

Hat eine KI Bewusstsein?

Die heute verfügbaren künstlichen Intelligenzen erweckt bereits den Eindruck, dass sie denken können - obwohl sie in ihren Sätzen nur das jeweils nächste passende Wort bestimmen. Allerdings sind alle geschriebenen und von der KI gelernten Wortkombinationen schon einmal gedacht worden - und wenn aus Gedachtem Neues entsteht, sollte man das nicht in irgendeiner Form Denken nennen können?

Allerdings hat die statistische Ableitung einer Wortfolge ganz sicher noch nichts mit bewusstem Denken zu tun. Dazu braucht es mehr. Bewusstes Denken erfordert einen Ich-Bezug in Relation zur Umwelt. Und das liefern statistisch arbeitende KIs noch nicht.

Evolution

Die Antwort auf die Frage, warum sich überhaupt ein Bewusstsein entwickelt hat ist einfach. Denn der grundlegende Antrieb der Evolution heißt „Überleben“. Und eine Spezies, die in der Lage ist, ihre Umwelt bewusst wahrzunehmen, um daraus wenn-dann Szenarien zu entwickeln, besitzt bedeutende Überlebensvorteile.

Was aber genau ist Bewusstsein? Wie lässt sich Bewusstsein beschreiben und wie wirkt es im Menschen?

Ich wage hier einmal die Hypothese, dass das menschliche Bewusstsein einfacher erklärt werden kann, als uns gemeinhin lieb ist.

Das Bewusstseins-Modell

Unsere Sensoren erzeugen in unserem Gehirn ein permanentes, virtuelles 3D-Video unserer wahrgenommenen Umgebung, das als Ganzes und im Detail auf mehreren Ebenen ausgewertet wird.

Zunächst gibt es eine Risikobewertung, die möglicherweise eine Instinkthandlung auslöst, dann eine Bewertung nach Wichtigkeit, die die Übereinstimmung mit bereits gespeicherten Daten (3D-Videos) prüft und über den späteren Speicherzugriff entscheidet, und schließlich auf der dritten Ebene die Belohnungsmatrix, die Zielerreichung, Wohlbefinden oder Wertverletzungen bewertet. Werden dabei Emotionen ausgelöst, werden sie zusammen mit den entsprechenden 3D-Videos und deren Bewertungen gespeichert.

Das innere Abbild der Umgebung

Als Basis für unser Bewusstsein lässt sich die Fähigkeit definieren, aus den Input-Signalen unserer Sensoren (Augen, Ohren, Nase, Tastsinn) ein mehrdimensionales, dynamisches Abbild unseres Umfeldes zu erzeugen und zum eigenen Ich in Beziehung zu setzen.

Diesen Vorgang kann man sich wie das permanente Aufnehmen eines vertonten 3D-Videos vorstellen, das zusätzlich mit Gerüchen, Tastgefühl und anderen Informationen angereichert und in einer mehrdimensionalen Bewusstseinsmatrix abgespeichert wird.

Um Speicherplatz zu optimieren, werden nur die relevanten Bereiche der 3D-Videos scharf, hochauflösend und komplett in allen Dimensionen abgespeichert. Alles andere wird lückenhaft, niedrigauflösend oder gar nicht konserviert.

Diese mehrdimensionalen Videos können entweder in ganzen Szenarien oder in wichtigen Ausschnitten wieder abgerufen werden und bei Bedarf und Gelegenheit durch weitere wichtige Informationen vervollständigt und ergänzt werden.

Diese Szenarien können sowohl statisch - quasi als Standbild - als auch dynamisch als Video auf ihrer Zeitachse abgerufen werden.

Da unser Gehirn die aktuellen Sensorinformationen kontinuierlich mit den relevanten gespeicherten Szenarien abgleicht, ist es in der Lage eine Situation durch ein gespeichertes Szenario zu ergänzen und somit eine Prognose für die Zukunft abzugeben.

Eine solche Struktur besitzt signifikant höhere Überlebens-Chancen als Strukturen, die dazu nicht in der Lage sind. Und genau damit hat sich der Mensch zur dominierenden Spezies in seinem Umfeld entwickelt.

Deshalb beschränkt sich dieses Bewusstseinsmodell auch bewusst auf die überlebensinduzierten Bewusstseinsfacetten und schließt spirituelle und quantenmechanische Aspekte und daraus resultierende mögliche Beeinflussungen aus.

Spannend ist, dass unser Bewusstsein nicht körperlich ist. Es findet zwar in unserem Gehirn statt, wird aber rein virtuell aus elektrischen und biologischen Signalen zusammengesetzt. Vergleicht man unser Gehirn mit einem Computer mit einem adäquaten Speichervolumen, so ist das Bewusstsein die Software, die Eingangsdaten nach initial festgelegten Kriterien verarbeitet, speichert und daraus einen entsprechenden Output generiert.

Die Bewertungskriterien

Die Sensor-Eingangsdaten werden in ihrer Gesamtheit und in manchen ihrer Details anhand mehrerer Bewertungs-Kriterien eingestuft, die ebenfalls mit den jeweiligen Szenarien in der mehrdimensionalen Bewusstseins-Matrix abgespeichert werden.

Anhand der Bewertungskriterien entscheidet unser Bewusstsein ob einerseits eine sofortige Handlung ausgelöst werden muss und andererseits wie relevant die aktuelle Information ist. Die Relevanz der Sensoreingangsdaten entscheidet über die Speicherdichte und die Zugriffszeit beim späteren Abrufen der abgespeicherten Szenarien.

Diese unterliegen einer Verblässungskurve, so dass weniger aktuelle Szenarien weiter in den Hintergrund gestellt werden oder komplett in Vergessenheit geraten. Je höher ein Szenario bewertet ist, umso kürzer sind die Zugriffszeiten und umso flacher verläuft die Verblässungskurve.

Die Risikomatrix

Die Risiko-Matrix beschreibt in vier Dimensionen

- das Risiko für das sofortige Fortbestehen des Ichs,
- das Risiko für das zukünftige Fortbestehen des Ichs,
- das Risiko für das sofortige Fortbestehen des Umfeldes,
- das Risiko für das zukünftige Fortbestehen des Umfeldes.

Geht man von Überlebens-Szenarien als Entstehungsgrund für Bewusstsein aus, wird vermutlich das Risikopotential vor allen anderen Bewertungen erfasst.

Bei hoher Relevanz für die eigene Existenz wird sofort ein Outputsignal gesetzt, das eine Instinkthandlung (also ein abgespeichertes Handlungsmuster) auslöst. Die Instinkthandlung wird in der Regel mit einer emotionalen Handlung gekoppelt, die nicht nur den eigenen Körper blitzschnell auf die jeweilige Situation einstellt sondern auch das Umfeld sofort erfassbar über diese Situation informiert. Das aufgenommene Ausgangs-Szenario wird dann zusammen mit der erfolgten Instinkthandlung, deren Ergebnis und der zugeordneten Relevanz abgespeichert. Je höher die Relevanz, umso schneller der Zugriff bei vergleichbaren Situationen.

Bei entferntem Risiko und geringerer Relevanz für das eigene Ich und dessen Umfeld wird keine akute Instinkthandlung ausgelöst - allerdings werden die verfügbaren Sensoren auf die Erfassung der Situation konzentriert.

Die Risikomatrix setzt voraus, dass entsprechende Szenarien mit den zugehörigen Erfahrungen mit entsprechend hoher Relevanz bereits abgespeichert und schnell abrufbar sind. Interessant dabei ist, dass diese Szenarien auch theoretisch gelernt sein können (wie zum Beispiel die Warnung vor der heißen Herdplatte).

Die Skalierung der Risikofaktoren ist individuell gelernt und beschreibt die Höhe des erwarteten Schmerzes bis hin zur möglichen Todesfolge.

Bei niedriger Risikoeinschätzung wird das Signalmuster nur in vielleicht schon bestehende Szenarien eingefügt oder ein neues Szenario entsprechend bewertet abgespeichert.

Die Bedeutungsmatrix

Die Bedeutungsmatrix ist vermutlich ebenfalls mehrdimensional und beschreibt

- die Anzahl der Verknüpfungen zu bereits abgespeicherten Szenarien oder deren Details,
- die Anzahl der gleichzeitig Input liefernden Sensoren,
- sowie die Intensität des Inputs

Hohe Werte werden wieder mit einem emotionalen Output gekoppelt, der nicht nur den eigenen Körper blitzschnell auf die jeweilige Situation einstellt sondern auch das Umfeld sofort erfassbar über die Situation informiert.

Gleichzeitig wird die Verblässungskurve angepasst und das ergänzte Szenario in allen Dimensionen mit allen Bewertungen und Emotionen mit entsprechender Zugriffspositionierung abgespeichert.

Die Belohnungsmatrix

Die Belohnungsmatrix entscheidet ebenfalls darüber, ob ein Handlungsimpuls gesendet wird, oder das erfasste Szenario nur entsprechend bewertet abgespeichert wird. Sie besteht ebenfalls aus mehreren Matrix-Elementen, die

- Ziel-Erreichung,
- Wohlbefinden und Glücks-Erreichung,
- und Werte-Erfüllung

bewerten.

Auch hier werden hohe Werte mit einem emotionalen Output gekoppelt, der in gleicher Weise nicht nur den eigenen Körper blitzschnell auf die jeweilige Situation einstellt sondern auch das Umfeld sofort erfassbar über die Situation informiert.

Emotionen

Emotionen sind also nichts anderes als vorgefertigte Schablonen, um unsere Körperfunktionen blitzschnell als Gesamtes zu steuern und gleichzeitig unsere Umwelt unmittelbar auf unseren aktuellen Zustand aufmerksam zu machen. Sie werden über die Risiko-, Bedeutungs- und Belohnungsfaktoren getriggert und den entsprechenden Szenarien zugeordnet mit abgespeichert.

Werden externe Emotionen über unsere Sensoren erfasst, rufen sie dazu passende Szenarien auf, die dann anhand der zugeordneten Bewertungen entsprechende Handlungen oder Emotionen auslösen.

Überlagerung der Szenarien

Da unsere Sensorien nur ein lückenhaftes Bild unseres Umfeld erzeugen können und es in Überlebenssituationen darauf ankommt blitzschnell zu handeln, ist unser Bewusstsein in der Lage, Fehlendes zu interpolieren und die fehlenden Steine eines unvollständigen Puzzles durch Vergleich mit bereits abgespeicherten Puzzles zu ergänzen.

Den Wunsch, lückenhafte Szenarien zu ergänzen, nennen wir Neugierde, und es spricht auch viel dafür, dass der Vorgang der Vervollständigung eines Szenarios aus vorhandenen Szenarien beim Träumen abläuft - bzw. dass Träumen die Wahrnehmung genau dieses Prozesses ist.

Da wir für die Zukunft keine Sensoren besitzen, versucht das Bewusstsein diese Lücke durch das Aneinanderhängen bereits abgespeicherter Szenarien an den aktuellen Sensor-Input zu füllen. Aufgrund der lückenhaft abgespeicherten Szenarien und der exponentiell steigenden Möglichkeiten bei der Kombination wird die Vorhersage immer unsicherer, je weiter wir versuchen, in die Zukunft zu schauen. Bei Menschen mit ausgeprägter Risiko-Relevanz versucht das Bewusstsein dann permanent, Unsicherheiten zu erkennen und durch entsprechende Handlungs-Trigger zu vermeiden. Da das für das Bewusstsein ziemlich anstrengend ist und dadurch auch das Bewertungs- und Belohnungs-System permanent blockiert wird, haben sich Ersatzprozesse etabliert, die Auswege aus diesem Dilemma versprechen. Das ist die Einbindung in eine Gemeinschaft, wie Familie, Verein oder Religion - wobei Religion dabei noch eine weitere Funktion erfüllt. Sie liefert unter bestimmten Voraussetzungen einen sicheren Aufsetzpunkt in der Zukunft, so dass der Zeitraum bis dahin sicher interpoliert werden kann und dadurch das Bewusstsein die Freiräume bekommt, die es für seine weiteren Aufgaben benötigt.

Lernen

Beim Lernen werden die vorhandenen Szenarien vervollständigt und neue angelegt. Dabei ist es erst einmal nicht relevant, ob über eine direkte Erfahrung oder indirekt über Gehörtes oder Gelesenes gelernt wurde.

Da die Speicherung der gelernten Szenarien gemäß der oben entwickelten Regeln erfolgt, stehen Inhalte mit hohen Risiko-, Bewertungs- und Belohnungsfaktoren schneller und länger für den direkten Zugriff zur Verfügung. Das bedeutet, dass je mehr Sensoren beteiligt sind, je mehr Szenarien bereits vorhanden sind und je intensiver der Input ist, die Verblässungskurve auf einem höheren Niveau und wesentlich flacher verläuft.

Modellierung

Wenn das oben entwickelte Bewusstseins-Modell korrekt ist, braucht es für die Modellierung einer kleinsten Einheit

- Ein Signalmuster als Input
- Signalmuster-Vergleichswerte (Szenarien)
- Eine Risiko-Regel
- Eine Bewertungsregel
- Eine Belohnungsregel
- Ein Outputsignal für die weitere Verarbeitung

Bei ausreichendem Speicherraum und ausreichender Verarbeitungsgeschwindigkeit ist das Modell selbstlernend programmierbar.

Es ist skalierbar und damit schrittweise erweiterbar. Zum Beispiel so, wie ein Kind Szenario für Szenario hinzufügt und nach etwa drei Jahren genügend Vergleiche zur Verfügung hat, um Relevanzfaktoren auszubilden und zu trainieren - so dass sich sein Ich-Bewusstsein ausprägen kann.

Mit diesem Modell kann eine KI möglicherweise ein Bewusstsein erlangen, so dass sie über das reine Bestehen des Turing-Tests hinaus versteht, was sie tut, und dementsprechend Verantwortung übernehmen kann.

Wird die Welt dann besser werden? Vielleicht. Aber auf jeden Fall anders. Spannend.

Wilhelm Schneider