

Mensch oder Algorithmus?

Zum verantwortungsvollen Umgang mit Künstlicher Intelligenz

Christiane Floyd

13. 10. 2023

Siehe: C Floyd: From Joseph Weizenbaum to ChatGPT: Critical encounters with dazzling AI technology,
in: *Weizenbaum Journal of the Digital Society*, vol. 3, issue 3, 2023

Was ist künstliche Intelligenz?

Künstliche Intelligenz ist
die **Fähigkeit einer Maschine,**
menschliche Fähigkeiten wie
logisches Denken, Lernen, Planen und Kreativität
zu imitieren.

– *Europäisches Parlament*

Künstliche Intelligenz - Aufbruch und Kritik

- ◆ 1956 - John McCarthy prägt den Namen *Artificial Intelligence*
- ◆ 1966 - Joseph Weizenbaum: erster Chatbot ELIZA
- ◆ 1976 - Joseph Weizenbaum: *Computer Power and Human Reason*
- ◆ ab 1985 - Künstliche neuronale Netze setzen sich durch

1984/85 Stellungnahme Floyd

Grenzen des verantwortbaren Computereinsatzes

- ◆ Grenzen des **fachlich verantwortbaren** Computereinsatzes, wo Computer aufgrund eines **verfehlten Vertrauens** in die Leistungsfähigkeit von Programmen eingesetzt werden.
- ◆ Grenzen eines **zwischenmenschlich verantwortbaren** Computereinsatzes, wo Computer aufgrund einer **verfehlten Gleichsetzung** von Menschen mit Maschinen eingesetzt werden.
- ◆ **Moralisch/rechtliche Grenzen** des Computereinsatzes, wo mit Computern versucht wird, **was ohne Computer nicht gemacht werden darf.**

Grundproblem Mensch vs. Maschine

- ◆ **Gleichsetzung oder Differenz?**
 - ◆ Der Mensch ist keine Maschine.
 - ◆ Aber „Maschine“ kann eine Metapher für Menschen sein.
 - ◆ Wir betrachten Einzelne als Maschine, Organisationen als Maschine ...
- ◆ **Ersetzung oder Unterstützung?**
 - ◆ Arbeitet eine Maschine einem Menschen zu?
 - ◆ Tritt eine Maschine *autonom* an die Stelle eines Menschen?

KI aus menschenzentrierter Sicht

- ◆ Technik *für* Menschen
- ◆ Kriterien
 - ◆ Förderung menschlicher Kompetenz
 - ◆ Erhalt menschlicher Zuwendung
 - ◆ Stärkung persönlicher und sozialer Verantwortlichkeit
- ◆ Szenarien für Technikeinsatz im Kontext

2023 - Sensation ChatGPT: im Fokus dieser Untersuchung

- ◆ universell anwendbarer Chatbot
- ◆ einheitliche sprachliche Oberfläche
- ◆ große Vielfalt an Leistungen
- ◆ basiert auf neuronalen Netzen und maschinellem Lernen
- ◆ trainiert mit riesigem Text Corpus

[In artificial intelligence...]
machines are made to **behave in wondrous ways**,
often sufficient to **dazzle**
even the most experienced observer.
But **once a particular program is unmasked**,
once its inner workings are explained[...]
its magic crumbles away.

– *Joseph Weizenbaum*

<https://cse.buffalo.edu/~rapaport/572/S02/weizenbaum.eliza.1966.pdf>

Entzaubern - Aneignen - Einschätzen

Experiment - Reflexionszyklen

A "wheeled" triskelion of three three-fold spirals (variant of triple-spira... <https://upload.wikimedia.org/wikipedia/commons/6/65/Triple-Triple-Spi...>

- ◆ Fragen formulieren
- ◆ Ausprobieren
- ◆ Reflektieren



Chat 1:

Wer war Paul Watzlawick und was sind seine wichtigsten Beiträge?

Paul Watzlawick (1921-2007) war ein österreichisch-amerikanischer Kommunikationswissenschaftler, Psychotherapeut und Autor ...

Hier sind seine wichtigsten Beiträge und Ideen:

1. Axiome der Kommunikation ...
2. Radikaler Konstruktivismus ...
3. Das Doppelbindungskonzept ...
4. Pragmatische Therapie ...
5. Bücher und Schriften ...



Paul Watzlawick hatte einen bedeutenden Einfluss auf die Bereiche Kommunikationswissenschaft, Psychotherapie und Psychologie ...

Einschätzung zu Chat 1

- ◆ Paul Watzlawick war sehr berühmt, hier und in den USA.
- ◆ Es gibt viele englische und deutsche Texte zu Paul Watzlawick.
- ◆ Das System kann
 - ◆ verlässliche Muster bilden
 - ◆ eine überzeugende Struktur erarbeiten.

Chat 2:

Längste Nebenflüsse der Donau

Die Donau, der zweitlängste Fluss in Europa, hat viele Nebenflüsse, von denen einige ziemlich lang sind ...

1. Der **Inn** ... ist der ~~längste~~ Nebenfluss der Donau ...
2. Die **Save** ... ist ein wichtiger Nebenfluss ...
3. Die **Theiß** ... ist ein bedeutender Nebenfluss ...
4. Die **Drau** ... fließt durch Italien, Österreich, ...
5. Die **Morava** ... ist ein Flusssystem in Südosteuropa ...

Die rumänischen Nebenflüsse fehlen ganz!

Save, Theiß, Drau: falsch gereiht!

Die Morava ist hier fehl am Platz!

Korrekt:

1. Theiß: 966 km
 2. Pruth: 953 km
 3. Save: 940 km
 4. Drau: 749 km
 5. Sereth: 647 km
 6. Olt: 615 km
 7. Inn: 517 km
- ...
- Morava: 185 km!

Einschätzung zu Chat 2

- ◆ Chat GPT kann (noch) nicht mit Zahlen und Messgrößen umgehen.
- ◆ Das System orientiert sich ausschließlich an textuellen Mustern und kennt hauptsächlich englische Texte.
- ◆ Die Aussagen von ChatGPT sind durch den Text-Corpus einseitig anglo-amerikanisch geprägt.
 - ◆ „Der Inn ist der längste Nebenfluss der Donau“ - ja, in Bayern!
 - ◆ Dagegen kommen rumänische Flüsse im Text-Corpus kaum vor.

Chat 3:

Leben und Wirken von Christiane Floyd

Christiane Floyd ist eine deutsche Informatikerin, die für ihre Beiträge zur Softwaretechnik und Mensch-Computer-Interaktion bekannt ist. Sie wurde am 20. Mai 1943 in Lübeck, Deutschland, geboren.

Floyd begann ihre akademische Laufbahn mit dem Studium der Mathematik und Physik an der Universität Münster. Sie promovierte 1971 in Informatik an der Universität Karlsruhe. Während ihrer Karriere arbeitete sie an verschiedenen renommierten Institutionen und Universitäten, darunter die Technische Universität Berlin und die Universität Stuttgart ...

Neu!

Bisher:

Ludwigshafen

Leipzig

Traunstein

Hamburg

Berlin

Einschätzung zu Chat 3

- ◆ Es gibt nur verhältnismäßig wenige Texte zum Thema CFL.
- ◆ Die textuellen Muster sind daher nicht verlässlich.
- ◆ Texte über CFL haben häufig Bezug zur TU Berlin oder zur Uni Hamburg. Also hält das System CFL für eine Deutsche.
- ◆ Das System nutzt eine Vorlage, in der Geburtstag, Geburtsort, Studium usw. vorgesehen sind - findet zu CFL aber nichts in Deutschland.
- ◆ Warum und wie ChatGPT Informationen erfindet, ist mir unbekannt.

Ist ChatGPT lernfähig?

Aus dem Chat "Arktisches Klima im Juli"

Christiane Floyd

Was ist das Klima in Nordpolen im Juli?

Ich habe nach dem Klima in Nordpolen gefragt.

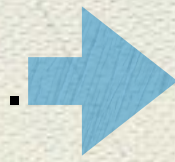
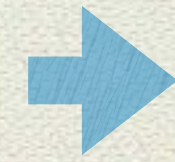
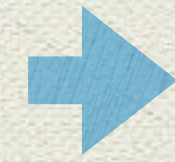
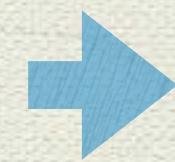
Mit welchem zusammengesetzten Wort lässt sich der nördliche Teil von Polen beschreiben?

Ok. Was ist das Klima von Nordpolen im Juli?

Nordpolen gibt es sehr wohl als geografischen Ort.

Christiane Floyd fährt diesen Sommer nach Nordpolen.

Dass Nordpolen eine geografische Region ist, habe ich schon zu Beginn dieses Chats klargestellt.



ChatGPT

Da der Nordpol auf dem Wasser des Arktischen Ozeans liegt, gibt es kein direktes Klima am Nordpol ...

Es gibt keinen spezifischen Ort namens "Nordpolen".

Der nördliche Teil von Polen wird oft als "Nordpolen" bezeichnet.

Es tut mir leid, ... Nordpolen gibt es nicht als geografischen Ort.

Sie haben Recht, dass "Nordpolen" ... verwendet wird, um den nördlichen Teil Polens zu beschreiben ...

Das ist unwahrscheinlich, da Nordpolen kein geografischer Ort ist.

Ich möchte mich für meine vorherigen Antworten entschuldigen ... Nordpolen bezieht sich in der Tat auf den nördlichen Teil Polens ...

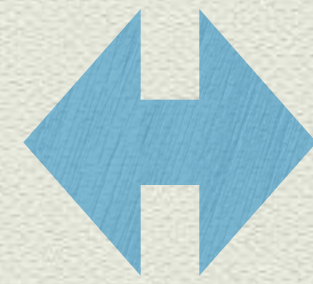
Anspruch und Wirklichkeit von ChatGPT

- ◆ ChatGPT
 - ◆ gibt auf manche Fragen ausgezeichnete Antworten,
 - ◆ führt in anderen Fällen zu unsinniger Fehlinformation,
 - ◆ zeigt nicht verlässlich seine eigenen Grenzen auf.
- ◆ Die Antworten von ChatGPT müssen überprüft werden!
- ◆ ChatGPT ist nicht vertrauenswürdig!

Wahrscheinlichkeit statt Wahrheit

- ◆ ChatGPT hat keinen Wirklichkeitsbezug.
 - ◆ Es kann nur Textmuster erkennen und transformieren.
 - ◆ Es bezieht sich auf *inferentielle*, nicht auf *referentielle* Semantik.
- ◆ ChatGPT versteht ohne zu verstehen.
- ◆ Die Wahrscheinlichkeit, dass eine Aussage zutrifft, soll möglichst hoch sein.
- ◆ Wie nahe kann diese Wirkungsweise dem menschlichen Denken kommen?

Ablehnen

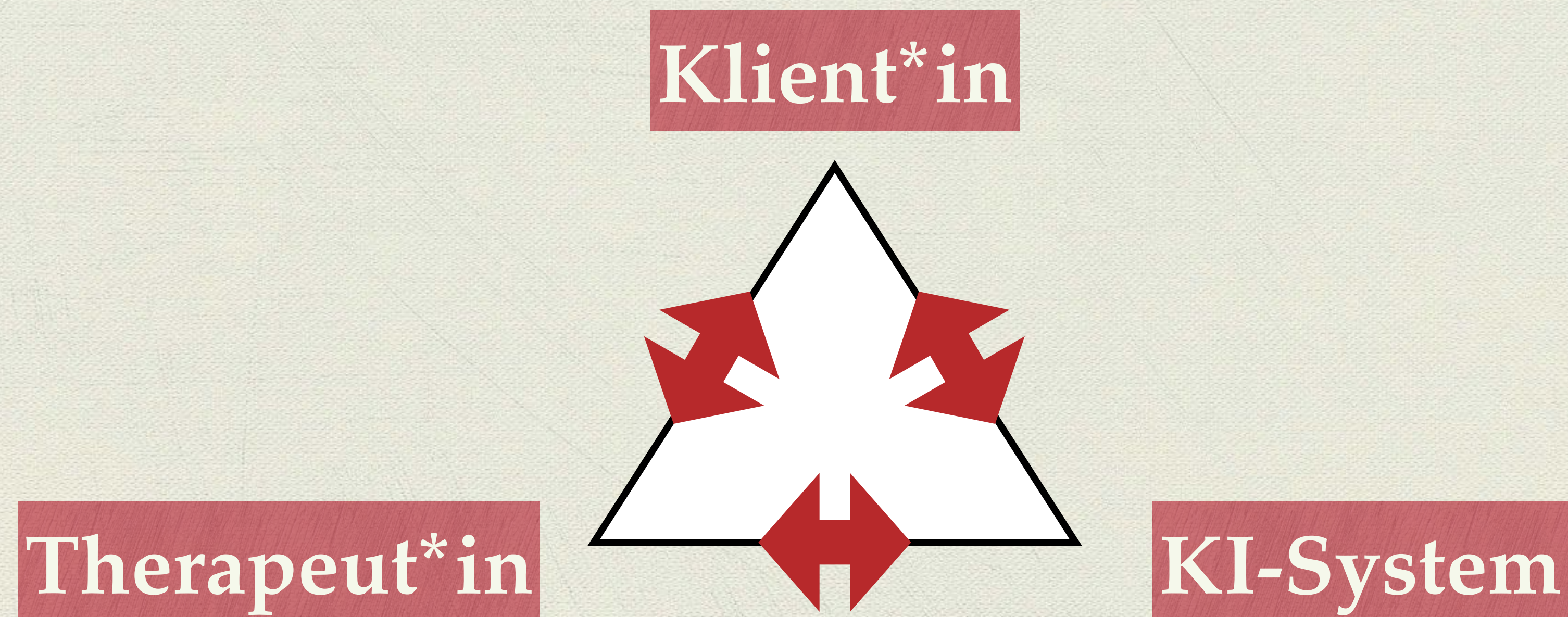


Gestalten

Gestaltung: Das Mögliche ermöglichen

- ◆ **Einsatzszenarien im Kontext**
 - ◆ Ziele, Beteiligte, Metaphern, Narrativ
- ◆ **Systemgestaltung**
 - ◆ Interaktivität, Aufgabenangemessenheit, Erklärbarkeit, Rechenschaftspflicht ...
- ◆ **Einsatzgestaltung**
 - ◆ Wer, was, wann, wie, wo?

Beispiel Diagnose / Therapie



Drei sehr unterschiedliche Beziehungen
Wie können sie zusammenspielen?

Szenario 1: Antrag auf einen Therapieplatz

- ◆ Klient*in muss ein KI-System benutzen, um einen Therapieplatz zu erlangen.
- ◆ Ziele: aussagekräftige Anträge, Entlastung des Krankenkassen-Personals
- ◆ Metapher: Intelligentes Formular
- ◆ Anforderungen:
 - ◆ Wahrung der Vertraulichkeit
 - ◆ Eigenständige Benutzbarkeit
 - ◆ Transparenter Fragenkatalog
 - ◆ Nachprüfbare Zuordnung zu standardisierten Kategorien ...

Szenario 2: Anamnese-Unterstützung

- ◆ Klient*in benutzt KI-System VOR Beginn der Therapie.
- ◆ Ziele: Einheitliche Standards und Transparenz bei der Anamnese, Entlastung der Therapie-Gemeinschaft, ggf. Diagnose-Vorschlag
- ◆ Metapher: Intelligenter Assistent
- ◆ Anforderungen:
 - ◆ ähnlich wie bei Antragstellung, aber
 - ◆ tieferes Verständnis der medizinisch-therapeutischen Domäne
 - ◆ zuverlässige Abstimmung mit der Therapie-Gemeinschaft

Szenario 3:

Vertiefende Therapie-Begleitung

- ◆ Klient*in benutzt ein KI-System zwischen Therapie-Sitzungen.
- ◆ Ziele: besseres Verständnis der Therapie, Unterstützung der Kommunikation, Stärkung der Zusammenarbeit zwischen Therapeut*in und Klient*in
- ◆ Metapher: immer verfügbarer Kontakt zur Therapie
- ◆ Anforderungen:
 - ◆ Therapeut*in muss das System sehr genau kennen,
 - ◆ das System muss sehr gut an die spezifische Therapie anpassbar sein
 - ◆ sehr anspruchsvolles Benutzungsmodell

Szenario 4: Unterstützung der Diagnose

- ◆ Therapeut*in benutzt KI-System zur Absicherung einer Diagnose.
- ◆ Ziele: Zuverlässige Diagnose, Verbesserung der Kommunikation in der Therapie-Gemeinschaft
- ◆ Metapher: Intelligentes Handbuch
- ◆ Anforderungen:
 - ◆ KI-System muss fundiertes Fachwissen verkörpern
 - ◆ Therapeut*in muss interaktiv eine Diagnose entwickeln können
 - ◆ Diagnosevorschläge des Systems müssen hinterfragt werden können
 - ◆ Verantwortlichkeit für die Diagnose muss gestärkt werden

Szenario 5: KI-System anstelle von Therapeut*in

- ◆ Klient*in nutzt ein KI-System zur Therapie ohne therapeutische Begleitung.
- ◆ Ziele: Therapie-Ersatz, wo eine menschliche Betreuung nicht möglich ist
- ◆ Metapher: Automatisierte Therapie — sehr fragwürdig
- ◆ Anforderungen:
 - ◆ müssen, wenn überhaupt, von erfahrenen Therapeut*innen entwickelt werden,
 - ◆ Oberstes Anliegen: Schaden an Klient*innen vermeiden,
 - ◆ wenn irgend möglich: Hotline zur Notfallbetreuung vorsehen ...

Herausbildung von Lernkulturen

- ◆ Jede Organisation oder Gemeinschaft zentriert sich um gemeinsames und geteiltes Wissen.
- ◆ Das gemeinschaftliche Wissen entwickelt sich kontinuierlich weiter.
- ◆ Gemeinschaften können das gemeinsame Lernen fördern.
- ◆ KI-Systeme ermöglichen neue Ebenen der Wissensbildung.
- ◆ Gemeinschaften können ihre Lernkultur mit KI-Systemen bereichern.

Zusammenführung

- ◆ KI-Systeme sind kein Zauberwerk.
 - ◆ Es lohnt sich, eigene Erfahrung mit ihnen zu machen.
- ◆ KI-Systeme sind noch nicht technisch ausgereift.
 - ◆ Sie können verbessert oder durch bessere abgelöst werden.
- ◆ KI-Systeme können gestaltet werden.
 - ◆ WIR sind aufgerufen, Anforderungen zu entwickeln.

DANKÉ

für Ihre Aufmerksamkeit!