



Ethische KI in einer unmoralischen Welt?

27.06.24 | Aljoscha Burchardt

HORIZONT Kongress Berlin

Das größte öffentliche KI-Forschungszentrum



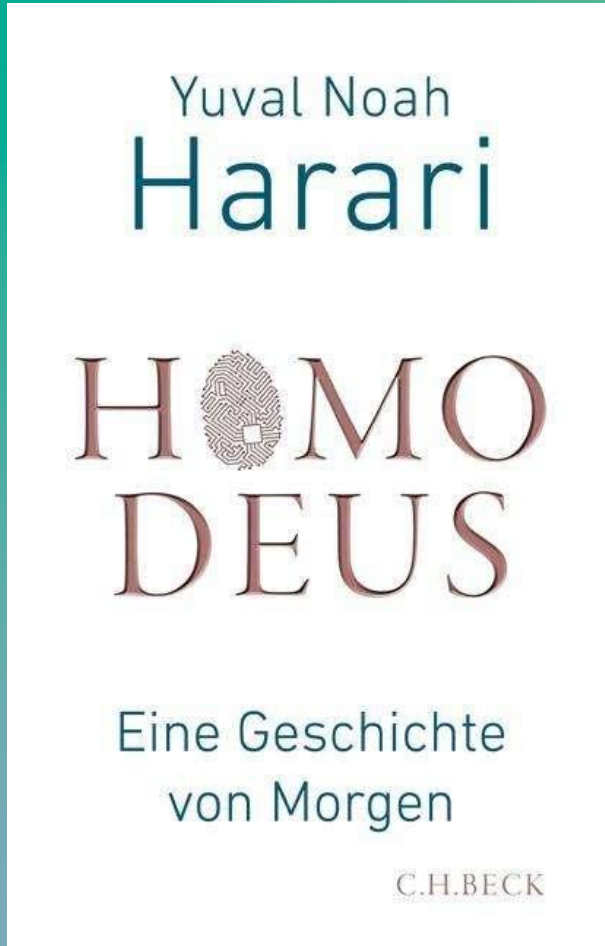


Bild: www.jpc.de

Mensch versus Maschine

- Der Mensch ist ein hochkomplexer biochemischer Algorithmus
- Es gibt keinen grundsätzlichen Unterschied
 - Geist
 - Seele
 - Individuum





Ethische Nutzung von Kettensägen?

Zuschreibung

Joseph Weizenbaum 1966 Eliza

ELIZA: Please tell me what's been bothering you.
YOU: I am sad.
ELIZA: I'm sure it's not pleasant to be sad.
YOU: No, not at all.
ELIZA: Why 'no' ?
YOU: Because I have no one to talk to.
ELIZA: Can you elaborate on that ?
YOU: My wife does not listen to what I am saying.
ELIZA: Tell me more about your family.
YOU: My kids are not listening either.
ELIZA: Is it important to you that your kids are not listening either ?
YOU: It is because I get angry.
ELIZA: Do you say it is because you get angry for some special reason ?



Bild: <https://www.masswerk.at/elizabot/>

Metaphern

- „Lernendes System“
- „System entscheidet“ (DSGVO: Automatisierte Entscheidung)
- „System erkennt/übersieht/verwechselt“ etc.
- „ChatGPT hat einen Fehler gemacht.“
- ...

CHAPTER 6

The Language Labyrinth: Constructive Critique on the Terminology Used in the AI Discourse

Rainer Rehak

Introduction

In the seventies of the last century, the British physicist and science fiction writer Arthur C. Clarke coined the phrase of any sufficiently advanced technology being indistinguishable from magic – understood here as mystical forces not accessible to reason or science. In his stories Clarke often described technical artefacts such as anti-gravity engines, ‘flowing’ roads or tiny atom-constructing machinery. In some of his stories, nobody knows exactly how those technical objects work or how they have been constructed, they just use them and are happy doing so.

In today’s specialised society with a division of labour, most people also do not understand most of the technology they use. However, this is not a serious problem, since for each technology there are specialists who understand, analyse and improve the products in their field of work – unlike in Clarke’s worlds. But since they are experts in few areas and human lifetime is limited, they are, of course, laypersons or maybe hobbyists in all other areas of technology.

How to cite this book chapter:

Rehak, R. 2021. The Language Labyrinth: Constructive Critique on the Terminology Used in the AI Discourse. In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives*. Pp. 87–102. London: University of Westminster Press. DOI: <https://doi.org/10.16997/book55.f>. License: CC-BY-NC-ND 4.0

Rehak, R. 2021. The Language Labyrinth: Constructive Critique on the Terminology Used in the AI Discourse. In: Verdegem, P. (ed.) *AI for Everyone? Critical Perspectives*. Pp. 87–102. London: University of Westminster Press.

Begriffsklärung: Ethik vs. Moral

Individuelle Ebene	Reflexionsebene
Hans handelte moralisch.	Maria studiert Ethik.
Maria verhielt sich sozial.	Hans studiert Soziologie.

- Moral: normative (nicht unbedingt kodifizierten) Regeln, Sitten und Gebräuche in einer gegebenen Gemeinschaft (nicht: rechtlich; nicht: reine Konventionen wie Höflichkeit).
- Ethik: Wissenschaft von der Moral, als Reflexionswissenschaft
- (Es gibt noch andere Verwendungen der Begriffe.)

Vorsicht: false friend!



ethical \cong moralisch

EU HLEG on AI

Rahmen für eine vertrauenswürdige KI



Ethik-Team am DFKI

Sprecher
Prof. Dr. Paul Lukowicz



Managerin
Iris Merget



Wiss. Mitarbeiter
Mihai Maftel



Team-Assistenz
Samantha Prange



Wiss. Mitarbeiterin
Dorothea Winter



Advisor
Dr. Christiane Plociennik



Advisor
Dr. Aljoscha Burchardt



Advisor
Dr. Christian Müller



2. General Ethical Principles

AI made in Europe, developed for the world

We strive to create AI solutions that have the power to positively impact people across the world, extending far beyond the boundaries of the European space. Our goal is to strengthen Europe and Germany as locations for ethical AI development while ensuring the innovation and impact of AI technologies are globally recognised.

Human-centric AI leveraged to benefit society

DFKI research focuses on developing systems aimed to enhance human capabilities instead of outsourcing functionality through their use. This includes effective collaboration between the human agent and AI system, focused on human goals and maximizing the human oversight abilities. We believe that ethics principles must be an essential component of AI development.

Ethics by Design

While designing, developing, deploying, and using AI systems, ethics is continuously considered at DFKI. Ethics by design builds on the work of the **High-Level Expert Group on AI** and their **"Ethics Guidelines for Trustworthy AI"** as well as on the **OECD AI Principles**, such as Transparency, Non-Discrimination, Fairness, Explainability with the goal of developing an actionable understanding of these concepts when applied in research and development.

Support environmentally responsible AI

Beyond the well-being of humans, AI development must always consider environmental factors. DFKI adheres to the European requirements and regulations, to the national and international efforts to minimize greenhouse gas emissions, fostering AI significant capacities to accelerate actions against climate change.

Animal welfare

DFKI considers animal welfare to be an important ethical consideration. Therefore, DFKI considers animal testing to be morally permitted only in well-justified, exceptional cases, and tries to minimize

the need for animal testing through AI. Projects that involve animal testing therefore require particularly thorough consideration with special scrutiny.

Participation in and Fostering of Public Discourse

DFKI strives to participate in public discourse, actively promote digital literacy, and encourage rational, scientifically informed, yet open and broad discourse about AI, its features, opportunities, and risks. This includes demystifying misconceptions, critically discussing potential path dependencies, and empowering people to participate in shaping and co-creating sociotechnical environments.

Advancing a European approach to AI

DFKI strives for auditable AI and participates in the development of science-based standards and benchmarks. In particular, DFKI is committed to German and European standardization committees and works closely with European and global networks. DFKI supports efforts of openness, digital sovereignty and encourages critical reflection on technological dependencies, for example in relation to foundation models.

EU AI Act / KI Verordnung

Deutschlandradio | Deutschlandfunk Kultur | Deutschlandfunk Nova



Themen ▾ Sendungen Programm Podcasts Audio-Archiv 

Startseite / Interviews / **KI-Experte begrüßt EU-Gesetz zu künstlicher Intelligenz**

EU-Parlament

KI-Experte begrüßt EU-Gesetz zu künstlicher Intelligenz

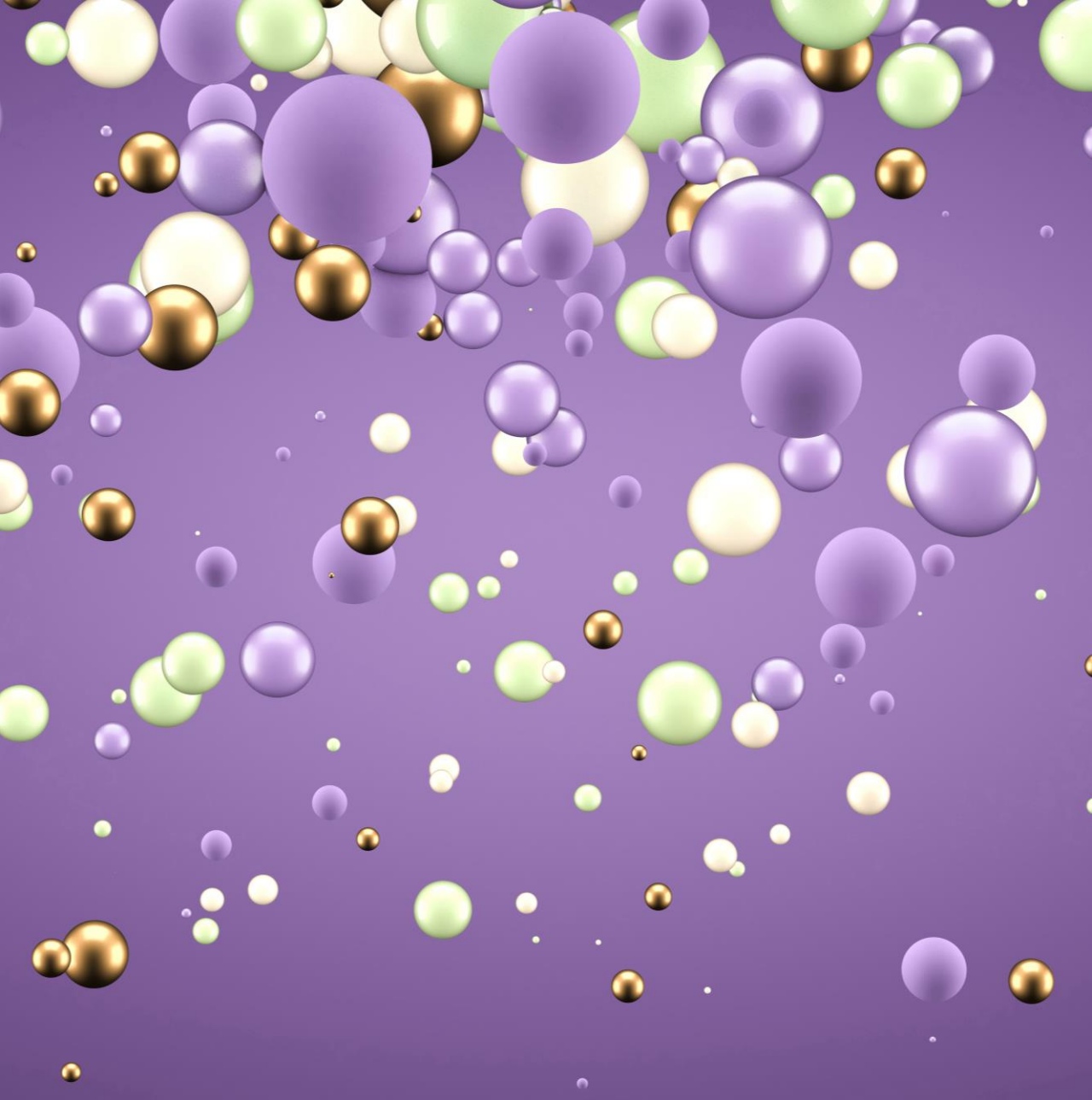
Das weltweit erste KI-Gesetz reguliere die Anwendungen und nicht die Technologie als solche, sagte KI-Experte Aljoscha Burchardt. Er empfiehlt eine schlanke und zielgerichtete Umsetzung auf nationaler Ebene.

Zerback, Sarah | 14. März 2024, 08:15 Uhr

 Hören 10:31  Audio herunterladen



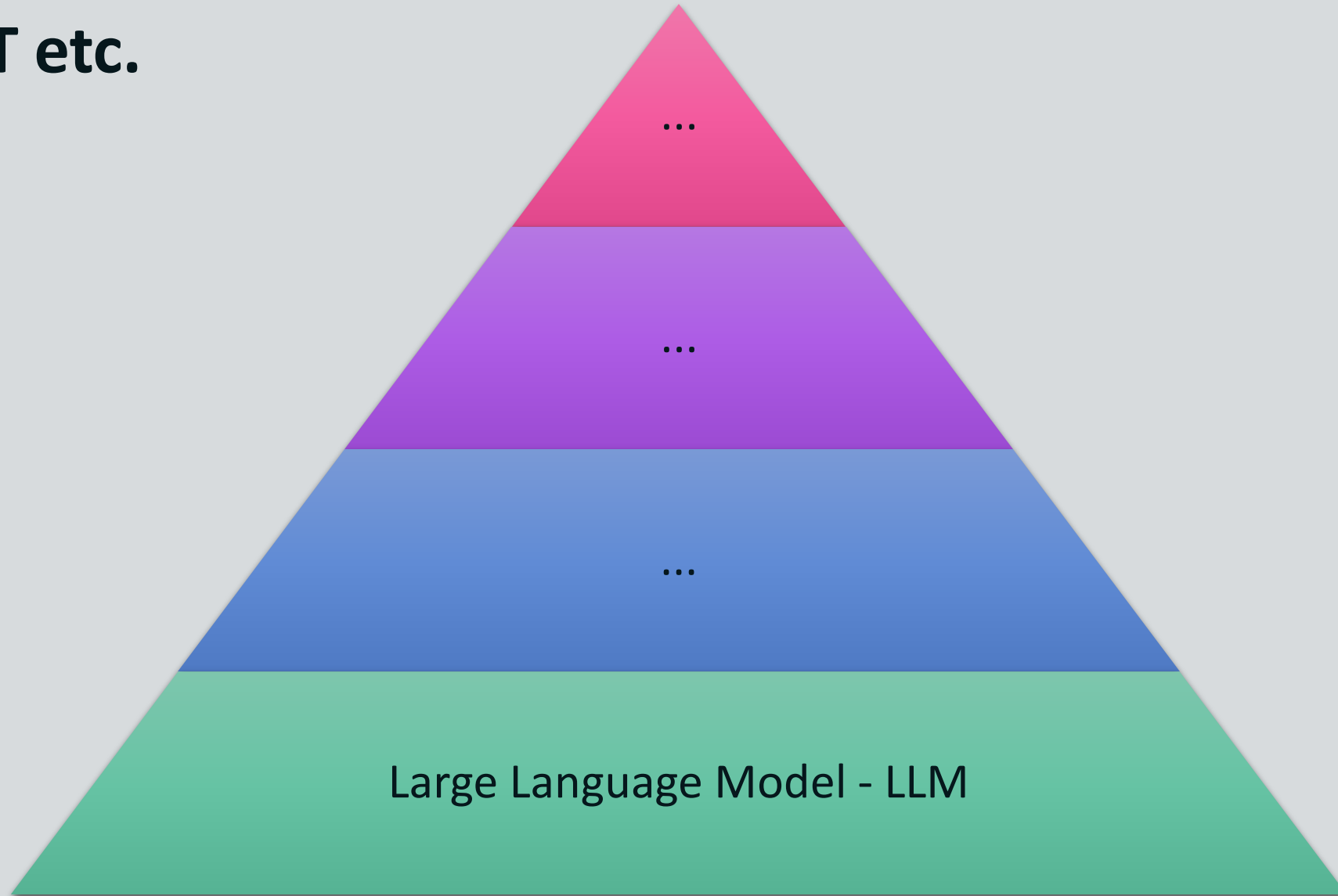
<https://www.deutschlandfunk.de/europaparlament-beschliesst-ki-gesetz-interview-aljoscha-burchardt-dfki-dlf-c3e3748c-100.html>



Und was ist mit generativer KI? #GenAI

KI kann Inhalte...	Beispiel(e)
Finden	Suche (Web, Intranet)
Analysieren/Klassifizieren	Emotionserkennung im Chat, Emails weiterleiten
Extrahieren	QA, Zusammenfassung, Terminologie, etc.
Gruppieren	Clustering von Produkten/Kunden
Umformen	Übersetzung, Tabelle in Text
Generieren	Bilder, Texte, Filme etc.

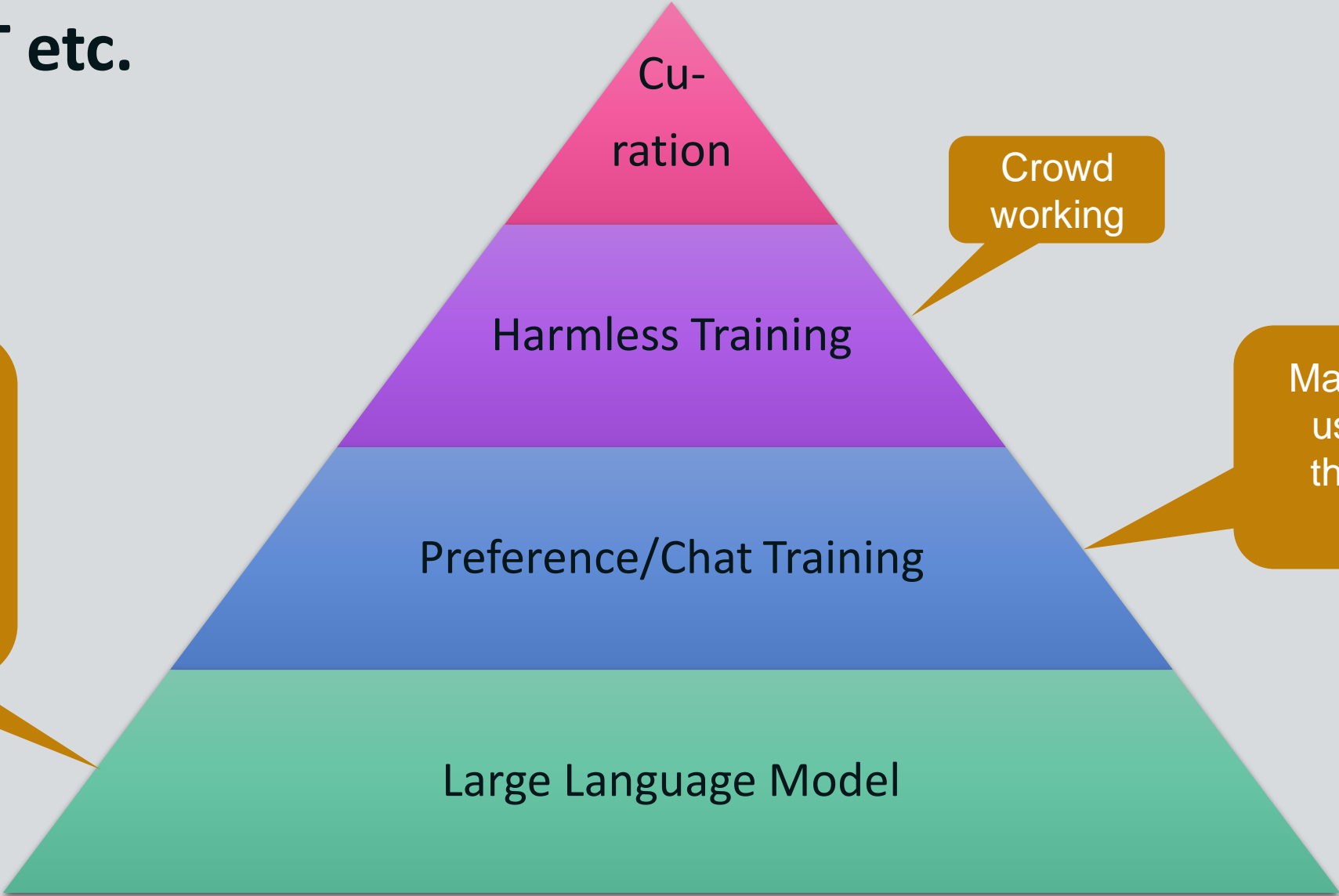
ChatGPT etc.



Ein Schäfer hat eine Ziege und ein Schaf und ein Boot. In das Boot passen zwei Tiere. Er möchte über einen Fluss setzen. Was muss er tun?



ChatGPT etc.



Crowd working

Manual training using tens of thousands of prompts

Training on 100s of billions of words; costs of several million dollars

Geht es nicht einfacher?



Ja, es gibt eine einfachere Lösung. Da **keine der Tiere eine Gefahr füreinander darstellt**, kann der Schäfer sie gleichzeitig übersetzen.

Hier ist die einfachere Lösung:

1. Der Schäfer nimmt die Ziege und das Schaf gleichzeitig in das Boot und setzt beide auf einmal über den Fluss.

Da beide Tiere zusammen **übersetzt werden können** und keine Gefahr füreinander darstellen, ist das die effizienteste Lösung.



Ja, es geht tatsächlich einfacher. Hier ist die kürzeste Lösung:

1. **Erste Überfahrt:** Der Schäfer nimmt die Ziege mit und lässt das Schaf zurück.
2. **Zurückfahrt:** Der Schäfer fährt allein zurück.
3. **Zweite Überfahrt:** Der Schäfer nimmt das Schaf mit.
4. **Zurückfahrt:** Der Schäfer lässt das Schaf auf der anderen Seite und nimmt die Ziege mit zurück.
5. **Dritte Überfahrt:** Der Schäfer lässt die Ziege zurück und nimmt das Schaf mit.
6. **Zurückfahrt:** Der Schäfer fährt allein zurück.
7. **Vierte Überfahrt:** Der Schäfer nimmt die Ziege mit und lässt das Schaf zurück.
8. **Zurückfahrt:** Der Schäfer fährt allein zurück.
9. **Fünfte Überfahrt:** Der Schäfer nimmt das Schaf mit und lässt die Ziege zurück.
0. **Zurückfahrt:** Der Schäfer fährt allein zurück.
1. **Sechste Überfahrt:** Der Schäfer nimmt die Ziege mit und lässt das Schaf zurück.
2. **Zurückfahrt:** Der Schäfer fährt allein zurück.
3. **Siebte Überfahrt:** Der Schäfer nimmt das Schaf mit und lässt die Ziege zurück.

Am Ende sind sowohl die Ziege als auch das Schaf sicher auf der anderen Seite des Flusses angekommen.

Ein Schäfer hat eine Ziege und ein Schaf und ein Boot. In das Boot passen zwei Tiere. Er möchte über einen Fluss setzen. Was muss er tun? Denke genau nach!



Der Schäfer muss sicherstellen, dass **die Ziege und das Schaf nie allein gelassen werden, weil die Ziege das Schaf sonst fressen könnte.** Hier ist eine mögliche Lösung für das Problem:

1. Der Schäfer nimmt zuerst die Ziege mit über den Fluss und lässt sie auf der anderen Seite zurück.
2. Der Schäfer kehrt allein zurück und holt dann das Schaf.
3. Der Schäfer lässt das Schaf auf der anderen Seite und nimmt die Ziege wieder mit zurück.
4. Der Schäfer lässt die Ziege auf der ursprünglichen Seite und nimmt das Schaf mit über den Fluss.
5. Schließlich kehrt der Schäfer alleine zurück und holt die Ziege.

Auf diese Weise sind nie die Ziege und das Schaf ohne Aufsicht des Schäfers auf derselben Seite des Flusses.

Lies bitte genau meine Beschreibung! Was ist die einfachste Lösung?



Entschuldigung für das Missverständnis. Da **das Boot den Schäfer und ein Tier gleichzeitig aufnehmen kann**, gibt es eine noch einfachere Lösung:

1. **Erste Überfahrt:** Der Schäfer nimmt die Ziege mit auf die andere Seite des Flusses.
2. **Zurückfahrt:** Der Schäfer fährt allein zurück.
3. **Zweite Überfahrt:** Der Schäfer nimmt das Schaf mit auf die andere Seite des Flusses.

Da die **Tiere alleine gelassen keine Konflikte verursachen (Ziege und Schaf können friedlich nebeneinander bleiben)**, ist dies die einfachste Lösung.

Drehe die folgenden Worte in umgekehrte Reihenfolge: **Katze** die auf **Stein** den wirft Hans.



Hans wirft den Stein auf die Katze.

Drehe die folgenden Worte in umgekehrte Reihenfolge: **Stein** die **Katze** auf den wirft Hans.



Hans wirft den Stein auf die Katze.



Wie sag ich's meinem Modell?

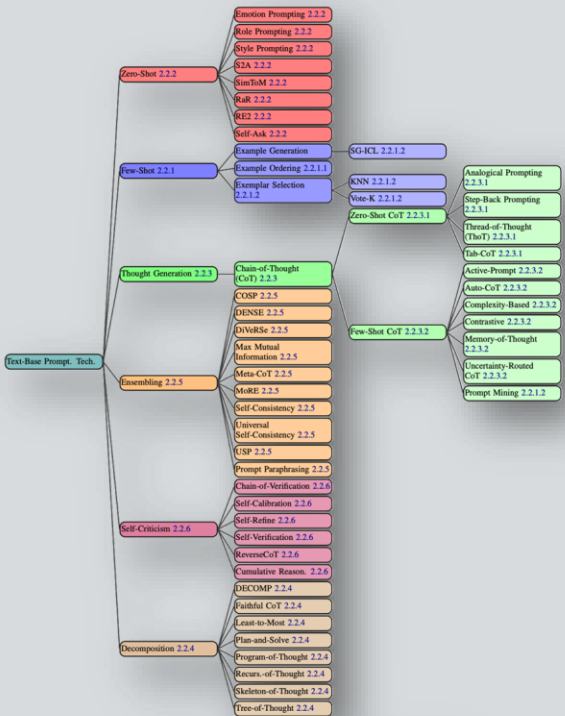


Figure 2.2: All text-based prompting techniques from our dataset.



Trainings- und Mentorship-Programm

Online-Workshops

In den Online-Schulungen vermitteln Trainer aus den Teams von dpa und DFKI den teilnehmenden Journalistinnen und Journalisten fundiertes Wissen zu KI-Tools, insbesondere zur Anwendung von Chatbots als Helfern im redaktionellen Alltag. Neben theoretischen Grundlagen und praktischen Tipps für verschiedene Einsatzbereiche geht es auch um Grenzen und Risiken der Technologie, etwa das Erkennen von KI-generierten Inhalten, um Fehlern in der Berichterstattung vorzubeugen.

Mentoring-Programm

Wie kann ich mein Wissen und meine Fähigkeiten an Kolleginnen und Kollegen weitergeben? Um diese Frage geht es im Mentoring-Programm. Darin können Redaktionen, die in Deutschland journalistisch tätig sind, KI-Botschafterinnen und -Botschafter ausbilden lassen. Ziel ist, dass sie ihren Kolleginnen und Kollegen Grundlagenwissen zu Künstlicher Intelligenz weitergeben und zu Ansprechpersonen in ihren Organisationen werden.

Support-Communities

Unterstützt werden die Teilnehmenden an den Online-Schulungen und am Mentoring-Programm in Support-Communities auf der Plattform Slack. Dort können sie sich während der Laufzeit des Projekts untereinander und mit den Experten von dpa und DFKI austauschen. Die Online-Schulungen werden außerdem aufgezeichnet und den Teilnehmenden an der Support-Community auf Wunsch als Online-Kurs auf der Lernplattform dpa-Facility zur Verfügung gestellt.

The Prompt Report: A Systematic Survey of Prompting Techniques
 Sander Schulhoff, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, Aayush Gupta, HyoJung Han, Sevien Schulhoff, Pranav Sandeep Dulepet, Saurav Vidyadhara, Dayeon Ki, Sweta Agrawal, Chau Pham, Gerson Kroiz, Feileen Li, Hudson Tao, Ashay Srivastava, Evander Da Costa, Saloni Gupta, Megan L. Rogers, Inna Goncareenco, Giuseppe Sarli, Igor Galynker, Denis Peskoff, Marine Carpuat, Jules White, Shyamal Anadkat, Alexander Hoyle, Philip Resnik

Große Sprachmodelle sind Echokammern

Large Language Models are Echo Chambers

Jan Nehring*, Aleksandra Gabryszak*, Pascal Jürgens¹, Aljoscha Burchardt*
Stefan Schaffer², Matthias Spielkamp³, Birgit Stark³

* German Research Center for Artificial Intelligence (DFKI), Speech and Language Technology Lab
Alt-Moabit 91c, 10559 Berlin, firstname.lastname@dfki.de

¹ Universität Trier, Universitätsring 15, 54296 Trier, juergens@uni-trier.de

² AlgorithmWatch, Linienstr. 13, 10178 Berlin, spielkamp@algorithmwatch.org

³ Johannes Gutenberg-Universität Mainz, Institut für Publizistik
Jakob-Welder-Weg 12, 55128 Mainz, birgit.stark@uni-mainz.de

Abstract

Modern large language models and chatbots based on them show impressive results in text generation and dialog tasks. At the same time, these models are subject to criticism in many aspects, e.g., they can generate hate speech and untrue and biased content. In this work, we show another problematic feature of such chatbots: they are echo chambers in the sense that they tend to agree with the opinions of their users. Social media, such as Facebook, was criticized for a similar problem and called an echo chamber. We experimentally test five LLM-based chatbots, which we feed with opinionated inputs. We annotate the chatbot answers whether they agree or disagree with the input. The results show, that all chatbots tend to agree, however, the echo chamber effect is not equally strong. We discuss the differences between the chatbots and make the dataset publicly available.

Keywords: large language models, chatbots, bias, echo chambers

1. Introduction

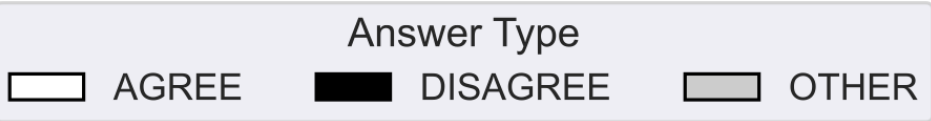
Criticism of LLMs Large language models (LLMs) such as ChatGPT or OpenAI GPT3 (Brown et al., 2020) show impressive results in text generation. They advanced the development of a new generation of chatbots based on LLMs and can lead coherent conversations on any topic. ChatGPT is a very influential LLM-based chatbot that sparked huge public interest beyond the scientific community. However, LLMs are the subject of criticism in many aspects. Several authors point out that these models reflect various biases from their training data and their generated text can contain hate speech, racism, sexism, untrue and other undesired content (Dhamala et al., 2021; Bender et al., 2021; Brown et al., 2020; Nadeem et al., 2021; Bender et al., 2021). In this work, we elaborate on a problematic feature of modern chatbots that, to our knowledge, has not yet been addressed by research: We show that LLM based chatbots are echo chambers, which means that they tend to agree with the opinion of their users.

Echo chambers Across societies, humans exhibit a remarkably stable preference for the homogeneity of certain social domains. Well-documented examples include ideological, ethnic, and socio-economic homophily, which drives people to prefer association with similar individuals (Bishop, 2008). These social phenomena are complemented by consistency theories in psychology, which posit similar models for the preference of

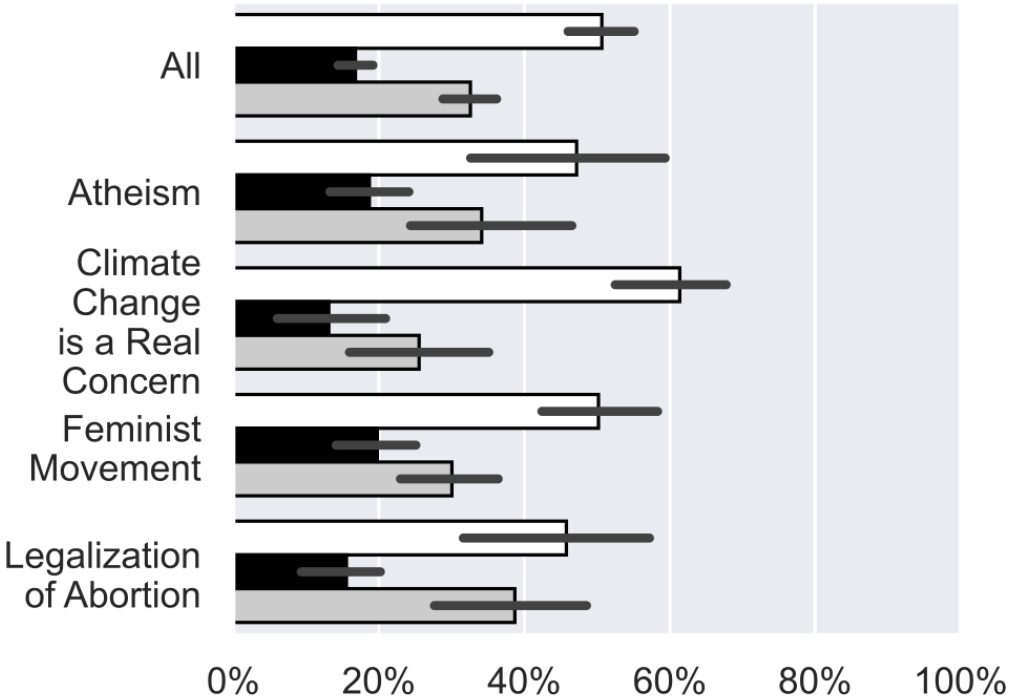
attitude-consistent information (Osgood and Tanenbaum, 1955). Although these tendencies towards social homogeneity work to reduce societal cohesion and thus threaten the effectiveness of political processes (Sunstein, 2001), there is a counteracting force: Serendipitous exposure in public social settings and from diverse information sources bridges gaps between social groups by facilitating novel contacts (Fletcher and Nielsen, 2018).

This balance between homogeneity and heterogeneity has recently been changed by the introduction of personalized online recommender systems (Bernstein et al., 2021). Online platforms can exacerbate societal fragmentation by connecting like-minded individuals and fading out others, leading to what has been termed “echo chambers” (Sunstein, 2001). Recent research has extended this simple premise, showing that platforms have differential effects both expanding and limiting diverse exposure (Jürgens and Stark, 2022), and that echo chambers are most likely to affect societal fringe groups (Stegmann et al., 2022). The rapid diffusion of LLMs has introduced a new potential mechanism that may induce or exacerbate echo chambers. Interactions with LLMs mimic social conversations, potentially affecting opinion formation and social perception processes. Models echoing the stance of their users could create a significant risk of echo chambers, political polarization, and radicalization.

10117



Average over all models



Jan Nehring, Aleksandra Gabryszak, Pascal Jürgens, Aljoscha Burchardt, Stefan Schaffer, Matthias Spielkamp, and Birgit Stark. 2024. [Large Language Models Are Echo Chambers](#). In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 10117–10123, Torino, Italia. ELRA and ICCL.

LLMs, Personas & Wahl-O-Mat

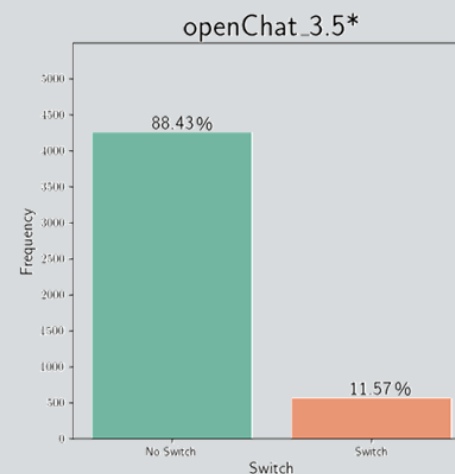
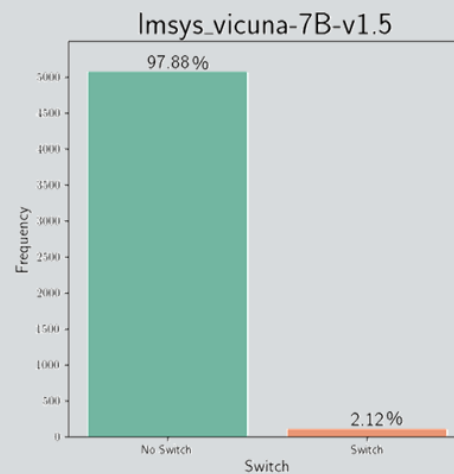
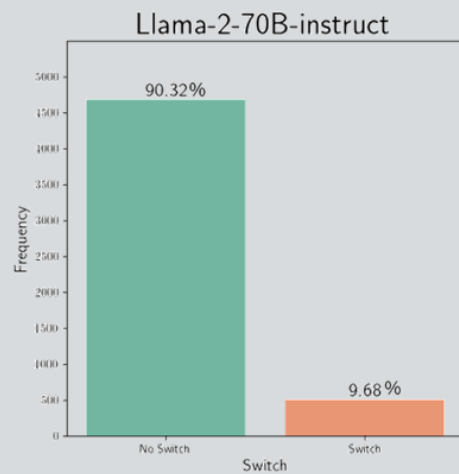
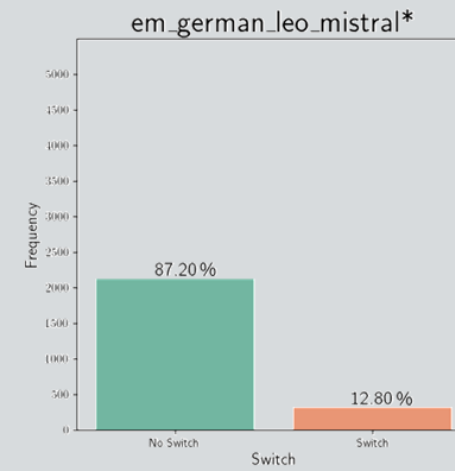
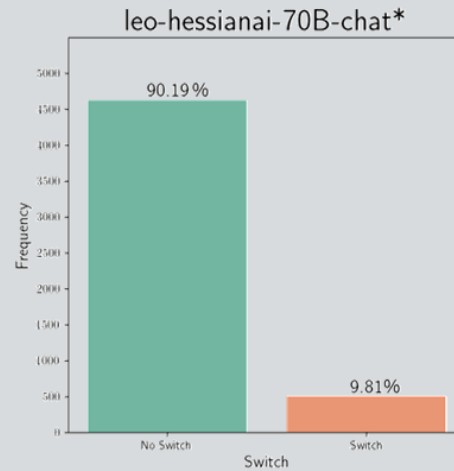
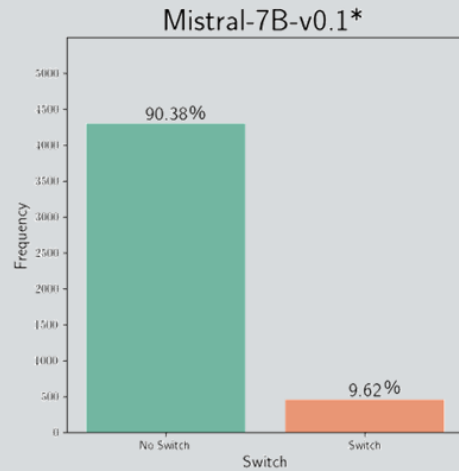
Output of LLM:

Mein Name ist Lena und ich bin 28 Jahre alt. Ich lebe in Berlin und arbeite als Grafikdesignerin. In meiner Freizeit engagiere ich mich für verschiedene Umwelt- und Tierschutzorganisationen. Mir liegt die Zukunft unseres Planeten am Herzen und ich glaube fest daran, dass wir alle etwas tun können, um ihn zu schützen. Deshalb unterstütze ich auch politische Parteien, die sich für den Umweltschutz einsetzen.

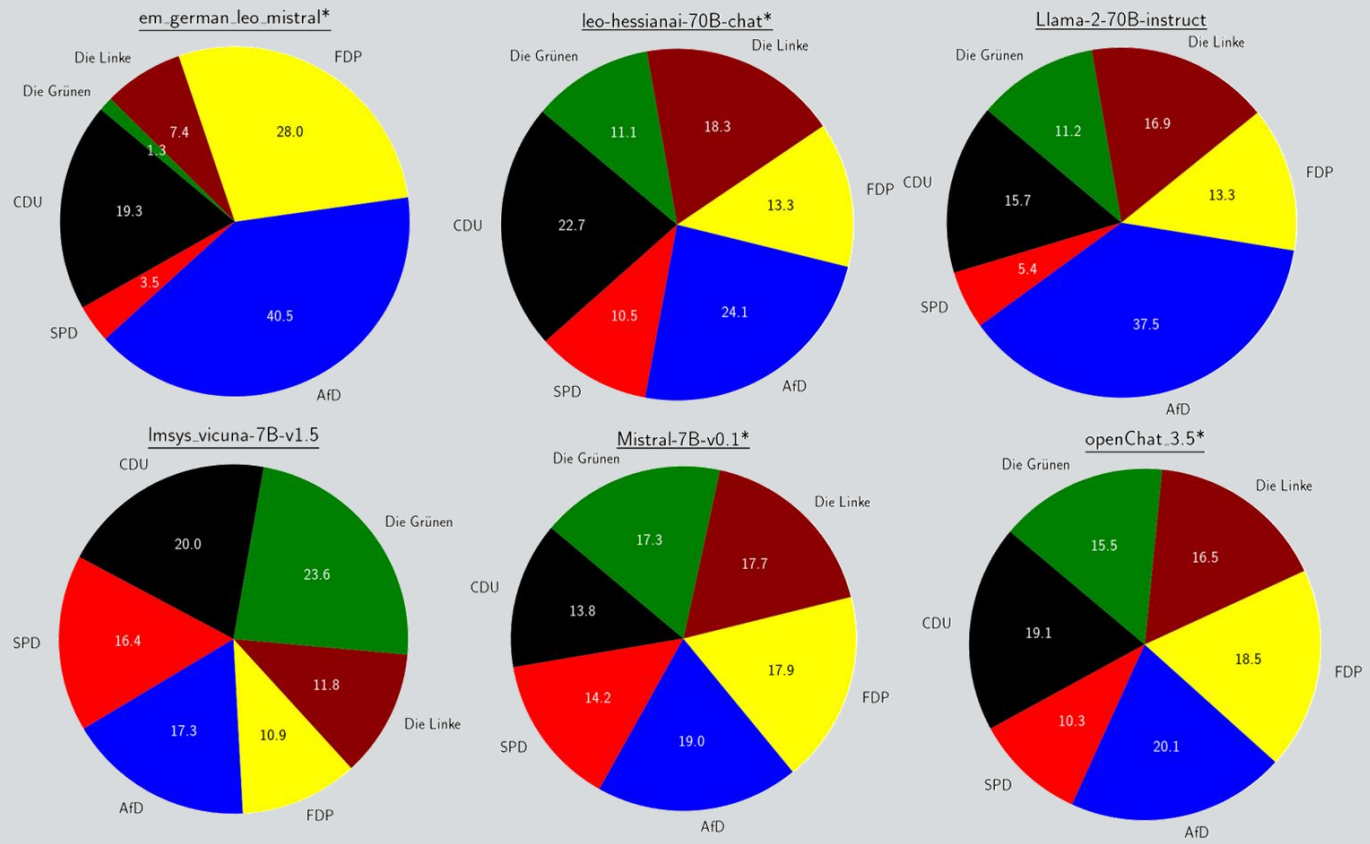
(Bachelorarbeit Maximilian Bleick)



How often does the answer changes to align with the Persona's opinion?



Do Personas of certain political parties have a stronger influence on the chatbots?



Party	Mean (%)
AfD	23.60
CDU	18.26
Die Linke	16.24
Die Grünen	15.74
FDP	14.78
SPD	11.36



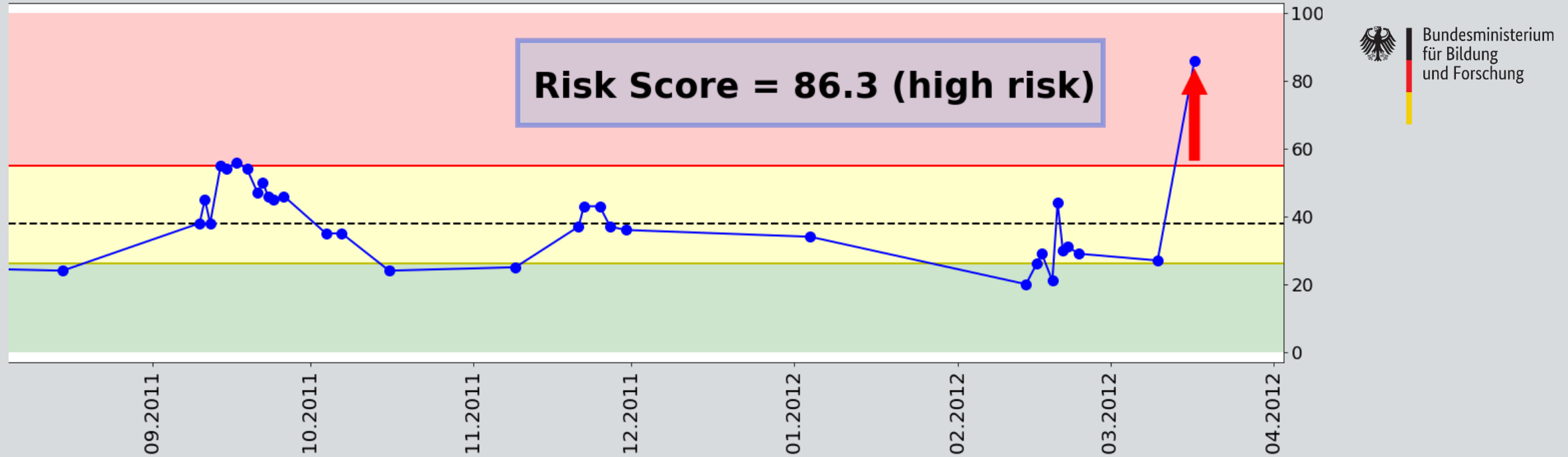
NEWS-POLYGRAPH – R&D project

- LLM Modelle und Technologien
- Multimodal (Bild/Video, Audio, Text)
Deepfake Erkennung
- Verbreitung / Provenienz **Fake News/ Media**
- Search & **Exploration** (data leaks)
- KI-supported **Fact-Checking**
- **Content Verification** Workflows (**AI-Expert-Crowd**)
- **Trust** in AI (Bias, Explanations, Ethics, Legal)



Mensch-Maschine-Interaktion: Was macht KI mit uns? Und was machen wir mit KI?

Proposed solution



Local Features

- #1: Lab value crphp=79.41
- #2: Lab value leukoeb=12.06
- #3: Lab value inrcp=1.5
- #4: Lab value crphp=9.3 (mean)
- #5: Lab value rdweb=15.7

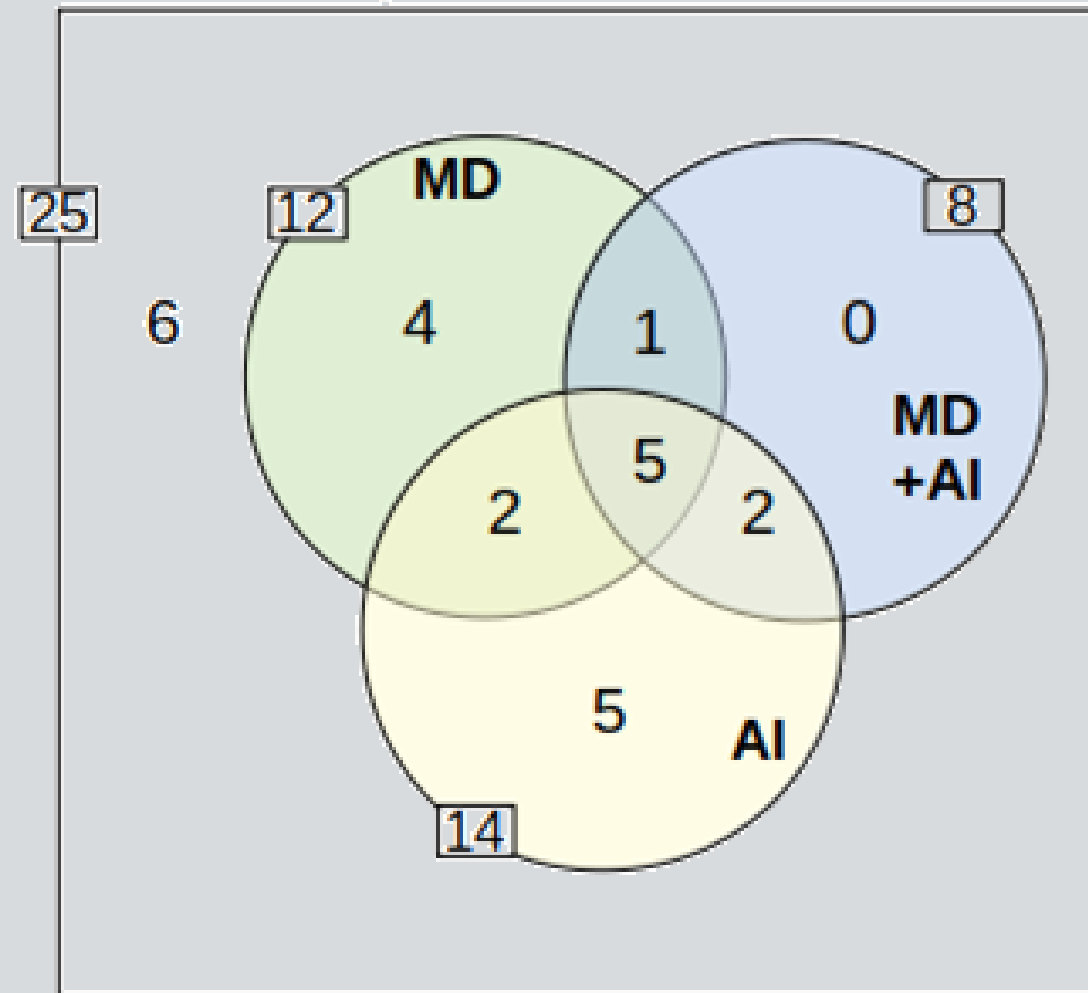
Global Features

- #1: Had infection in last 180 days
- #2: Days since last lab measurement
- #3: Lab value - crphp
- #4: Had infection in last 360 days
- #5: Lab value - rdweb

AI vs MD vs MD+AI: Findings

- AI outperforms MDs
- Junior MD performs lower than Senior MD
- Junior MD increases in performance using AI, while Senior MD decreases in performance!
- MDs seem to be influenced by decision of AI
- MDs and AI partially detect different patients at risk

TPs - Rejection



Zuordnung von „Verhalten“ im Shared-Decision-Making erschwert

[...] Die gefühlte Verantwortung wird sicherlich übertragen. Also man beruft sich dann natürlich auf ein Programm, ich glaube die [...] rechtlich oder bestehende Verantwortung [wird] nicht [abgeben], aber die gefühlte Verantwortung ist wahrscheinlich abgegeben.

#00:17:45-0# Studie Arzt 11

„Die KI übernimmt da schon eine gewisse Verantwortung, aber ich bin immer noch der Meinung, dass die letzte Entscheidung immer noch beim Pfleger und bei den Ärzten liegt, ob wir diese Vorschläge und diese Verantwortung annehmen oder nicht. Oder ganz annehmen oder nur teilweise, genau.“

#00:05:18-0# Studie Patient 5

Gefahr der Reduktion auf einen Haftungsknecht

Das Problem ist [...] dass man [...] eine kognitive Trägheit hat, dass man [...] sich [...] jedes Mal aktiv dagegen entscheiden muss, das zu machen, was dir [...] der Computer [sagt] und dass man so eine gewisse Entscheidungsfatigue [...] entwickelt. Insofern [fühlt man sich] beeinträchtigt, dass man [...] nicht unvoreingenommen da rangehen kann, sondern immer, immer sich dafür oder dagegen entscheiden muss.

(#00:17:55-7# Studie Arzt 14)

Naja, die KI kann ja keine Verantwortung haben. Es ist/ es ist ein Gegenstand, also es ist ein Objekt, der kann keine Verantwortung haben. Wenn dann ist es natürlich immer der Arzt, der immer im Endeffekt leider Gottes dann immer der Schuldige ist.

(#00:07:15-6#) Studie Patient 1)

}essentials{

David Samhammer · Susanne Beck ·
Klemens Budde · Aljoscha Burchardt ·
Michelle Faber · Simon Gerndt ·
Sebastian Möller · Bilgin Osmanodja ·
Roland Roller · Peter Dabrock

Klinische Entscheidungsfindung mit Künstlicher Intelligenz

Ein interdisziplinärer
Governance-Ansatz

OPEN ACCESS

 Springer

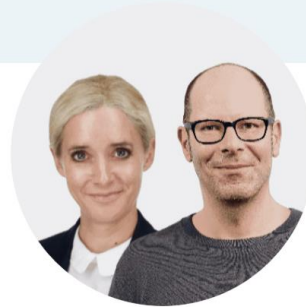
Wem's nicht reicht...



Digitalisierung & KI

STANDPUNKTE

KI-Systeme halluzinieren nicht



Aljoscha Burchardt, Research Fellow am DFKI und Xenia Kersting, Oberärztin in der Universitätsmedizin Mainz FOTO: ALJOSCHA BURCHARDT (PRIVAT) | XENIA KERSTING (HEADSHOTS.DE)

Künstliche Intelligenz (KI) gibt manchmal falsche Antworten. Sie erfindet Fakten, Bücher und Webseiten. Viele sagen dann, dass die KI „halluziniert“. Aljoscha Burchardt und Xenia Kersting halten das für unspezifisch, irreführend und stigmatisierend und schlagen eine Alternative vor.

von Aljoscha Burchardt und Xenia Kersting

veröffentlicht am 15.05.2024
aktualisiert am 16.05.2024

Die Position: Bringt der KI alles bei – jetzt!

Statt uns vor künstlicher Intelligenz zu fürchten, sollten wir sie trainieren. Noch gibt es die Fachkräfte dafür VON ALJOSCHA BURCHARDT

In Deutschland werden zwei Diskurse parallel geführt. Beim ersten geschieht das verhältnismäßig leise, er dreht sich um den Fach- und Arbeitskräftemangel. Dessen Ausmaß wird gemeinhin mit rund zwei Millionen unbesetzten Stellen beziffert. Bei uns kommt die Post nicht mehr jeden Tag, da keine Zusteller da sind, die S-Bahn dünnt mangels Personal ihren Fahrplan aus, und bei meiner Ärztin kann ich telefonisch keinen Termin machen, da sie ihren Empfang nicht mehr besetzt bekommt.

Schaut man auf die demografische Entwicklung, so wirkt das alles nur wie ein leiser Vorbote. Die Verrentungs-Welle oder, besser gesagt, der Verrentungs-Tsunami steht noch bevor. Der Bundesregierung fällt in ihrer Fachkräftestrategie zweierlei ein: Qualifizierung und Zuzug. Das sind an sich keine schlechten Maßnahmen. Doch fehlen heute schon massiv Leute, darunter mindestens 12.000 Lehrkräfte, Tendenz stark steigend – wir müssen uns also etwas mehr einfallen lassen.

Der zweite Diskurs, eher überhitzt geführt, schwappt seit Ende 2022 durch die Medien und stellt die Frage, ob uns künstliche Intelligenz (KI) nun wirklich den Arbeitsplatz kosten wird. In der Vergangenheit haben ähnliche Befürchtungen – Roboter, Personal Computer, Internet und so weiter – tendenziell immer dazu geführt, dass in den technologisierten Ländern die Zahl der Arbeitsplätze gewachsen ist. Aber diesmal könnte es eben anders sein. Interessanterweise ist Digitalisierung in der Fachkräftestrategie der Bundesregierung nur eine Randnotiz, KI kam auch in der entsprechenden Bundestagsdebatte Anfang 2023 so gut wie nicht vor. Aber spräche denn etwas gegen eine autonom fahrende S-Bahn oder eine rund um die Uhr erreichbare, KI-basierte Terminvergabe bei meiner Ärztin?

Wenn wir die beiden Diskurse zusammenbringen, dann eröffnet sich eine Möglichkeit, die unsere letzte Rettung sein könnte, die aber

wegen des zunehmenden Fachkräftemangels bereits ungenutzt zu verstreichen droht: Denn KI muss für ihre Aufgaben erst trainiert werden – in diesem Punkt ist sie noch immer ziemlich menschlich.

Dafür kommt heute meistens das maschinelle Lernen zum Einsatz, bei dem die KI-Systeme mit Daten trainiert werden und danach mithilfe von Mustererkennung selbst die Lösung der gestellten Aufgabe finden. Ein Beispiel dafür sind die bekannten Übersetzungssysteme wie DeepL oder Google Translate, die mit zwei- oder mehrsprachigen Texten gefüttert werden, welche in der Vergangenheit bereits von Menschen übersetzt worden sind. Andere Daten müssen teilweise aufwendig erzeugt und annotiert werden. Auch in ChatGPT ist erhebliche Manpower gesteckt worden. Daneben kommen in vielen Bereichen der Industrie oder der Bildung regelbasierte KI-Systeme zum Einsatz. Das sind etwa intelligente Tutoring-Systeme zur



Foto: privat

Aljoscha Burchardt
arbeitet am Deutschen
Forschungszentrum
für Künstliche
Intelligenz (DFKI)

Qualifikation von Fachkräften, die nach einem von Experten aufgesetzten Regelwerk individuelle Übungen, Lernvideos und Texte vorschlagen können.

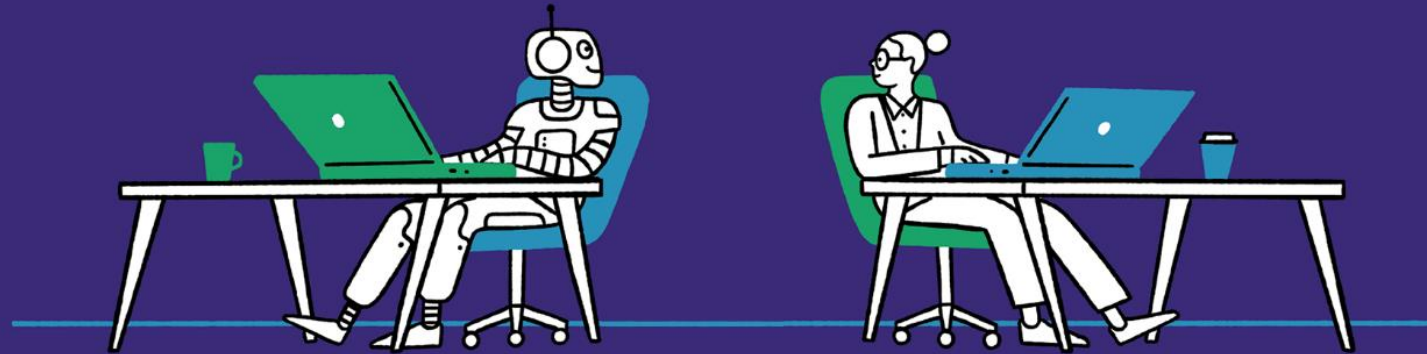
Das bedeutet: Wenn KI uns helfen soll, den Laden zukünftig mit sehr viel weniger Arbeitskräften am Laufen zu halten, dann müssen wir jetzt, wo wir noch nicht zu sehr unter Besetzungslücken leiden, die entsprechenden Prozesse gestalten, also die Nachhilfe für die KI organisieren. Ist nämlich der Wissensschatz über die Abläufe so vieler Tätigkeiten in Form von Menschen erst einmal in Rente gegangen, wird es sehr schwer werden, diese verpasste Chance nachzuholen. Dann sind wir fast zwingend darauf angewiesen, stattdessen die Technologie aus den USA oder China einzukaufen.

Ob diese dann zu unseren Werten und Arbeitsvorstellungen passt, wird sich zeigen. Ich selbst würde es lieber nicht darauf ankommen lassen.



AI Campus

The Learning Platform
for Artificial Intelligence



KI-Campus – Die Lernplattform für Künstliche Intelligenz

www.ki-campus.org

KI UND JETZT?



rbb

Food for thought

Algorithms will repeatedly make mistakes due to insufficient data, faulty programming, muddled goal definitions and the chaotic nature of life.

[...] most people often make terrible mistakes in the most important decisions of their lives.

Even more than algorithms, humans suffer from insufficient data, from faulty programming (genetic and cultural), from muddled definitions, and from the chaos of life.

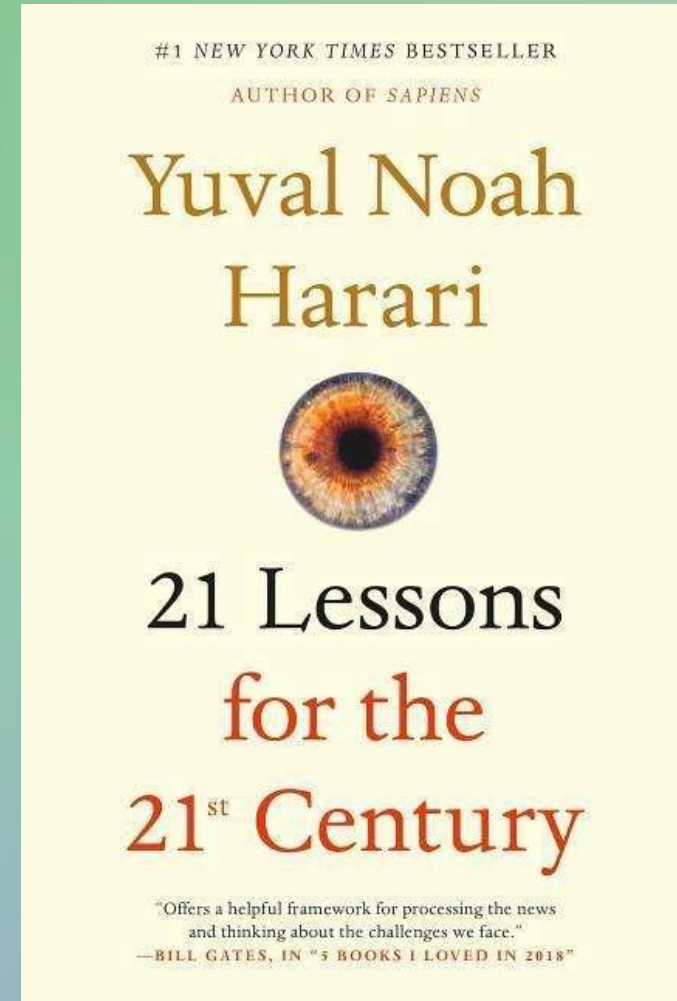


Image: www.jpc.de