

# Digitaler Humanismus

Julia Neidhardt

9. Oktober 2024





**Wir leben in einer Zeit, in der Technologie allgegenwärtig ist.**

Bildquelle: Canva



# Die digitale Landschaft unserer Zeit

## ■ Digitale Transformation

- Steht für die Informatisierung unserer Gesellschaften
- Ein komplexer Prozess, der sowohl uns als auch die Welt verändert
- Umfasst sozioökonomische und technische Aspekte

## ■ Internet und Web

- Dienen als globale Infrastruktur des digitalen Zeitalters
- Informatik und ihre Werkzeuge fungieren als das „Betriebssystem“ unserer Gesellschaft

## ■ Lehren aus der Pandemie

- Haben die Macht und das Potenzial des Internets/Webbs deutlich gemacht
- Zeigten jedoch auch dessen Grenzen und Schwächen auf

## ■ Künstliche Intelligenz (KI)

- Stellt den logischen nächsten Schritt in der digitalen Entwicklung dar
- Treibt die weitere Transformation maßgeblich voran

-> **Zwischen Utopie und Dystopie**



# Chancen von Digitalisierung und KI

- **Nahtloses Funktionieren durch KI**
  - KI optimiert Prozesse und sorgt für Effizienz in Systemen
- **Förderung von Teilhabe**
  - Digitale Plattformen können die politische Teilhabe stärken
- **Lösungen für globale Herausforderungen**
  - KI hilft bei der Bewältigung komplexer Probleme wie Klimawandel, Gesundheit und Bildung



Bildquelle: Wikipedia



# Risiken von Digitalisierung und KI

- **Monopole und Machtkonzentration**
  - Algorithmen und digitale Plattformen verstärken die Marktdominanz von Tech-Giganten
- **Souveränität und Überwachung**
  - Smart-City-Technologien oder Gesichtserkennung im öffentlichen Raum werfen Fragen zu Kontrolle, Privatsphäre und staatlicher Überwachung auf
- **Informationskrise**
  - Algorithmen schaffen Echokammern und verstärken Fake News, was den politischen Diskurs verzerrt
- **Entscheidungsfindung und Kriegsführung**
  - Autonome Systeme und automatisierte Entscheidungsprozesse gewinnen an Einfluss, auch in militärischen Kontexten
- **Automatisierung am Arbeitsplatz**
  - Automatisierung verändert Arbeitsmärkte und schafft neue Herausforderungen für Arbeitnehmer, insbesondere durch den Einsatz von KI

**“The system is failing”** – Tim Berners-Lee, 2018



# Workshop Digitaler Humanismus 2019

- Über 100 TeilnehmerInnen – weltweit – Informatik, Philosophie, Geschichte, Anthropologie, Recht, Ökonomie, Politikwissenschaften, Mathematik, Soziologie
- Drei Themenblöcke:
  - Geschichte und den Auswirkungen der Informationstechnologien
  - Gesellschaft, Künstlicher Intelligenz und Ethik
  - Dynamik und Zukunft des Sektors





# Ergebnis: Vienna Manifesto on Digital Humanism

- Das Manifest ist ein **Aufruf zur Reflexion und Handlung** (gerichtet an Wissenschaft und Entscheidungsträger:innen), ebenso ein Forschungsprogramm.
- **Einige Kernprinzipien:**
  - Digitale Technologien sollten so gestaltet werden, dass sie **Demokratie und Inklusion** fördern.
  - **Fairness, Verantwortung und Transparenz** von Softwareprogrammen und Algorithmen.
  - Maßnahmen/Eingriffe **gegen Technologiemonopole**.
  - Entscheidungen, die Menschenrechte betreffen, müssen **von Menschen getroffen** werden.
  - Die **Verbindung verschiedener wissenschaftlicher Disziplinen** ist wesentlich.
  - **Universitäten** schaffen neues Wissen und fördern kritisches Denken; sie **haben eine besondere Verantwortung**.
  - Akademische und industrielle Forscher\_innen müssen einen offenen **Dialog mit der Gesellschaft** pflegen.
  - Die akademische Lehre **muss Geisteswissenschaften, Sozialwissenschaften und Ingenieurwesen verbinden**.

<https://caiml.org/dighum/dighum-manifesto/>



# Ergebnis: Vienna Manifesto on Digital Humanism

- Das Manifest ist ein **Aufruf zur Reflexion und Handlung** (gerichtet an Wissenschaft und Entscheidungsträger\_innen), ebenso ein Forschungsprogramm.
- **Einige** **Digitaler Humanismus** analysiert und, noch wichtiger, beeinflusst das komplexe Zusammenspiel von Technologie und Menschheit, für eine bessere Gesellschaft und ein besseres Leben, unter voller Achtung der universellen Menschenrechte.
  - Digitaler Humanismus fördert Inklusion.
  - Fairness und Transparenz von Algorithmen.
  - Maßnahmen zur Erreichung von Digitaler Inklusion.
  - Entscheidungen müssen offen werden.
  - Die Digitalisierung muss menschlich sein.
  - Universalität: Digitaler Humanismus muss eine globale Perspektive einnehmen.
  - Akademische Lehre muss Geisteswissenschaften, Sozialwissenschaften und Ingenieurwesen verbinden.
- Die akademische Lehre muss **Geisteswissenschaften, Sozialwissenschaften und Ingenieurwesen verbinden**.

<https://caiml.org/dighum/dighum-manifesto/>



# The Poysdorf Declaration (2021)

- Poysdorf Declaration on Digital Humanism: “A Compass for Citizens during the Digital Transformation”
- Unterzeichnet von den Außenministern von
  - Österreich
  - Tschechische Republik
  - Slowakische Republik

## The Poysdorf Declaration

### Digital Humanism: A Compass for Citizens During the Digital Transformation

While frontier technologies are celebrated around the world as being part of the solution for many of our global challenges, we also need to be aware of the potential impact of technology on our privacy, our freedom, our democratic institutions.

Concerns about novel technologies are now shaping our world today: Increasing automation could soon make a large part of our economy redundant. The gap between the beneficiaries of digitalization and those who feel marginalized by it is growing. In our economy our privacy no longer seems to be guaranteed.

Our social life is increasingly shifting into virtual spaces, creating dependencies and transforming us from involved citizens to passive consumers. Our democratic institutions are slowly being undermined by digital bubbles and echo chambers. Face recognition and other technologies are being put into the hands of authoritarian dictatorships.

Our critical infrastructure is vulnerable to cyberattacks and our commitment to resilience building and improving cybersecurity seems to have reached a point where it calls into question what it means to be a human being.

Digital humanism is a response to fears among citizens about digital technology. Digital humanism provides us with a framework for digitalization towards a human-centric approach. It allows us to use technologies like artificial intelligence and safeguards our human rights against unacceptable biases, intransparency and other threats to human autonomy when confronted with systems based on artificial intelligence.

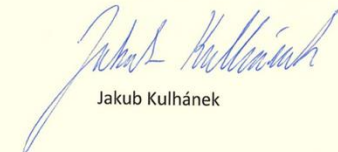
instruments, including the protection of digital human rights in the European Union and in all international fora, such as the UN, ITU, UNESCO, the OSCE or the Council of Europe.

For the Federal Ministry of European and International Affairs of the Republic of Austria



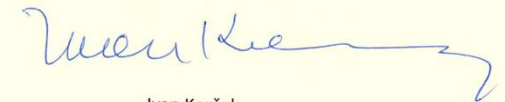
Alexander Schallenberg

For the Ministry of Foreign Affairs of the Czech Republic



Jakub Kulháněk

For the Ministry of Foreign and European Affairs of the Slovak Republic



Ivan Korčok



# Online und Offline Aktivitäten

- Etablierte **Online-Vortragsreihe** (über 60 Vorträge)
- **DigHum YouTube-Kanal**
  - <https://www.youtube.com/digitalhumanism>
  - Aufzeichnungen aller Workshops, Veranstaltungen und weiterer relevanter Vorträge
- Zwei oder mehr **internationale Workshops** pro Jahr
- Drei **DigHum Summer Schools**
  - Gefördert von der Stadt Wien und ACM Europe

**Zentraler Info-Hub:** <https://dighum.org>

JULY 3-5

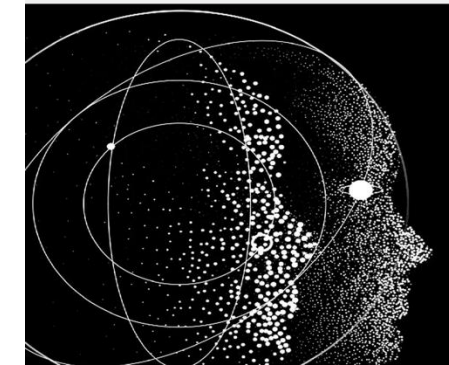
JOSEPHINUM  
WÄHRINGER  
STRASSE 25  
A- 1090 WIEN

The release of ChatGPT has stirred worldwide enthusiasm as well as anxieties. It has triggered popular awareness of the far-reaching impact of the latest generative AI, which ranges

DIGITAL  
HUMANISM  
SUMMIT



**A.I.  
AND  
DEMOCRATIC  
SUSTAINABILITY**



November 16<sup>th</sup> –  
17<sup>th</sup>  
2023

All day event.

MuseumsQuartier, Architekturzentrum Wien

1070 Vienna, Museumsplatz 1, Hof 7

Can Machines  
Save the World?

## Research, Innovation, and Education Roadmap for Digital Humanism

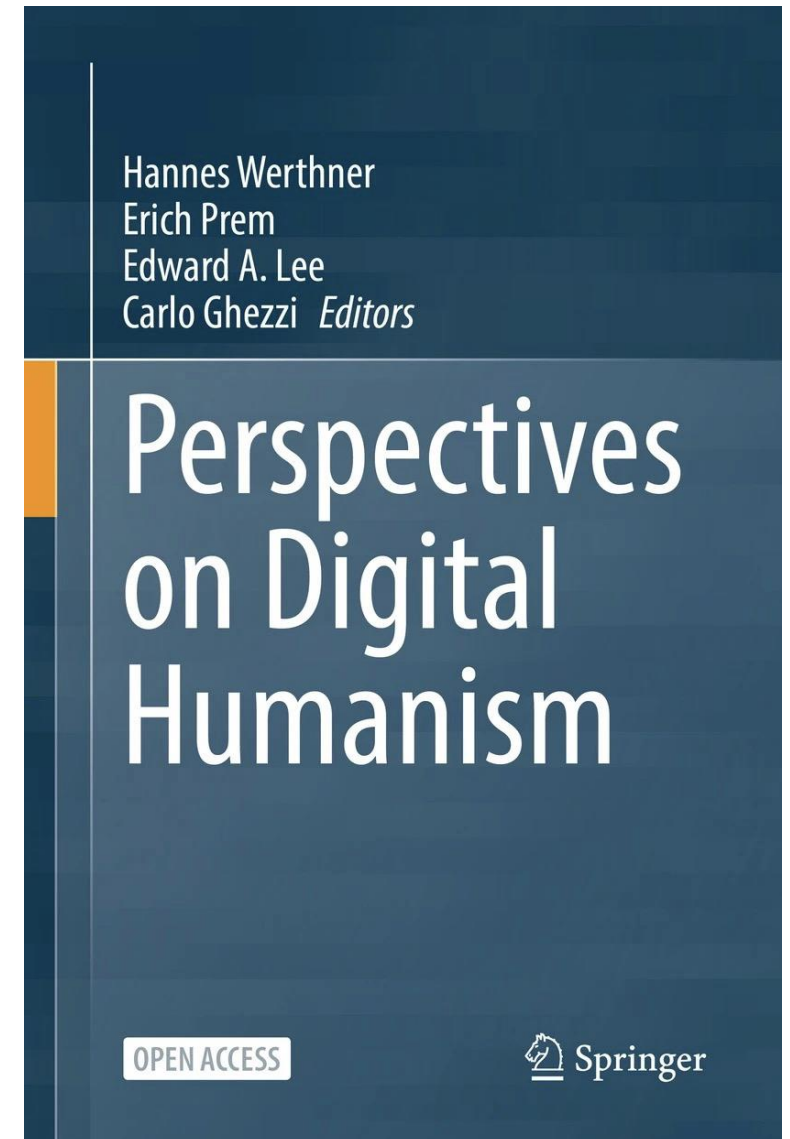
Erich Prem, Lynda Hardman, Hannes Werthner, Paul Timmers (eds.), The Digital Humanism Initiative.

*With contributions from Carlo Ghezzi, Peter Knees, Julia Neidhardt, Walter Palmetshofer, Viola Schiaffonati, Stefan Woltran, and all participants of the Fourth Workshop on Digital Humanism, Vienna, March 2022.*



# Perspectives on Digital Humanism

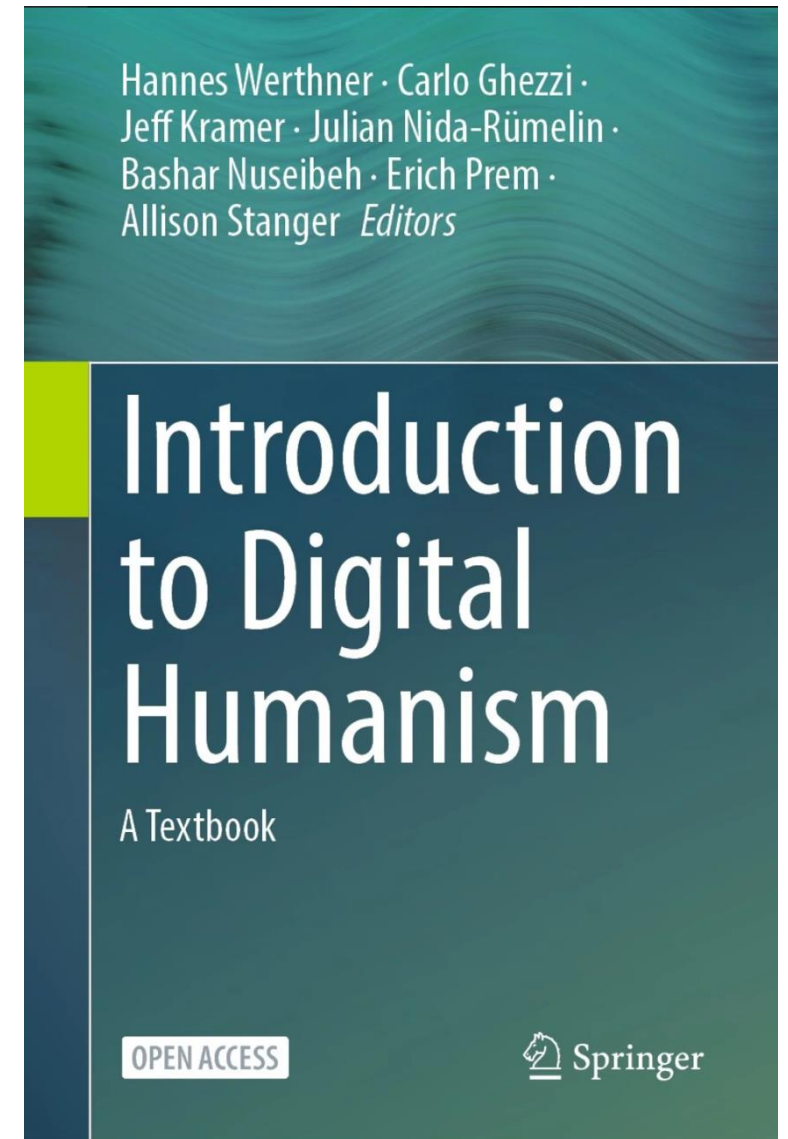
- **Herausgeber:** H. Werthner, E. Prem, E. A. Lee, C. Ghezzi
- **Ziel:** Entwicklung einer Agenda für Forschung und Handeln im aufstrebenden Bereich des Digitalen Humanismus
- **Inhalt:** Kurze Essays von ausgewählten Expert:innen aus den Bereichen Informatik, Recht, Geistes- und Sozialwissenschaften
- **Umfang:** 46 Kapitel
- **Verfügbarkeit:** Open Access
- <https://link.springer.com/book/10.1007/978-3-030-86144-5>





# Introduction to Digital Humanism

- **Herausgeber:innen:** H. Werthner, C. Ghezzi, J. Kramer, J. Nida-Rümelin, B. Nuseibeh, E. Prem, A. Stanger
- **Inhalt:** Diskutiert Digitalen Humanismus in Kontexten wie Künstlicher Intelligenz, Plattformmacht, Überwachung, Demokratie und Technikethik verfasst von Expert:innen auf dem jeweiligen Gebiet
- **Struktur:** Jedes Kapitel widmet sich einem spezifischen Thema und enthält zu beantwortende Fragen sowie eine kommentierte Leseliste
- **Umfang:** 39 Kapitel
- **Verfügbarkeit:** Open Access
- <https://doi.org/10.1007/978-3-031-45304-5>





# UNESCO Chair für Digitalen Humanismus (2023)


- Erster **UNESCO Lehrstuhl an einer Technischen Universität** in Österreich
- **Ziele:**
  - Zentrum für Forschungs- und Lehrkooperationen
  - Aktivitäten mit Partneruniversitäten, insbesondere im globalen Süden
  - Entwicklung von Lehrplänen und Forschungsrichtlinien basierend auf den Empfehlungen der UNESCO zur KI-Ethik
  - Förderung von Diversität in der Informatik, Wissenschaft und Industrie
- **Chair und Co-chair:** Peter Knees und Julia Neidhardt





# DigHum GPT

Explore the world of Digital Humanism with the help of Generative AI.

Interested in how this works? 

Not sure about what to ask? Here are some examples:

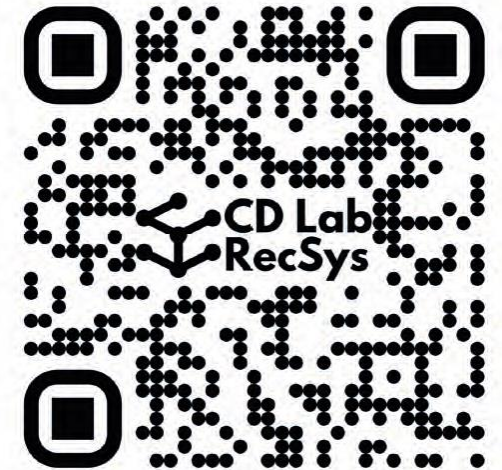
Describe the concept of digital humanism for a ten-year-old child.

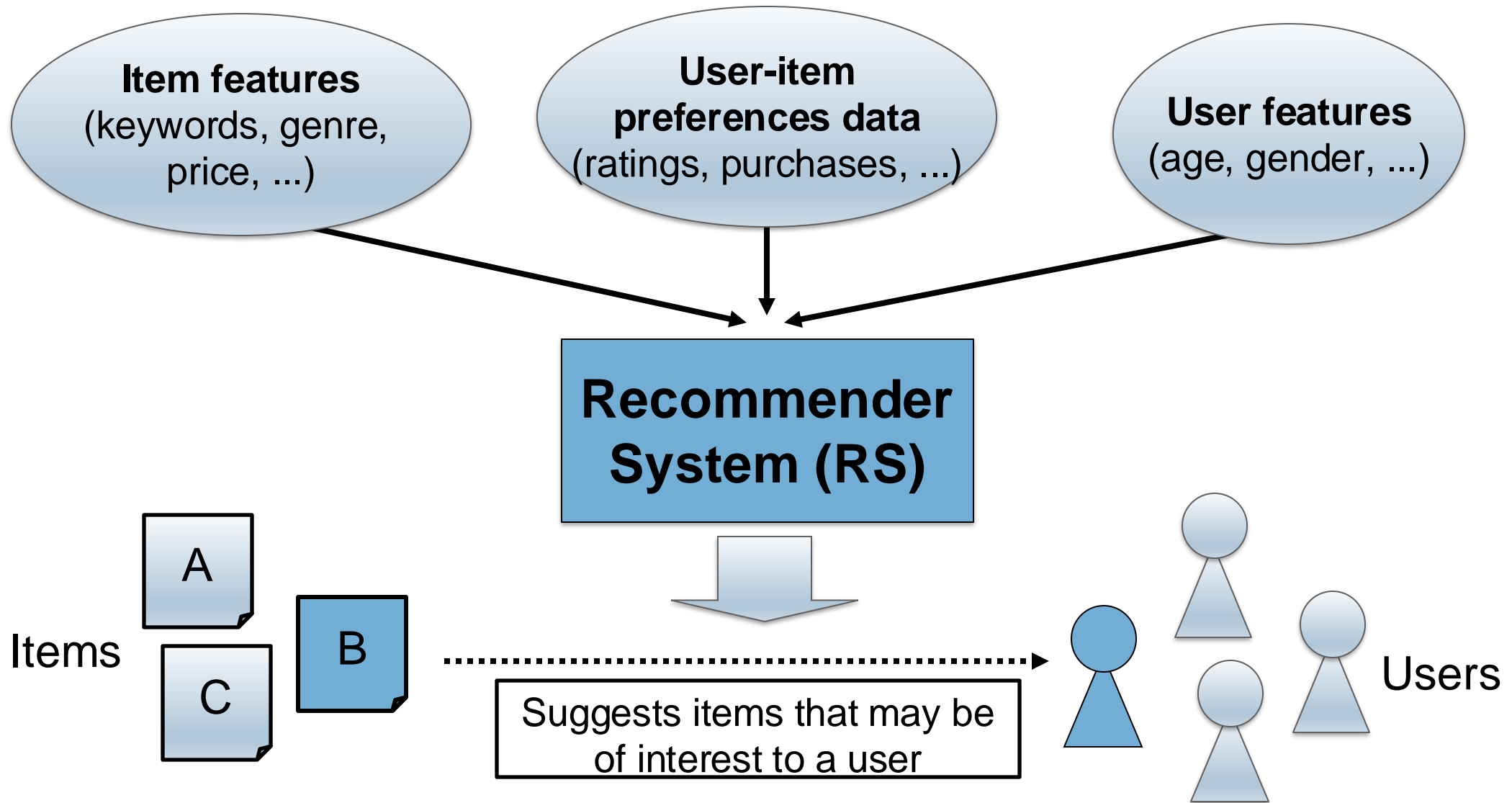
What is Digital Humanism about?

What are the key principles of Digital Humanism?



Welcome to the DigHum GPT App! The app is designed to help you with your questions about Digital Humanism. Feel free to ask me anything! Please do not add any personal data or sensitive information!






- **Bedeutung:** Recommender-Systeme sind heute unverzichtbare KI- und ML-Anwendungen in verschiedenen Bereichen.




# CDL-RecSys



## Module I


- Different stakeholders' perspectives and domain specifics
- **Objective:** Create adaptive, cross-domain recommenders using multi-modal content
- News Domain


FALTER MEDIA  
FALTER VERLAG DIE BESTEN SEITEN ÖSTERREICHS  
**derStandard.at**



## Module II

- Advanced user modeling with social context and item insights
- **Objective:** Enhancing shopping experience through conversational methods in groups
- Fashion and Lifestyle

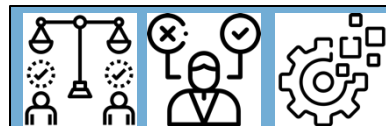
 **YouKnowMeBest**



## Module III

- Conversational approaches for preference elicitation
- **Objective:** Eliciting user preferences at various levels for tailored recommendations
- E-commerce, Price comparison

**Geizhals**  
Preisvergleich



### Additional Priorities:

Beyond-Accuracy, Fairness, Integration of LLMs, Digital Humanism





# Modul I: Schwerpunkte

- **Multi-Domain Perspektive:** Framework für Empfehlungen in einzelnen und mehreren Bereichen, darunter:
  - **Nachrichten:** Methoden für mehrere Stakeholder
  - **Podcasts** zu Nachrichten: Matching durch Text-/Audioanalyse
  - **Restaurants:** Vergleich von Basis- und fortgeschrittenen Methoden zur Erfassung latenter Merkmale
- **Verantwortungsvolle Empfehlungen:**
  - Fokus auf faire und vielfältige Nachrichtempfehlungen
  - Interdisziplinäre Studien zur Wirkung digitaler Anreize auf breiteren Nachrichtenkonsum und deren Einfluss auf das Wohlbefinden
- **Emotionale Dynamik und Klassifikation:**
  - Empfehlungsmodell berücksichtigt emotionale Hinweise in Nachrichten, verbessert die Leistung, reduziert jedoch Inhalts- und Emotionsvielfalt -> „Emotionsblase“-Effekt
  - Kontextsensitive Emotionsklassifikation kombiniert neue psychologische Konzepte mit fortgeschrittenen Berechnungsmethoden

**FALTER.at** ABONNIEREN EINLOGGEN

ZEITUNG ESSEN & TRINKEN EVENTS KINO KINDER BEST OF VIENNA SHOP SERVICE MEHR

**AKTUELLE AUSGABE**  
Klicken Sie hier, um den FALTER digital zu lesen »

**faltershop.at**  
1.5 Millionen Bücher, CDs und DVDs

**POLITIK**  
**Wie weit darf Israels Selbstverteidigung gehen? Zehn Antworten zur aktuellen Lage**  
Seit mehr als drei Wochen dauert nun der jüngste Krieg zwischen Israel und der Hamas. Zehn Antworten zur aktuellen...  
NINA BRNADA

**CHARIM**  
**Integration in diesen Zeiten**  
Insgesamt also ein Dilemma, aus dem keiner wirklich einen Ausweg weiß. Nicht nur hierzulande. Man schau nach...  
ISOLDE CHARIM

**FALTER.morgen**  
**Gaza-Krieg: Mehr Antisemitismus in Wien – und ein Lichtermeer für verschleppte Geiseln**  
Seit dem Terrorangriff der Hamas auf Israel steigt in Österreich der Antisemitismus – aber auch die Solidarität mit...  
FALTER.MORGEN VOM 03.11.2023

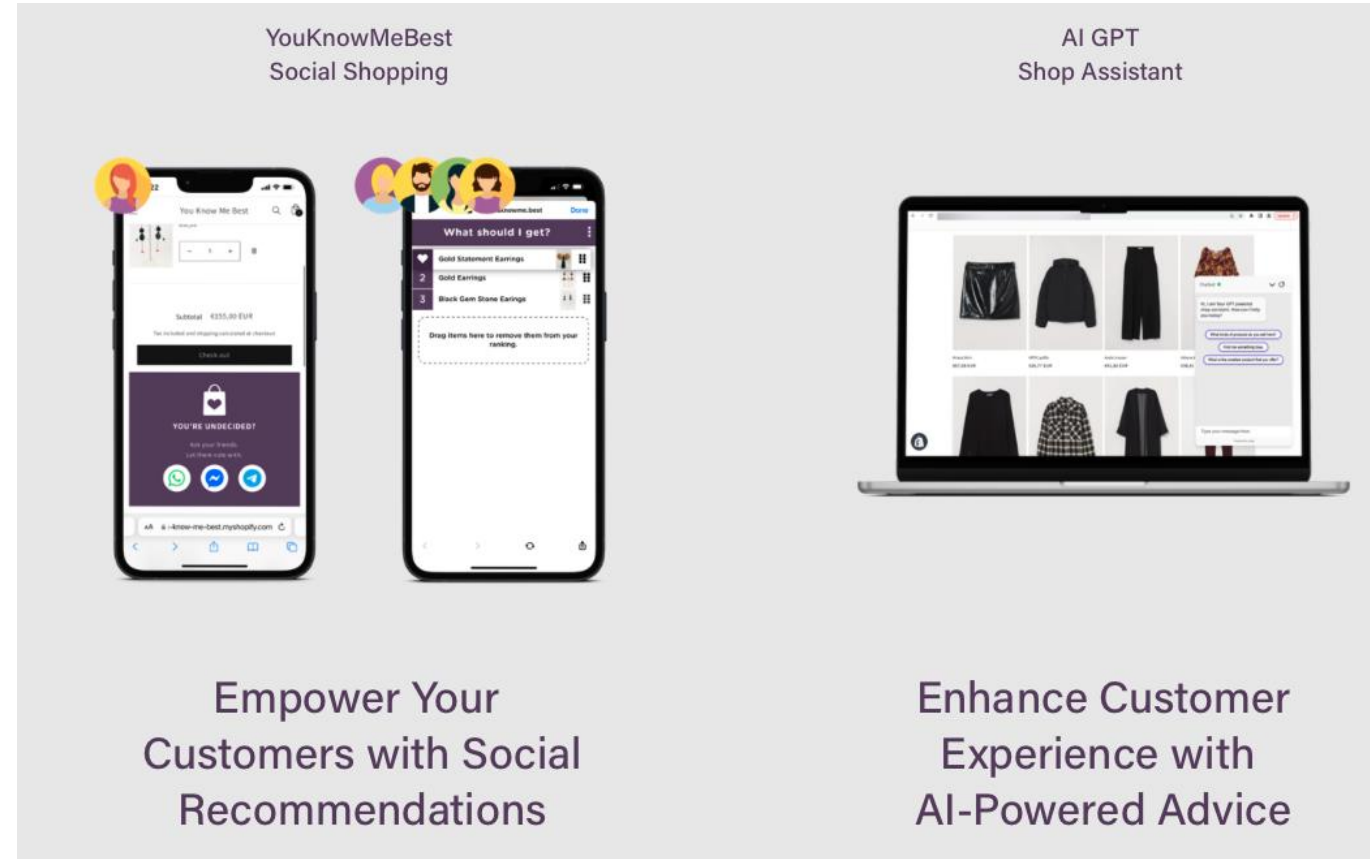
**FEUILLETON**  
**FALTER:WOCHE**

**FALTER MEDIA**  
**FALTER VERLAG**  
DIE BESTEN SEITEN ÖSTERREICHS



# Modul II: Schwerpunkte

- **Sozialsensitive Empfehlungen:**
  - System für sozial eingebundene Empfehlungen: Kombiniert **Echtzeit-Feedback aus dem sozialen Umfeld** mit Userpräferenzen für kontextbezogene Modeempfehlungen
  - **UX-Analyse:** Persönliche Merkmale wie Alter, Geschlecht und Chatbot-Erfahrung beeinflussen die Interaktion stark
- **Large Language Models (LLMs):**
  - **Integration:** LLMs verbessern Empfehlungssysteme, indem sie breite Erkenntnisse mit präzisen lokalen Daten kombinieren
  - **Anwendungspotenzial:** LLMs zeigen vielfältige Einsatzmöglichkeiten in verschiedenen Empfehlungsszenarien
  - **Nutzen für RSs:** LLMs optimieren Eingaben, Abruf und Personalisierung



 **YouKnowMeBest**



# Modul III: Schwerpunkte

- **Userintention erkennen**, Usermodelle erstellen und Empfehlungen generieren, um die Entscheidungsfindung während einer Konversation zu unterstützen
- **Potenziale von LLMs** zur Unterstützung der Empfehlungspipeline erkunden
- Untersuchen, wie die Präferenzermittlung über die Zeit **Bias und Ziele wie Diversität, Neuheit und Artikelabdeckung** beeinflusst
- Produktsuchen durch Konversationen vereinfachen und dabei potenzielle Verzerrungen verringern

The screenshot shows the Skinflint Price Comparison website. The navigation bar includes categories like Hardware, Phones, Video, Cameras & TV, Audio & HiFi, Home & Living, Health & Beauty, Sports & Leisure, DIY & Garden, Car & Motorcycle, Toys & Scale Models, and Games & Consoles. Below the navigation bar, there are several product category tiles, each with a list of sub-categories. For example, the 'Hardware' tile lists Graphics Cards, Monitors, Hard Drives & SSDs, Notebooks, Processors (CPUs), Motherboards, Input Devices, Cases/Enclosures, Tablets, Memory (RAM), Network, Heatsinks & Fans, Power Supplies & UPS Systems, and Water Cooling. The 'Top-10 products' section at the bottom displays four items: 1. AOC Q27G2E/BK, 27"; 2. AMD Ryzen 7 7800X3D, 8C/16T, 4.20-5.00GHz, boxed without cooler; 3. Lexar Professional NM710 1TB, M.2; 4. LG UltraGear 27GN800P-B, 27".

**Märkte:** Österreich, Deutschland, Großbritannien, Polen





# User-Sessions auf Preisvergleichsplattformen

- **Verstehen des Userverhaltens** bei der Interaktion mit Plattform
- Daten meist auf einzelne Sitzungen beschränkt
- **Ziel:** Personalisierte Empfehlungen unter Wahrung der Privatsphäre
- **Clustering** auf Basis von Sessions statt Usern



Quick Peek



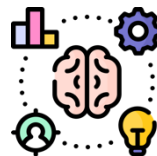
Constraint-based browsing



Query and Browse



Major Purchase



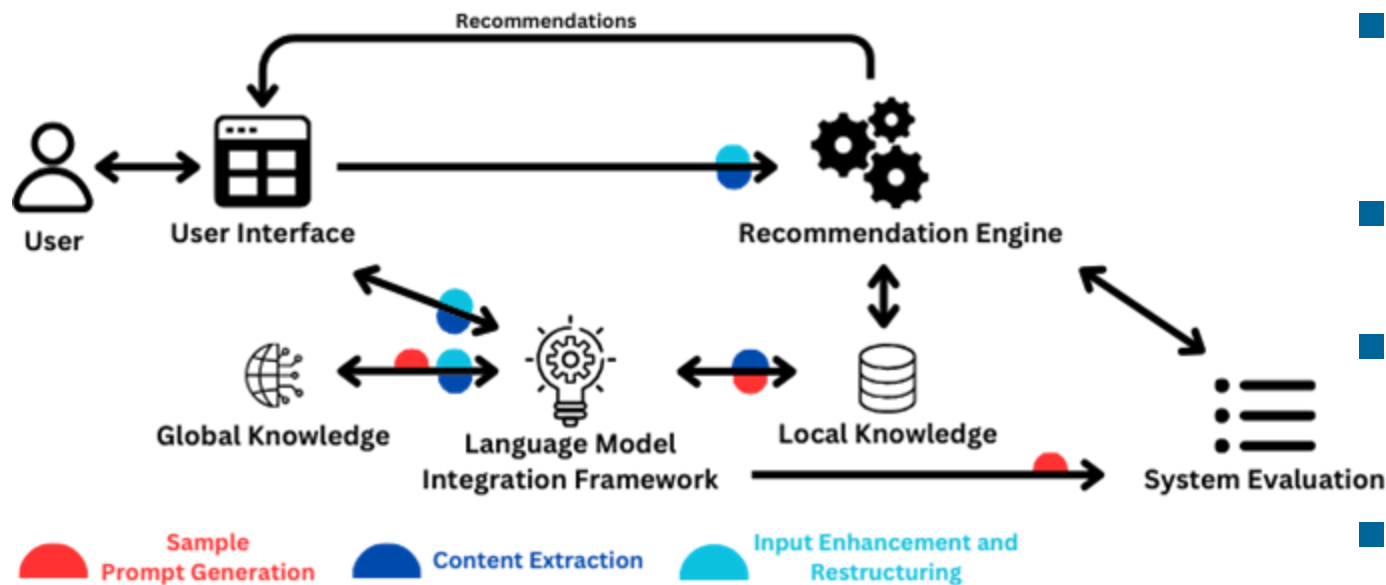
Knowledge Seeking



Heavy Browsing

A. Wagne and J. Neidhardt, "Potentials of Combining Local Knowledge and LLMs for Recommender Systems," in Proceedings of the 18<sup>th</sup> ACM Conference on Recommender Systems (Bari, Italy), 2024, forthcoming

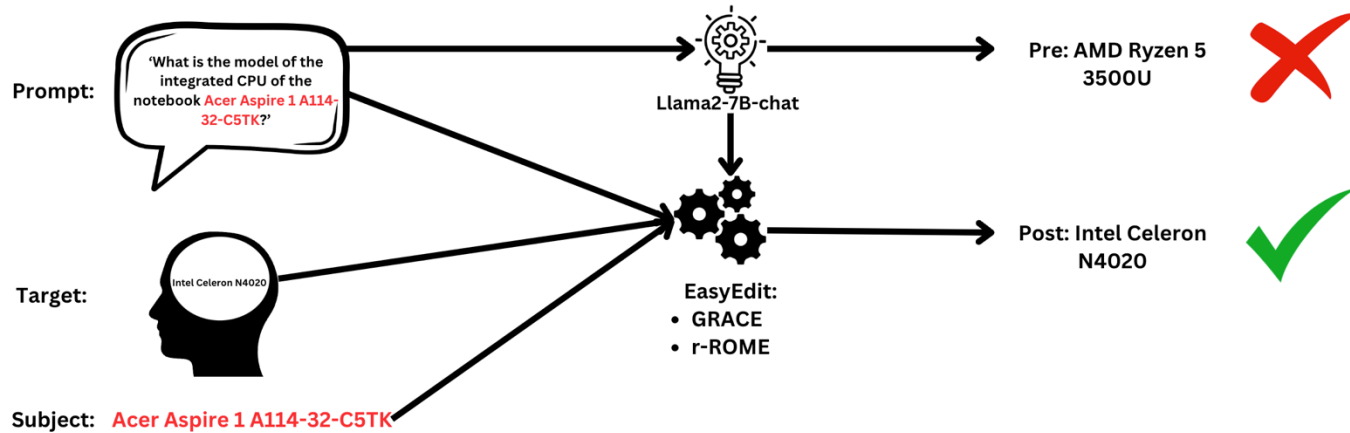
# Potentiale der Kombination von lokalem Wissen und LLMs für Recommender-Systeme



- **Kombination von LLMs und lokalem Wissen** verbessert RSs
- Nutzung von globalen und spezifischen Daten
- Lokales Wissen **reduziert Halluzinationen** und sorgt für präzisere Empfehlungen
- LLMs verbessern Anfragen, Eingaben und Personalisierung
- **Herausforderungen:** Prompt-Gestaltung, Integration von lokalem und globalem Wissen

Kolb, T. E., Wagne, A., Sertkan, M., & Neidhardt, J. 2023. *Potentials of Combining Local Knowledge and LLMs for Recommender Systems* (forthcoming). *Proceedings of the 5th Edition of Knowledge-Aware and Conversational Recommender Systems (KaRS) in Conjunction with ACM RecSys (2023)*.

# Produkte in KI integrieren: Wissensbearbeitung zur Abstimmung von LLMs mit Produktkatalogen



- Darstellung von Artikeln in dialogbasierten RSs
- Integration von lokalem Wissen in LLMs ohne erneutes Training durch Knowledge Editing
- Reale Produktkataloge sind dynamisch und begrenzt
- Vermeidung externer Datenabfragen während der Inferenz
- Präzise Darstellung von Produkteigenschaften im Modell

Method	Pre-Acc	Post-Acc	Edit time	Inference time
No edits	-	-	0	7,078
GRACE	0,022	0,21	22,942	7,000
r-ROME	0,449	0,868	24,953	5,800

A. Wagne, and J. Neidhardt, "Can we integrate items into models? Knowledge Editing to Align LLMs with Product Catalogs" in Proceedings of the 6th *Edition of Knowledge-Aware and Conversational Recommender Systems (KaRS) in Conjunction with ACM RecSys, 2024, forthcoming.*

# LLM-basiertes Conversational Recommender-System

## ■ Integration von LLMs in RSs:

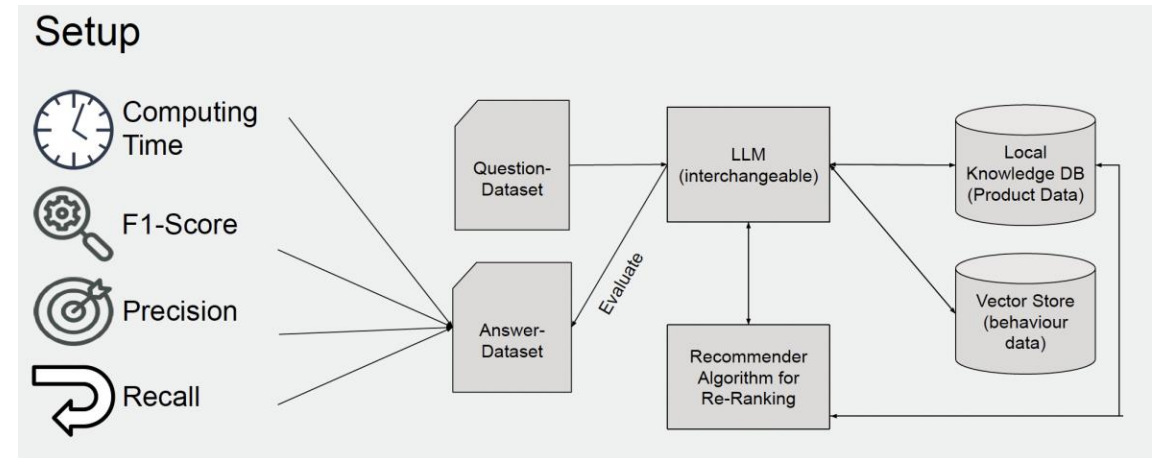
- Erste Implementierungsschritte zeigen, wie LLMs zur Personalisierung von Empfehlungen eingesetzt werden

## ■ Proof-of-Concept & User-Studie:

- Technische Machbarkeit demonstriert
- Erste Ergebnisse einer User-Studie bestätigen den erfolgreichen Einsatz von LLMs zur Verarbeitung von Userinteraktionen und zur Generierung kontextbezogener Empfehlungen

## ■ Verbesserte Nutzererfahrung:

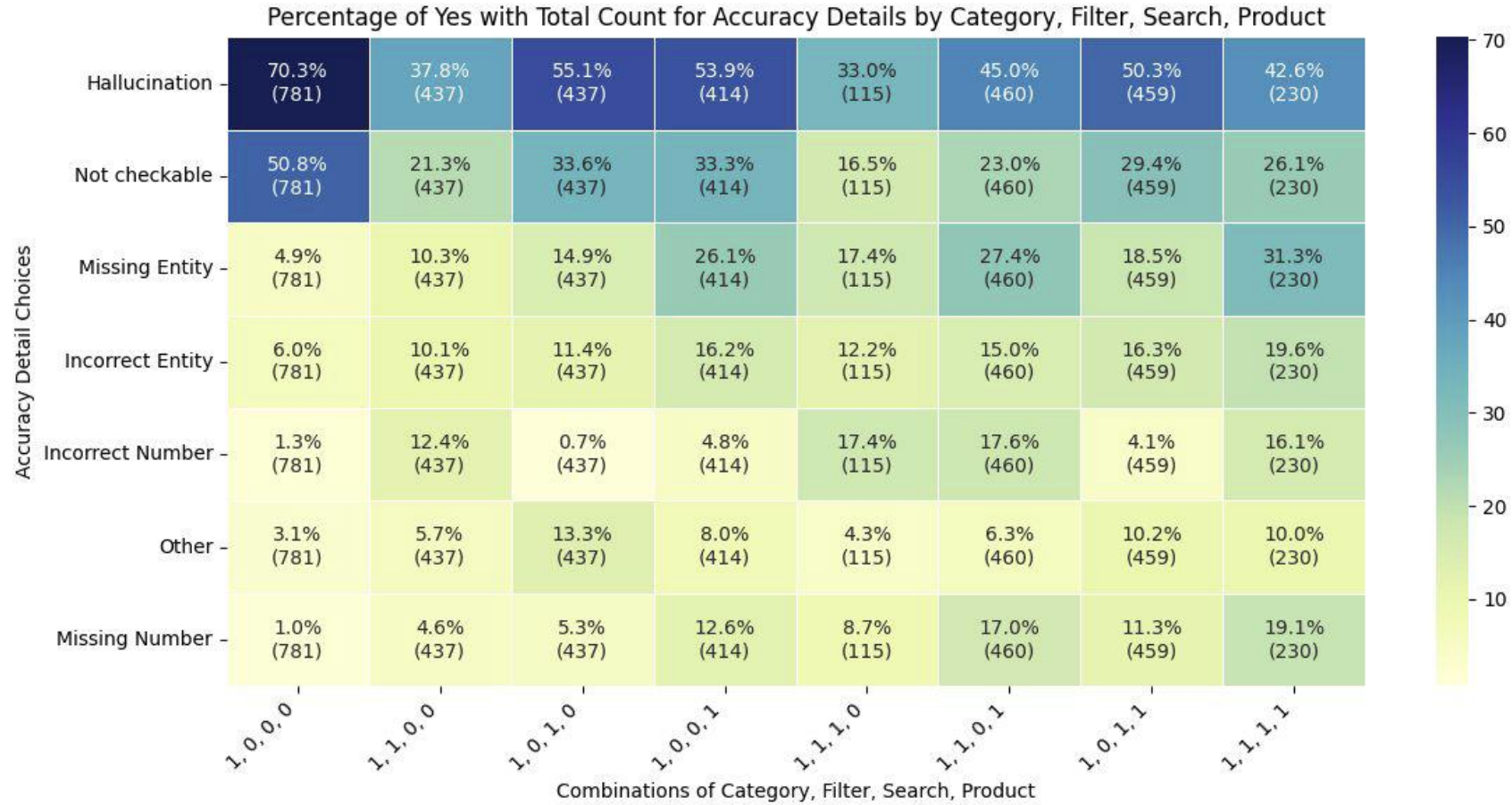
- Echtzeit-Vorschläge basierend auf Verhalten und Präferenzen
- Steigerung der Relevanz und Genauigkeit von Empfehlungen



Schmiedmayer, M. (laufend). The impact of using Large Language Models on the performance of Conversational Recommender Systems. CDL RecSys.



# LLM-basiertes Conversational Recommender-System (2)



Schmiedmayer, M. (laufend). The impact of using Large Language Models on the performance of Conversational Recommender Systems. CDL RecSys.





# Conclusion

- **Technologie mit menschlichen Werten:** Verbindung von RSs-Forschung und DigHum betont, dass RSs fair und inklusiv gestaltet werden sollten, um ethische Standards (Datenschutz, Inklusion) zu wahren - auch wenn nicht alle Forschung dies priorisiert
- **LLMs in RSs:** Integration von LLMs in RSs birgt sowohl Potenziale zur Verbesserung der Personalisierung als auch Risiken, wie Halluzinationen und mangelnde Transparenz
- **Multidisziplinäre Ansätze:** Enge Zusammenarbeit zwischen technischen und geisteswissenschaftlichen Disziplinen ist entscheidend, um ethische und gesellschaftliche Auswirkungen von KI zu verstehen und verantwortungsvoll zu gestalten
- **Menschzentrierte Innovation:** KI muss demokratische und ethische Werte fördern



**Julia Neidhardt**

Assistant Professor, TU Wien, Austria

Head of the CD Lab RecSys

UNESCO Co-Chair for Digital Humanism

[julia.neidhardt@tuwien.ac.at](mailto:julia.neidhardt@tuwien.ac.at)

<https://recsys-lab.at/>

# Christian Doppler Lab for Recommender Systems



**Julia Neidhardt**  
Head



**Irina Nalis**  
Post Doc



**Mete Sertkan**  
PhD Candidate



**Thomas Kolb**  
PhD Candidate



**Aayasha**  
PhD Candidate



**Ahmadou Wagne**  
PhD Candidate



**Michael Schmiedmayer**  
Student Employee



**Gwendolyn Rippberger**  
Student Employee



**Guilherme Oliveira**  
Student Employee



**Joanna Zamiechowska**  
Student Employee