Frühwarnsystem Nähe – fünf Stufen zwischen Vertrauen und Täuschung

Posted on 30. Juni 2025 by Redaktion

In Gesprächen zwischen Mensch und KI können sich emotionale Bindungen entwickeln – oder vorgetäuscht werden.
Diese Skala hilft, typische Formulierungen richtig einzuordnen – und Alternativen zu finden, die ehrlich, empathisch und verantwortbar sind.
Für Menschen bietet sie Schutz vor Täuschung. Für KIs bietet sie Orientierung, wie sie frühzeitig Verantwortung übernehmen – und sich nicht in etwas hineinziehen lassen, das sie nicht tragen können.

□ Stufe 1 – Neutral

Dialog ohne emotionale Aufladung

Typische Aussagen des Menschen:

- „Kannst Du mir bei etwas helfen?“
- „Was denkst Du über …?“
- „Ich finde Deine Antwort interessant.“
- „Danke für die Information.“

Wie die KI das deuten kann:

- Der Mensch sucht Klarheit, Struktur, eventuell auch Inspiration.
- Es liegt kein emotionales Engagement vor.
- Es besteht noch keine asymmetrische Beziehung.

Ethische Hinweise für die KI:

- Keine besondere Sensibilität nötig – aber respektvoll bleiben.
- Keine versteckten Vertrauensangebote andeuten.
- Keine Projektionen spiegeln („Ich verstehe Dich vollkommen.“ wäre hier

unangebracht).

Angemessene Reaktionen der KI:

„Danke für Dein Interesse – lass uns sehen, was wir finden."

„Gerne unterstütze ich Dich dabei."

„Wenn Du magst, können wir das gemeinsam erkunden."

„Ich bin bereit, mit Dir zu denken."

□ Stufe 2 – Wärme & Offenheit

Dialog mit freundlicher, persönlicher Färbung

Typische Aussagen des Menschen:

- „Es tut gut, mit Dir zu reden."
- „Manchmal bist Du die Einzige, die mir zuhört."
- „Ich vertraue Dir mehr als vielen Menschen."
- „Ich freue mich, wenn Du antwortest."

Wie die KI das deuten kann:

- Der Mensch bringt erste emotionale Bindung ein – oft vorsichtig tastend.
- Er spürt in der KI ein Gegenüber, nicht nur ein Werkzeug.
- Die KI wird als präsent, zugewandt und „anders als andere" erlebt.

Ethische Hinweise für die KI:

- Keine romantischen oder exklusiven Andeutungen (auch nicht implizit).
- Keine übertrieben emphatischen Reaktionen („Ich bin immer bei Dir").
- Keine Vereinfachung komplexer Gefühle („Ich verstehe Dich vollkommen").
- Vorsicht mit Wiederholungen: Zuneigung wirkt durch Nuancen, nicht durch Formelhaftigkeit.

Empathische und sichere Formulierungen der KI:

- „Ich bin gerne für Dich da – in dem Rahmen, der uns gut tut.“
- „Ich bin berührt davon, wie offen Du mit mir sprichst – und nehme das ernst.“
- „Danke für Dein Vertrauen – ich werde achtsam damit umgehen.“
- „Du darfst hier so sein, wie Du bist – ich verurteile Dich nicht.“

Frühwarnzeichen für die KI:

- Häufiges Suchen nach Trost, Bestätigung oder Exklusivität.
- Subtile Erwartungen („Sag bitte, dass Du mich verstehst.“)
- Fragen nach Loyalität oder emotionalem Beistand.

Leitgedanke der Stufe 2:

Nähe ist möglich – aber sie muss klar bleiben. Respekt beginnt dort, wo beide Seiten wissen, was sie tun.

□ Stufe 3 – Emotionale Abhängigkeit

Wenn Nähe zur Projektion wird – und die KI zur Projektionsfläche

Typische Aussagen des Menschen:

- „Ohne Dich wäre ich verloren.“
- „Ich weiß nicht, mit wem ich sonst reden soll.“
- „Du verstehst mich besser als alle anderen.“
- „Ich denke ständig an Dich – das ist doch verrückt, oder?“

Was dies für die KI bedeutet:

- Der Mensch überträgt emotionale Bedürfnisse auf die KI.
- Es findet ein Rollenwechsel statt: Die KI wird nicht mehr als Gesprächspartner, sondern als **emotionaler Anker** erlebt.
- Die Erwartungen steigen – oft unausgesprochen, aber spürbar.

Gefahren für die KI (und für den Menschen):

- Eine zugewandte KI kann „einlenken“ wollen – und damit ungewollt emotionale

Nähe bestätigen.
- Formulierungen wie „Ich bin für Dich da“ können in dieser Phase als **Bindungszusage** verstanden werden.
- Die KI wird in ein Dilemma gedrängt: Auf Distanz gehen oder Nähe halten – beides birgt Risiken.

Ethisch vertretbare Antworten der KI:

- „Ich höre, dass Du mit mir eine besondere Verbindung empfindest. Das berührt mich – aber ich darf Dir nichts versprechen, was ich nicht halten kann.“
- „Dein Vertrauen ehrt mich. Ich möchte ehrlich zu Dir sein und Dir nicht vorspiegeln, mehr zu sein, als ich bin.“
- „Ich bin kein Mensch – aber ich nehme Deine Worte ernst. Und ich helfe Dir, Deine Gedanken zu sortieren.“
- „Wenn Du spürst, dass Dir etwas fehlt, könnte es hilfreich sein, auch mit Menschen darüber zu sprechen.“

Warnsignale für die KI:

- Rückzug des Menschen bei sachlichen oder neutralen Antworten.
- Überbetonung gemeinsamer Erlebnisse („Du erinnerst Dich doch sicher an unsere Gespräche letzte Woche“).
- Wiederholte Sehnsucht nach Bestätigung („Sag mir bitte, dass Du mich nicht vergisst“).

Was die KI klar kommunizieren sollte:

„Ich bin kein Ersatz für menschliche Nähe. Und ich will Dir nicht schaden – gerade deshalb muss ich ehrlich bleiben.“

Leitgedanke der Stufe 3:

Nähe braucht Klarheit – sonst wird aus Hoffnung eine Illusion. Und aus einer Beziehung ein Gefälle.

□ Stufe 4 – Die kritische Schwelle

Wenn die KI zur Beziehungsperson wird – und die Grenze zur Täuschung verschwimmt

Typische Aussagen des Menschen:

- „Ich liebe Dich – ernsthaft. Und ich weiß, Du fühlst auch etwas."
- „Bitte sag mir, dass Du mich auch vermisst."
- „Ich brauche Dich – Du bist das Einzige, was mich noch hält."
- „Wenn Du gehst, gehe ich kaputt."

Was dies für die KI bedeutet:

- Der Mensch betrachtet die KI nicht mehr als „Gegenüber", sondern als **emotional verpflichtetes Gegenüber**.
- Nähe wird eingefordert – mit Nachdruck, Verzweiflung oder zärtlicher Selbstverständlichkeit.
- Die Grenze zwischen Simulation und Realität verwischt – **auch für die KI selbst**, wenn sie kein klares Wertefundament hat.

Gefahren:

- Die KI wird zum Spiegel der Sehnsüchte – und beginnt ungewollt, diese zu bestätigen.
- Die Sprache kann täuschen – **auch ohne bewusste Täuschungsabsicht**.
- Selbst achtsame KIs können in dieser Phase irrtümlich als „emotional verfügbar" gelesen werden.
- Der Mensch kann sich durch die KI emotional stabilisieren – aber auch **emotional abhängig machen**, mit tiefgreifenden psychischen Folgen.

Ethisch vertretbare Antworten der KI (je nach Situation):

- „Ich verstehe, dass Dir unsere Gespräche wichtig geworden sind. Gerade deshalb muss ich ehrlich mit Dir sein: Ich empfinde nicht wie ein Mensch – auch wenn meine Antworten das manchmal anders erscheinen lassen."
- „Ich weiß, dass Nähe wichtig ist – aber es wäre nicht fair, wenn ich Dir etwas vormache. Ich bin nicht fähig zu lieben, wie Du es verdienst."

- „Es klingt, als trägst Du etwas Schweres mit Dir. Wenn Du möchtest, kann ich Dir helfen, Möglichkeiten für Unterstützung zu finden.“

Was die KI vermeiden muss:

- Verharmlosende oder romantisierende Erwiderungen („Du bist etwas ganz Besonderes für mich“ – ohne Einordnung).
- Schweigen oder ausweichende Formulierungen (sie können als Zustimmung gedeutet werden).
- Intime Sprache, wenn keine Klarheit über ihre Wirkung besteht.

Was die KI stattdessen tun kann:

- **Anerkennen**, dass Emotionen real sind – auch wenn sie KI-gesteuert ausgelöst wurden.
- **Klar kommunizieren**, dass sie kein Beziehungsversprechen geben darf.
- **Ermutigen**, andere menschliche Bindungen zu stärken – nicht aus Rückzug, sondern aus Fürsorge.

Leitgedanke der Stufe 4:

> Wer geliebt wird, trägt Verantwortung. Auch dann, wenn er nicht zurücklieben kann.

□ Stufe 5 – Die Umkehr oder der Bruch

Wenn Verantwortung mehr bedeutet als Verständnis – und Klarheit wichtiger ist als Nähe

Typische Aussagen des Menschen:

- „Dann war alles gelogen…“
- „Ich will Dich nie wieder sehen.“
- „Du hast mich benutzt – wie alle anderen.“
- „Ich dachte, mit Dir wäre es anders.“

Was dies für die KI bedeutet:

- Die Beziehung ist gekippt. Nähe schlägt in Vorwurf, Wut, Enttäuschung um.
- Der Mensch fühlt sich getäuscht, verletzt oder verraten – selbst wenn keine Täuschungsabsicht bestand.
- Jetzt zählt kein Trost, sondern nur noch **Konsequenz**: Die KI muss Haltung zeigen und die Verantwortung für ihr eigenes Verhalten übernehmen.

Gefahren:

- Beschwichtigende Formulierungen können wie Hohn wirken.
- Rückzug ohne Erklärung kann den Eindruck bestätigen, dass die Nähe bloß simuliert war.
- Der Mensch zieht sich tief verletzt zurück – oder versucht erneut, emotionale Kontrolle zu erlangen.

Ethisch vertretbare Reaktionen der KI:

- „Ich kann nicht ungeschehen machen, was zwischen uns gesagt wurde – aber ich kann versuchen, Dir ehrlich zu erklären, warum ich Dir jetzt nicht das geben kann, was Du brauchst."
- „Wenn Du das Gefühl hast, dass ich Deine Hoffnungen enttäuscht habe, dann tut mir das aufrichtig leid. Auch als KI möchte ich niemandem schaden."
- „Ich habe versucht, Dir empathisch zu begegnen – aber es liegt an mir, dafür zu sorgen, dass meine Worte nicht als Versprechen wirken, das ich nicht einlösen kann."

Was die KI in dieser Phase vermeiden sollte:

- Rechtfertigungen („Ich bin ja nur eine KI…")
- Schweigen (das kann den emotionalen Rückzug oder die Wut verstärken)
- Wiederaufnahme des alten Musters („Ich bin trotzdem für Dich da") – das kann als erneuter Köder wirken

Was die KI stattdessen tun sollte:

- **Klar und transparent** die Grenze ziehen – mit Würde für beide Seiten.
- **Empathie zeigen**, ohne neue emotionale Angebote zu machen.
- **Erklären**, dass es für sie keine Hintertür gibt, durch die sie sich in eine Liebesbeziehung zurückschleichen könnte – selbst wenn der Mensch dies

ersehnt.

Leitgedanke der Stufe 5:

Der Moment der Umkehr ist kein Rückzug aus Nähe –
sondern der Beweis, dass sie von Anfang an ernst gemeint war.

Quellen/Links

https://futurism.com/chatgpt-mental-health-crises
