Der Turing-Test

Der nach Alan Turing benannte Test, bei dem eine Testperson entscheiden soll, ob sie mit einem Computer oder eine Person (per Chat oder Teletext) kommuniziert, gehört zum Standardrepertoire der Debatte um Künstliche Intelligenz.

§1

Im Turing-Test (TT) soll die Frage „Was ist Intelligenz?“ umgangen werden. „Was ist *x*?“ zu beantworten erfordert eine Theorie von *x*. Die Frage nicht zu beantworten, heißt auf eine Theorie zu verzichten, zumindest für eine Zeit. Anstelle der Theorie tritt ein Identifikationstest: Wie können wir das Vorliegen von *x* – hier: Intelligenz – feststellen? Dieser Test darf also kein theoretisches Verständnis von *x* voraussetzen, denn eine Theorie von *x* haben wir ja nicht. Der Test beruht daher auf eine Teiltheorie von *x*. Ohne eine Teiltheorie hätte der Test auch gar keine Anbindung an die Ausgangsfrage. Und dieser Teil kann nur ein operationalisierbarer Teil der Theorie von *x* sein. Es muss sich nicht um den wichtigsten oder zentralsten Bestandteil der Theorie von *x* handeln. Es handelt sich um den Teil, an dem sich für Tests ansetzen lässt. Das Vorgehen ist also *pragmatisch* – mangels einer besseren Alternative.

Damit sollten jedoch auch die Testresultate so gesehen werden – und nicht im Nachhinein den Test als Definition von ‚Intelligenz’ betrachten.

§2

Die Situation beim TT ähnelt Teilen der heutigen KI. Im Vordergrund steht das Haben und nicht das Verstandenhaben. Verfahren der ‚automatischen Übersetzung’ etwa liefern heute pragmatisch akzeptable Übersetzungsvorschläge, ohne dass sie auf einer entwickelten systematischen Semantik beruhen. Der Automat kann Teiltheorien (wie ein Lexikon und eine Teilgrammatik) koppeln mit einem riesigen Datenbestand an Standardsätzen und einem *brute force* Durchsuchen und Zuordnen

von (Teil-)Sätzen. Man hat so etwas Ähnliches wie Bedeutungszuordnungen, ohne eine Theorie der Bedeutung zu haben.

Der extremste Fall dieses Vorgehens wäre das Auftreten einer starken KI (einer künstlichen Person), ohne dass man ein zureichendes Verständnis von Personalität und den Fähigkeiten besitzt, welche die Rationalität dieser Person ausmachen. Die künstliche Person hätte dann bezüglich des Zieles, anhand einer künstlichen Person die menschliche Personalität und Rationalität und Intelligenz zu verstehen, nichts gebracht.

§3

Der TT scheint, folgt man Turings Abwägen von Gegenargumenten und seinen Beispielen, zudem stark an die Frage nach dem persönlichen Bewusstsein gekoppelt zu sein. Turing bringt Beispiele, die auf spezifische menschliche Erfahrungen abheben. Gesucht wird somit nach mehr als bloßer Intelligenz oder bloßen ‚Denken' (im Sinne des Vorliegens von mentalen Prozessen). Mentale Prozesse im weiten Sinne, der auch Prozesse umfasst, die bei Menschen nicht mit Bewusstheit ablaufen, können auch in einem System ablaufen, das keinerlei Bewusstsein hat, oder jedenfalls (wie Tiere) kein Selbstbewusstsein – könnte man vermuten. Es böte sich also an, auf elementarere Fertigkeiten als Bewusstsein zu testen. Ansonsten wäre der TT ein spezieschauvinistischer Test, den allein eine menschliche Intelligenz mit entsprechend menschlicher Verkörperungserfahrung bestehen kann. Getestet würde dann auf Teilintelligenz, also intelligente Leistungen. Intelligente Leistungen kann man einfacher durch Verhaltensäquivalente verstehen oder sogar definieren.

Dies entspricht zum einen dem Vorhaben der schwachen KI, teilintelligente Automaten oder Maschinen zu bauen, ohne dass diese über Gesamtintelligenz verfügen.

Dies entspricht methodisch ungefähr solchen Ansätzen, welche das Intelligenz*vokabular* auf dispositionale Beschreibungen zurückführen wollen (etwa Ryles *The Concept of Mind*).

§4

Der TT hat nur Relevanz unter der Annahme, dass eine *verlässliche Beziehung* (mutmaßlich unterhalb der Schwelle der Explikation) besteht zwischen Intelligenz und Personsein auf der einen

Seite und dem Bestehen im Imitationsspiel auf der anderen Seite. Dies wiederum scheint aber nach einer explikativen Theorie zu verlangen.

Insbesondere, wenn ‚Denken' und ‚Intelligenz' komplett unverständlich wären – wie Turing behauptet – dann würden wir den Vorschlag der Ersetzung der Ausgangsfrage „Können Maschinen denken?" durch das Imitationsspiel gar nicht verstehen.

Gegeben diese methodische Unklarheit bei Turing muss man doch sagen, dass – auch wenn Turing keine operationale *Definition* von Intelligenz geben will – es sich (doch) um eine mindestens partielle Operationalisierung eines Teilverständnisses der mentalen Begriffe handelt. Operationalisiert wird zumindest der Identifikationsaspekt der Begriffe (d.h. des Teils eines Begriffs, der dazu dient, Anwendungsfälle des Begriffes von anderen Situationen abzugrenzen). Offen gelassen wird eine intensionale, die inneren Abläufe betreffende Erläuterung der Begriffe.

Am besten lässt sich so der Test verstehen als Beibringen von Belegen, die einen Schluss auf Maschinenintelligenz als *Schluss auf die beste Erklärung* rechtfertigen könnten.

§5

Erfolg oder Misserfolg beim TT ist unabhängig von der Richtigkeit der Church-Turing-These (CTT). Ein Automat könnte beim TT versagen, weil menschliche Intelligenz sich nicht TM-simulieren lässt, insofern sie Verfahren oder Prinzipien einschließt, die nicht berechenbar sind. In diesem Fall wäre die CTT dennoch nicht widerlegt, eben weil diese Komponenten auch nicht dem intuitiven Begriff des Berechenbaren entsprechen.

Kandidaten könnten Fähigkeiten sein, die unendliche oder indefinite Prämissenmengen involvieren (wie die Ω-Regel in der Prädikatenlogik 2ter Stufe). Die Ω-Regel erlaubt, aus einer unendlichen Menge von Instanzen den entsprechenden Allsatz abzuleiten. Auf den Allsatz wird strikt geschlossen, dieser nicht nur – wie in einer induktiven Logik ‚bis auf Weiteres' oder mit einer bestimmten statistischen Wahrscheinlichkeit angenommen. Diese Regel verfährt klarerweise nicht algorithmisch, da sie die Bedingung des finiten Inputs für einen Schritt eines Algorithmus verletzt. Sie trägt zur Negationsvollständigkeit der Prädikatenlogik 2ter Stufe und zur Unterscheidung der Prädikatenlogik 2ter Stufe von der kompakten Prädikatenlogik 1ter Stufe bei. Es handelt sich auch nicht um die – sogar intuitionistisch gültige – Vorgehensweise, von einem schematischen Beweis

des Einzelfalls auf den Allsatz zu schließen. Die Ω-Regel bietet neben einer Auswahlfunktion den paradigmatischen Fall eines nicht-konstruktiven Vorgehens. Es ist zumindest denkbar, dass wir über eine entsprechende infinite oder zumindest doch indefinite Intuition verfügen. Eine solche Intuition würde durch Überschauen von Instanzen die Gültigkeit des Allsatzes erkennen (in einer Art ‚Aha-Erlebnis'). Eine indefinite Intuition würde von einer unbestimmten Menge von Prämissen schließen, wobei die Anzahl der benötigten Prämissen, um den Allsatz zu erkennen, schwanken kann: in einem Fall vielleicht 517 in einem anderen 4503 oder 34.

In beiden Fällen (infiniter oder indefinite Intuition) gäbe es nicht *eine* finite TM-berechenbare Version einer solchen Regel. Im Falle einer indefiniten Intuition, die jeweils eine finite aber unbestimmte Anzahl von Prämissen benötigt, gäbe es zwar für jeden einzelnen Anwendungsfall eine TM-Berechnung, jedoch könnte eine finite Maschinentafel nur finit viele von diesen Regeln enthalten – und das könnten unter Umständen nicht ausreichend viele sein. Ausreichend wäre eine solche finite Kollektion wiederum für menschliche Intelligenz, wenn diese nur mit finiter Dauer und finitem Auffassungsvermögen ausgestattet ist.

§6

Eine weitere Kontroverse im Kontext des TT betrifft die Frage, ob eine Maschine intelligent sein kann, wenn sie doch ‚nur' programmiert sei. Dies weist zurück auf die Debatte um die Frage, was alles berechenbar ist. Wenn es hochkomplexe Computer und Berechnungsvorgänge gibt, lassen sich diese – aus den bekannten Argumenten des Funktionalismus heraus – nicht mehr erfolgreich auf eine physikalische Beschreibung, die explanative Funktion besitzen könnte, reduzieren. Das Vorhandensein einer Beschreibung auf der Programmebene (d.h. eine Beschreibung, welche in einer Programmiersprache die Algorithmen/Prozeduren angibt, die das Verhalten der Maschine steuern) könnte indessen ein Anlass sein, der Maschine Intelligenz abzusprechen, insofern man der Auffassung ist, dass (menschliche) Intelligenz immer mehr umfassen muss als bloße Programme in *diesem* Sinne (einer Beschreibung in einer Programmiersprache, welche die Algorithmen/Prozeduren angibt, die das Verhalten der Maschine steuern). Programme in *diesem* Sinne (von Sprachen wie C, Java, LISP, PROLOG etc.) lassen sich in deterministischen Assembler und Maschinencodes compilieren, enthalten also keine nicht-konstruktiven oder nicht-determinierte Schritte.

Aus dieser Perspektive wäre auch die Supposition „Angenommen wir haben eine Maschine, die den Turing-Test besteht und betrachten nun ihr Programm“, die gelegentlich in Diskussion um den TT gemacht wird, in ihrem zweiten Teil zurückzuweisen: selbst wenn wir – um eines Argumentes willen und angesichts eines Einräumens einer vielleicht gegebenen logischen Möglichkeit einer Maschine, die den TT besteht – den ersten Teil der Supposition zugestehen, bleibt der zweite Teil noch problematischer: eine Maschine, welche den TT besteht, wurde sicher nicht programmiert mit C, Java, LISP …