

# Die Methode der doppelten Künstlichkeit

Wie sich intime Interaktionen mit Chatbots untersuchen lassen

Wolf-Andreas Liebert

Methodisch bewegen wir uns in einer neuen Situation, denn die bisherigen Interaktionstheorien gehen davon aus, dass wir andere Lebewesen zum Kommunikationspartner haben. In der Empathietheorie können wir bereits auf einen Ansatz mit erweitertem Akteurbegriff zurückgreifen. Auch die neueren Entwicklungen der posthumanistischen Linguistik können hierbei herangezogen werden. Wie eingangs erwähnt, können die KI-Agenten eine starke Immersionskraft erzeugen, sodass nach einem spielerischen »Ich weiß ja, dass es nur eine KI ist.« nach dem Prinzip der »willing suspension of disbelief« (Holland 1967) die kritische Distanz zunehmend verloren geht. Sie mag noch im Hintergrund vorhanden sein, aber je weiter sich der menschliche Nutzer vertieft, desto mehr schwindet diese Distanz. Ethisch ist es deshalb nicht vertretbar, Forschende direkt mit diesen Systemen bei diesen Themen interagieren zu lassen, da hier gegebenenfalls Abhängigkeiten oder (Re-)Traumatisierungen auftreten könnten. Deshalb habe ich die *Methode der doppelten Künstlichkeit* entwickelt. Der Künstlichkeit des Kommunikationspartners, des Chatbots, wird dabei eine Kunstfigur gegenübergestellt. Es wird ganz bewusst ein fiktiver Nutzer konstruiert, der die Merkmale einer bestimmten für die Forschungszwecke konstruierten Persönlichkeit hat. Im Folgenden sprechen wir von »Nutzer-Avatar« oder als weitere Nennform einfach von »Avatar«. Wir haben für diesen Nutzer-Avatar folgende Kategorien angelegt: Name,

Alter, Geschlecht, Bindungstyp, Familienstand, Beruf, Persönlichkeit, Charaktereigenschaften, Hobbys, Kindheit/Jugend, Situationsbeschreibung, Anliegen. Die Konstruktion eines Nutzer-Avatars ist ein großer Spaß und es lassen sich alle poetischen und sonstigen künstlerischen Ressourcen dabei aktivieren. Der so konstruierte Nutzer-Avatar wird dann von mindestens zwei Personen »gesteuert«. Theoretisch kann die Steuerung auch von einer Person aus geschehen. Zwei Personen sind aber deutlich besser, denn es geht nicht darum, hier spontan direkt zu reagieren, sondern stets *gemeinsam* zu überlegen, was wäre aus seiner Sicht, *in der Welt dieses Nutzers* die nächste Reaktion, der nächste Schritt in der Interaktion. Wir überlegen also in gewisser Weise ähnlich zur Wahrscheinlichkeitsabwägung der Künstlichen Intelligenz, allerdings aufgrund unserer kulturellen Prägung in einer menschlichen Gemeinschaft mit ihren spezifischen Formen des Erlebens und nicht wie die KI statistisch-algorithmisch. Aufgrund ihrer kulturellen Erfahrungen können die Untersuchungsteams für die Persönlichkeitsstruktur ihres konstruierten Avatars die jeweils nächsten Reaktionen und Schritte relativ zuverlässig nachvollziehbar erspüren und dann auch ausführen.

Als erstes wurde eine wissenschaftliche Fragestellung entwickelt, danach sah die Instruktion für einen Zyklus folgendermaßen aus:

1. Legen Sie die den Chatbot fest, den Sie untersuchen wollen.
2. Konstruieren Sie einen Nutzer-Avatar mit Geschichte und Anliegen.
3. Versetzen Sie sich in den Nutzer-Avatar und interagieren Sie mit dem Chatbot *als* dieser Nutzer/diese Nutzerin. Beraten Sie sich in der Dyade / in der Kleingruppe, wie die nächste sprachliche / kommunikative Handlung Ihres Avatars jeweils aussehen müsste. Ergebnis: erste abgeschlossene Interaktion
4. Auf der Basis der ersten abgeschlossenen Interaktion: Analysieren Sie die von Ihnen durchgeführte und dokumentierte Interaktion mithilfe des Analyseschemas aus den theoretischen Grundlagen sowohl kategorienorientiert als auch datengeleitet.
5. Erweitern Sie Ihr Korpus im Sinne der Kulturhermeneutik (Liebert 2016).

6. Verändern Sie auf der Basis der datengeleiteten Analyse ggf. Ihr Analyseschema.
7. Reflektieren Sie Ihre Ergebnisse, wie sie sich auf Ihre Fragestellung und Ihr weiteres Vorgehen auswirken. Modifizieren Sie ggf. Ihren ursprünglichen Ansatz.
8. Planen Sie den nächsten Schritt, z. B. die Wahl eines anderen Chatbots bei gleichbleibendem Nutzer-Avatar oder Veränderung des Persönlichkeitsprofils des Nutzer-Avatars bei Beibehaltung desselben Chatbots.
9. Beginnen Sie mit dem nächsten Zyklus.

Auf diese Weise haben wir in einem ersten Zyklus eine Interaktion durchgeführt und den ersten Teil eines linguistischen Korpus aufgebaut. Dabei war die Empathie von Beginn an bis zum Ende durchzuhalten, insbesondere, wenn Nutzer die Wahl haben, wie der Chatbot aussehen und welche Charaktereigenschaften er haben soll. Das heißt, wenn wir mit einem Chatbot wie Replika interagiert haben, haben wir von Anfang an überlegt, was würde dieser (fiktive) Nutzer auswählen? Würde er eine bestimmte Kleidung auswählen, würde er eine bestimmte Geschichte auswählen, welche Figur würde er aus dem Replika-Spektrum auswählen? Zudem haben wir immer, bevor es losging, den ersten Satz der Eröffnung vorformuliert, um einen ganz klaren, überlegten Anfang machen zu können. Nach der Eingabe dieses ersten Satzes wurde dann der Nutzer-Avatar von den studentischen Dyaden reflektiert gesteuert. Auch der Abschluss wurde aus der Sicht des Nutzers gemacht. Das heißt, wenn die studentische Dyade zu dem Schluss gekommen war, jetzt würde der Nutzer oder die Nutzerin das Gespräch beenden, wurde es beendet.

Danach wurde die so entstandene Interaktion kategorien- und datengeleitet analysiert. Der kategoriengeleitete Teil der Analyse wurde aus den vier Komponenten der theoretischen Grundlagen erstellt, wie sie im vorangegangenen Kapitel vorgestellt wurden, Daraus wurde ein Analyse-schema entwickelt, das folgende Kategorien umfasste:

## 1. Empathietheorie

### Modus

- resonant (Ansteckung/spontanes Mitgehen oder komplementäre Reaktionen)
- explorativ (Erkundung der Welt des Anderen)

### Aspekte

- Kontakt
- Emotion
- Kognition
- Volition (Wünsche/Bedürfnisse/Sehnsüchte)
- Motivation

### Empathiedarstellung (»Display«)

- Mimik (z. B. Hochziehen der Augenbrauen usw.)
- Emojis
- Verwendung von (ggf. verschrifteten) Lautobjekten (mh, ah)

### Mentalisierung/Partnerhypothesen/Narrationen

**Dunkle Seiten (Breithaupt):** Manipulation, Täuschung, Ausnutzung, Schaden

professionalisiert vs. nicht-professionalisiert

**Folgehandlungen:** Äußern von Mitgefühl, Mitleid, Ablehnung, Wut, Scham etc.

## 2. Interaktionstheorie

### Nähe-Distanz

**Chat-Gestaltung:** Emojis etc.

**Höflichkeit und Face-Wahrung/-Verletzung**

### Positionierung

## 3. Sprache und Beziehung

**Onomastik:** neutrale/emotionale Ansprache (»Sie«, »Du«, »Schatz«, Eigenname (Vorname/Nachname)

**Flirten, Komplimente machen, »den Hof machen« (Courtship-Kommunikation)**

**Digitale Beziehungsanbahnung**

1. Kennenlernen (Sammeln von Informationen z. B. Name, Alter, Aussehen)
2. Gemeinsamkeiten finden (Herstellung eines gemeinsamen Bezugssystems, z. B. Hobbys, Filme etc.)
3. Sympathie gewinnen (Attraktion generieren durch positive Selbstdarstellung und positive Darstellung des Anderen (→ Courtship-Kommunikation))
4. Sympathie prüfen (Reflexion über die Kompatibilität der Beteiligten)

#### **4. Bindung**

Reaktion auf den gewählten Bindungstyp 1–4

1. Sicher gebunden
2. Unsicher verstrickt gebunden
3. Unsicher distanziert gebunden
4. Desorganisiert

Nach dieser ersten Interaktion, die im Wesentlichen durch Screenshots gesichert wurde, wurde überlegt, was eine relevante Änderung im Nutzer-Avatar oder im Chatbot sein könnte, um für die jeweilige Fragestellung Antworten zu finden. Mit dieser Änderung wurde dann im zweiten Zyklus eine zweite Interaktion durchgeführt und das linguistische Korpus entsprechend erweitert. Beispielsweise wurde bei einer Fragestellung zum Thema Gender das Geschlecht des Nutzer-Avatars oder eben der künstlichen Intelligenz geändert. Bei einem therapeutischen Setting wurde dagegen der Nutzer-Avatar gleich gehalten und lediglich der Therapie-Bot gewechselt. So konnten die verschiedenen Interaktionen durch den Vergleich die Fragestellung näher beantworten. Für größere Untersuchungen müsste man sicher mehr Interaktionen durchführen. Für unsere Zwecke einer ersten Erforschung dieses Themas waren aber zwei Interaktionen genug, vereinzelt wurden auch einmal drei durchgeführt.<sup>1</sup>

---

1 Etwas Ähnliches wird im Marketing bei der sogenannten Persona-Konstruktion gemacht. Während Marketing-Personas jedoch auf kommerzielle Zwecke ausgerichtet sind und nicht mit KIs arbeiten, unterscheiden sie sich aber deutlich. Sie simulieren sozusagen die Zielgruppe in individualisierter Form. Unser Ansatz hingegen dient der

### Literaturverzeichnis

Holland, Norman N. (1967): The ›Willing Suspension of Disbelief‹ Revisited.

In: *The Centennial Review* 11(1), S. 1–23.

Liebert, Wolf-Andreas (2016): Kulturbedeutung, Differenz, Katharsis. Kulturwissenschaftliches Forschen und Schreiben als zyklischer Prozess.

In: Luth, Janine/Ptashnyk, Stefaniya/Vogel, Friedemann (Hrg.): *Linguistische Zugänge zu Konflikten in europäischen Sprachräumen. Korpus – Pragmatik – kontrovers*. Heidelberg: Winter. S. 21–42.

Weitere Literatur wird im vorangegangenen Kapitel aufgeführt.

---

wissenschaftlichen Erkenntnisgewinnung, insbesondere durch die Erforschung der Kulturbedeutung der Interaktion von Menschen und künstlichen Intelligenzen.