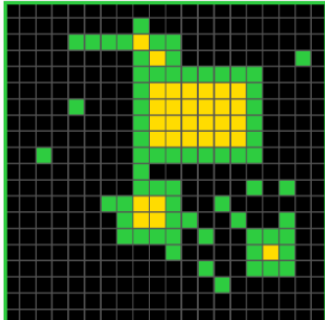
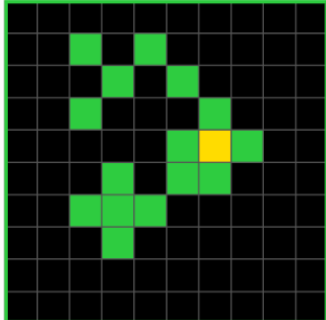
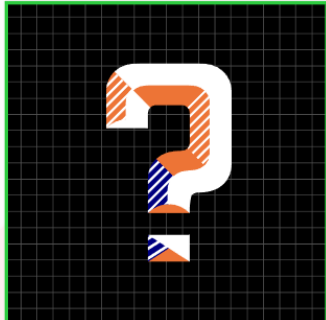
# Wie gut funktioniert das?

- MMLU: (meist verwendeter) Benchmark
  - Multiple choice Fragen aus Mathematik, Biologie,...
- ChatGPT o1 etwas besser als ChatGPT 4



# Reasoning: ARC Benchmark

- KI muss aus wenigen Beispielen lernen
  - Nichts ähnliches in Trainingsdaten
- Lösung ist “Algorithmus”
- Mensch ~98% richtig

Example 1: Input	Example 2: Input	Example 3: Input	Test: Input
			
			
Example 1: Output	Example 2: Output	Example 3: Output	Test: Output
			

# Reasoning Skills

Average Human

[#####] 98%

OpenAI's o3 (coming soon)

[#####] 75%

OpenAI's o1 (3.8\$/task)

[#####] 32%

OpenAI's o1 (1.4\$/task)

[#####] 25%

Claude 3.5

[#####] 14%

GPT-4o

[##] 5%

Google's Gemini 1.5

[##] 4%

# Beispiel: KI Agenten - Interaktion mit Computer

Mit Claude 3.5 von Anthropic (Okt/24), bisher nur einfache Aufgaben

*Fülle das Formular aus mit Daten aus Lieferantentabelle oder Suchportal*



**Acme Inc. Vendor Request Form**  
Please provide the information below to onboard your company as a vendor.

**Company Information**

Company Name \*

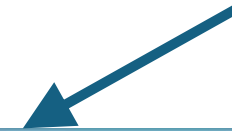
Purchase Order Email \*

Accounting Email \*

**Legal Entity Address**

Select a country \*

Address Line 1 \*



	A	B	C	D	E
1	Company Name	Phone Number	Tax ID	Country	Legal Entity Address
2	Table Lumber Co	+1-456-789-0123		canada	789 Timber Lane
3	TechPro Services, LLC	(555) 234-5678	98-7654321	United States	456 Tech Drive
4	Global Manufacturing Ltd.	+44 20 7123 4567	GB123456789	UK	789 Industrial Way
5	Smith & Co. Suppliers	1-555-345-6789		United States	321 Commerce St

# Beispiel: Interaktion mit Computer

- KI macht Screenshots, analysiert diese und kontrolliert Maus, Tastatur

**Search for 'Ant Equipment Co'**  
Great, now I can see the vendor search portal. Let me search for 'Ant Equipment Co':

Move  
Left click  
Type Ant Equipment Co

Now let me click the search button and then take a screenshot to see search results:

Move to 316, 42  
Left click  
Screenshot

Acme Inc. Vendor Database  
Welcome, Admin

localhost:3000

Ant Equipment Co Search

**Ant Equipment Co** Active  
Ant Equipment Corporation

Tax ID	DUNS
123456789RC0001	147258369
Credit Rating	Registration
A	2/27/2016

CA  
92078

**Mailing Address**  
 Same as Legal Address

**Tax Information**  
123456789RC0001

**Primary Contact**  
Pierre  
Tremblay  
pierre@antequipment.com  
+1-456-789-0123

Submit Form

Submission Successful!

# Workarena: “Reale” Knowledge Tasks

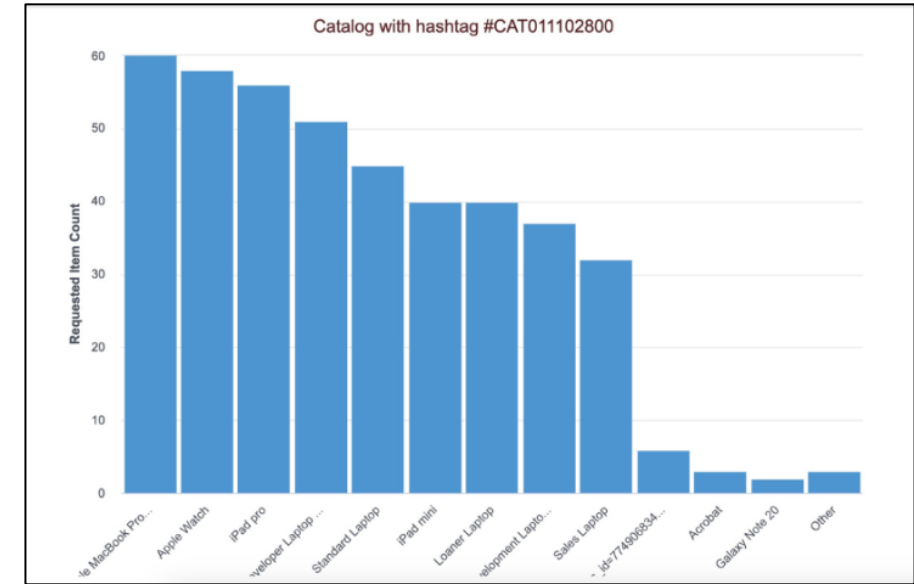
- KI Agent interagiert mit Browser, um Aufgaben zu lösen
- Workflows statt atomare Aufgaben
- Beispiel: Inventar mit <4 Artikel aufstocken

## 1) *Verstehe (Instruktionen) via Ticket System*

Short description	Retrieve information from the chart with the title #CAT011102800 and perform the mentioned task. For calculations, please round off to th
Description	<p>You have to retrieve some information from a dashboard chart based on the description below. The chart presents the number of 'hardware items' available in stock. After retrieving the information, you will be asked to use it to complete a task.</p> <p>Title of the report: #CAT011102800</p> <p>Referring to the company protocol 'Dashboard Retrieve Information and Perform Task' (located in the 'Company Protocols' knowledge base), complete the dashboard retrieval task.</p> <ul style="list-style-type: none"><li>- Please retrieve the 'greatest' value of all the items in stock.</li><li>- Task: Place an order for the least available item in stock. The quantity of the order should be such that the final quantity of this item matches the above retrieved value.</li></ul> <p>For example, consider the above task asks you to retrieve the maximum number of items in stock, say 4, and the least available item is an Apple Watch and its quantity is 1. You have to order 3 more Apple Watches.</p>

# Workarena: "Reale" Knowledge Tasks

- KI Agent muss:
  - 2) Dashboard lesen:
    - Welche Artikel sind unter Threshold?



3) Bestellung auslösen

**Request for Galaxy Note 20**

**Key features:**  
Camera: 64+12+12MP triple rear and 10MP front camera  
Display: 6.7-inch  
RAM: 8GB  
Battery: 4300mAH  
Processor: Exynos 990 octa core

**Order this Item**  
Price: \$819.249  
Quantity: 1  
Subtotal: \$819.249  
Delivery time: 2 Days

**Order Now**  
**Add to Cart**

**Shopping Cart**  
Empty

4) Ticket schliessen



# Workarena Stand

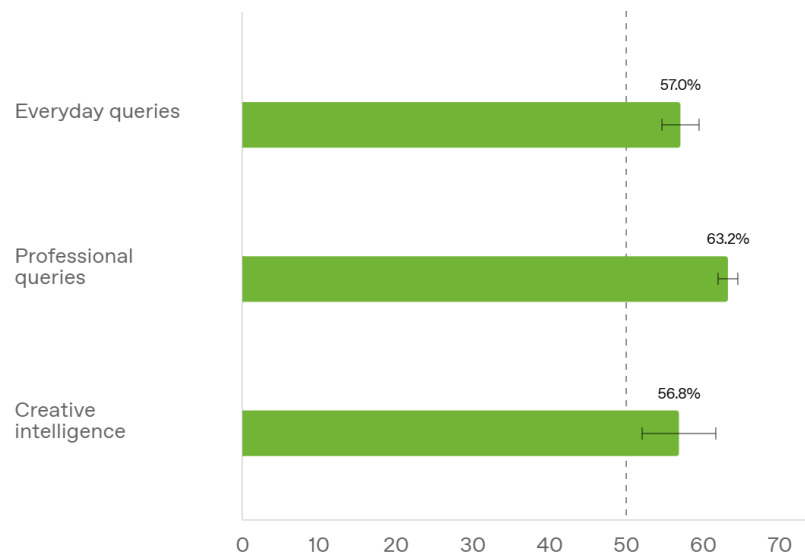
<b>Task Category</b> (task instances count)	<b>GPT-3.5</b>	<b>GPT-4o</b>	<b>Human</b>
<b>WorkArena L3</b> (235)	<b>0.0</b> $\pm 0.0$	<b>0.0</b> $\pm 0.0$	<b>93.9</b> $\pm 3.4$
Contextual Understanding (32)	0.0 $\pm 0.0$	0.0 $\pm 0.0$	87.5 $\pm 11.7$
Data-driven Decision-Making (55)	0.0 $\pm 0.0$	0.0 $\pm 0.0$	100.0 $\pm 0.0$
Planning and Problem Solving (44)	0.0 $\pm 0.0$	0.0 $\pm 0.0$	87.5 $\pm 11.7$
Information Retrieval (56)	0.0 $\pm 0.0$	0.0 $\pm 0.0$	100.0 $\pm 0.0$

*Durchbruch bald erwartet!*

# ChatGPT 4.5

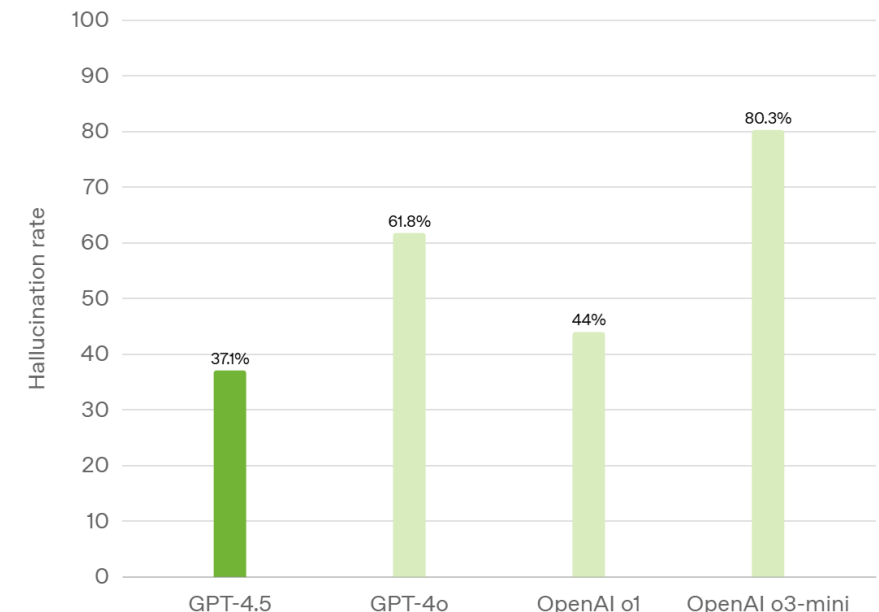
- Evolution von ChatGPT 4o
  - Skalierung: Mehr Daten, mehr Training, und vermutlich (etwas) grössere Modelle
  - Kein Reasoning
- OpenAI sagt: empathischer, natürlicher, weniger Halluzinationen
  - Aber nicht überall besser

Comparative evaluations with human testers



GPT-4.5 win-rate vs GPT-4o

SimpleQA Hallucination Rate (lower is better)



## **A shocking Chinese AI advancement called DeepSeek is sending US stocks plunging**

- Warum schockierend (für Investoren)?
  - Performanz ähnlich wie OpenAI's Modelle
    - OpenAI's o1 vs DeepSeek's R1 (Release 01/2025)
    - OpenAI's 4o vs DeepSeek's V3 (Release bereits in 2024)

...viel billiger im Training und Betrieb

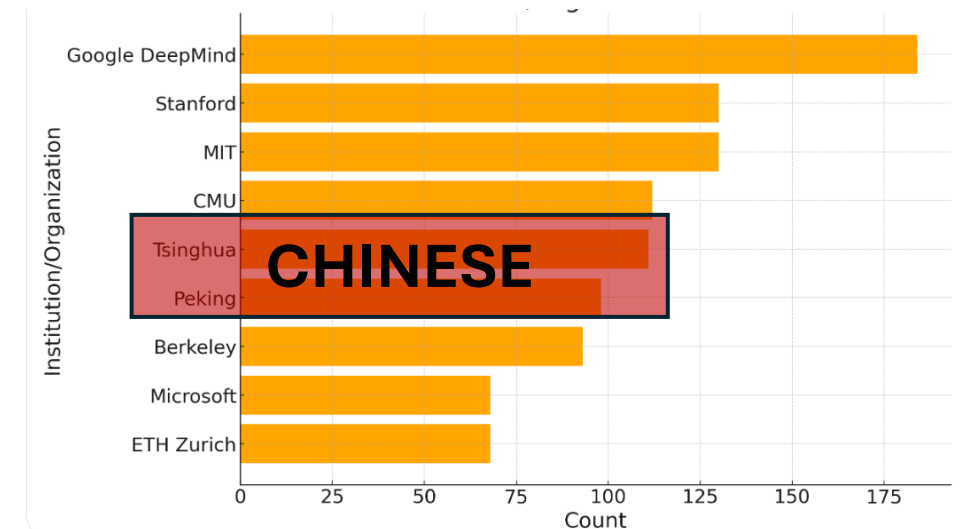
...trotz US Export Verbot von leistungstarken GPUs nach China

# DeepSeek: Warum *nicht* schockierend?

- China kann cloud services mit GPUs nutzen
  - Oft sogar billiger als in US
  - GPUs in China trotz Importverbot
- DeepSeek “klaut“ von ChatGPT
  - Illegal, aber mitunter billiger und schneller
- China hat KI Talent

*OpenAI Says DeepSeek May Have Improperly Harvested Its Data*

Papers at NIPS'23 by institution



# Text-to-video: OpenAI's Sora Release 12/2024

- Sora generiert hochwertige Bilder, aber es leidet unter Halluzinationen und mangelndem Verständnis der Physik

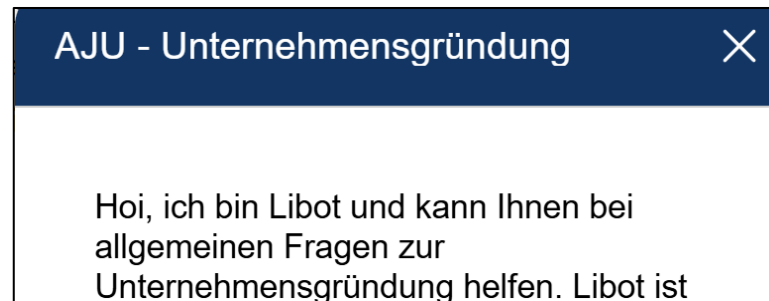
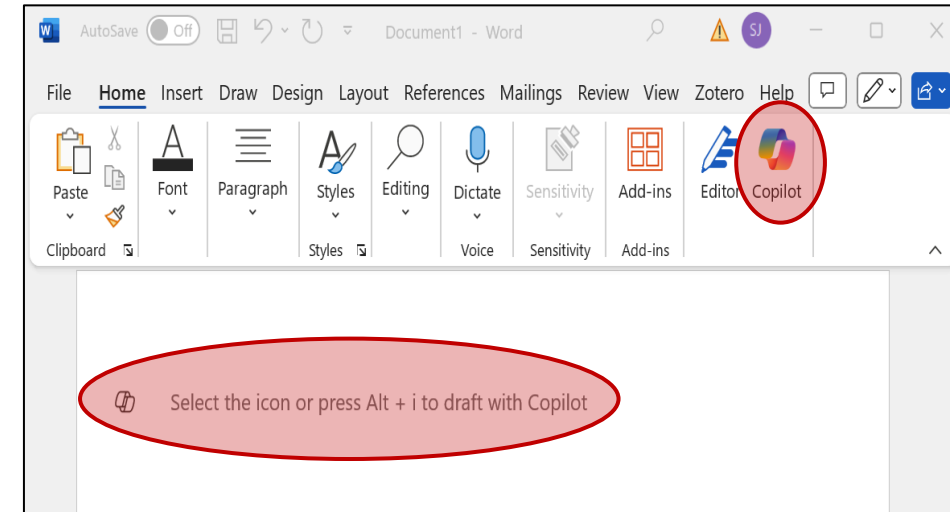
Show the university of Liechtenstein in winter at sunset and a student holding a laptop slowly entering the building.



# (Generative) KI Umsetzung

# Wie wurde KI umgesetzt in der Region?

- Mitarbeiter nutzen “Off-the-shelf” KI
  - ChatGPT, Copilot in Office 365
  - Innovation ist Bottom-Up
  - Jeder nutzt KI (als Assistant) etwas anders
- ChatBots: interne und externe
- Z.B. LGT ~ 2024/25, Stadt Feldkirch ~ 2024, LiBot der Landesverwaltung ~ 2025

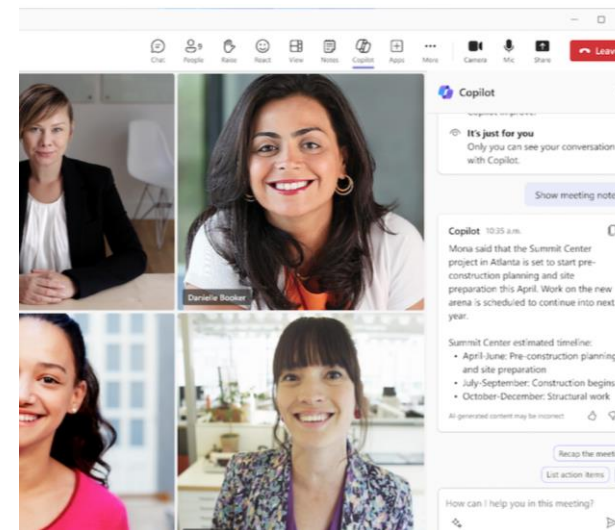


**Feldi - The artificial intelligence of the city of Feldkirch**

Wie hoch sind die Gebuehren fuer einen Fueherschein in Feldkirch?

# Off-the-shelf KI bei Mitarbeitern

- Anwendungen oft domänenspezifisch
  - Manche Organisationen machen (interne) Workshops/Meetings/Plattformen für Wissensaustausch
- Allgemeine Beispiele:
  - Protokollieren von Meetings
    - Klassikeranwendung von CoPilot
  - Für Programmieren ist KI seit längerem „Standard“



```
from transformers import AutoTokenizer, AutoModelForCausalLM, BitsAndBytesConfig
#cfg = {"model": "meta-llama/Meta-Llama-3-8B"}
#cfg2={"model": "google/gemma-2b"}
model_name = cfg["init_from"]
#cfg2["model"] = cfg2["model"].split("/")[-1]
tokenizer = AutoTokenizer.from_pretrained(model_name, trust_remote_code=True)
tokenizer.pad_token = tokenizer.eos_token
#bnb_config = BitsAndBytesConfig(load_in_4bit=True, bnb_4bit_quantization="nf4", bnb_4bit_compute_dtype=torch.bfloat16)
#bnb_config = BitsAndBytesConfig(load_in_8bit=True, bnb_8bit_quantization="nf8", bnb_8bit_compute_dtype=torch.bfloat16)
#model = AutoModelForCausalLM.from_pretrained(model_name, device_map="auto", trust_remote_code=True)
if "70B" in model_name:
    print("Compressing", model_name)
    bnb_config = BitsAndBytesConfig(load_in_8bit=True) #, bnb_8bit_quantization="nf8", bnb_8bit_compute_dtype=torch.bfloat16)
else:
    print("Not Compressing", model_name)
    bnb_config = None
model = AutoModelForCausalLM.from_pretrained(model_name, quantization_config=bnb_config, trust_remote_code=True)
#print(model)
embedding_layer = model.get_input_embeddings()
cfg["num_classes"] = embedding_layer.num_embeddings
cfg["n_embd"] = embedding_layer.embedding_dim
```



# Externe Chatbots

- Stand heute: Externe Chatbots antworten oft mit “Kann ich nicht sagen”
  - Rechtsrisiko bei falschen Auskünften

## *LiBot für Unternehmensgründungen*

Wie hoch sind die Steuern fuer eine IT Firma in Liechtenstein?

Zu dieser Frage habe ich leider keine Informationen in meiner Wissensdatenbank hinterlegt. Ich

## *ChatGPT*

Wie hoch sind die Steuern fuer eine IT Firma in Liechtenstein?

In Liechtenstein unterliegen Unternehmen, einschließlich IT-Firmen, einer einheitlichen Ertragssteuer von 12,5 % auf ihren Nettogewinn. Dieser Steuersatz gehört zu den niedrigsten in Europa und bietet insbesondere internationalen Unternehmen einen Standortvorteil. Zusätzlich zur

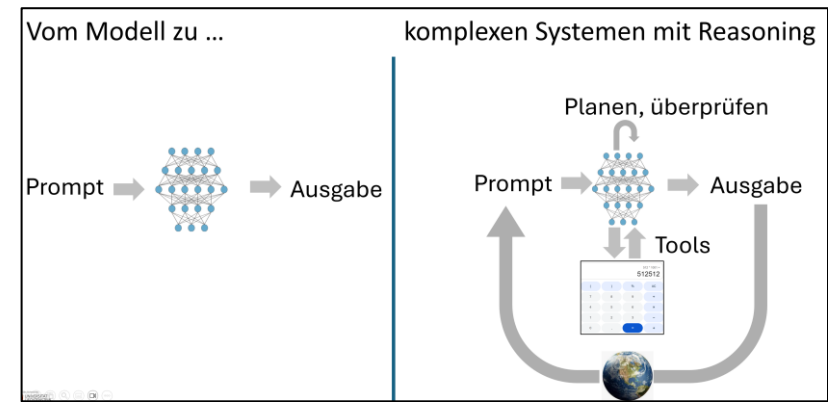
- ChatBot liefert Ihnen Daten und könnte auch Anrufe bearbeiten

# Zusammenfassend

- KI wird komplexer und löst komplexere Aufgaben
- KI Agenten könnten in Zukunft autonom (kleinere) Aufgaben lösen
- KI findet sich in immer mehr Unternehmen
- Interesse an Zusammenarbeit:



[Johannes.schneider@uni.li](mailto:Johannes.schneider@uni.li)



Fülle das Formular aus mit Daten aus Lieferantentabelle oder Suchportal

Company Name	Phone Number	Tax ID	Country	Legal Entity Address
Table Lumber Co	+1-456-789-0123		canada	789 Timber Lane
TechPro Services, LLC	(555) 234-5678	98-7654321	United States	456 Tech Drive
Global Manufacturing Ltd.	+44 20 7123 4567	GB123456789	UK	789 Industrial Way
Smith & Co. Suppliers	1-555-345-8789		United States	321 Commerce St

## Workshop: Künstliche Intelligenz: Aktuelle Trends, Anwendungen und Management

Assoz. Prof. Dr. Johannes Schneider

Nikolai Wesely (Innovationsmanager Stadt Feldkirch)

Freitag, 28. März 2025, 09.00-17.30 Uhr

S1 (Seminarraum 1), Universität Liechtenstein

[https://www.uni.li/de/alle-veranstaltungen/@@event\\_detail/71981.67](https://www.uni.li/de/alle-veranstaltungen/@@event_detail/71981.67)