Philosophie / Informatik
Sekundarstufe II

Können Computer denken?

„Künstliche Intelligenz“ als Thema
für einen fächerübergreifenden Unterricht

Unterrichtsmodul 1: Die Frage stellen

Amt für Schule
1995

Freie und Hansestadt Hamburg
Behörde für Schule, Jugend und Berufsbildung

Freie und Hansestadt Hamburg
Behörde für Schule, Jugend und Berufsbildung
Amt für Schule

Können Computer denken?

„Künstliche Intelligenz“ als Thema
für einen fächerübergreifenden Unterricht

Unterrichtsmodul 1:
Die Frage stellen

Fachreferenten:	Dr. Uwe Heinrichs	Amt für Schule S 13
	Ulrich Polzin	Amt für Schule S 13/33
Verfasser:	Reinhard Golecki	Gymnasium Klosterschule

Hamburg, 1995 (PDF-Version März 2000)

Inhaltsverzeichnis

Vorbemerkungen **1**

Unterrichtsmodul 1: Die Frage stellen **5**

Einführung **5**

1 ELIZA/DOCTOR **7**

2 Fragebogen **10**

3 Der Turing-Test **12**

4 Psychologischer Exkurs: Behaviorismus **16**

5 Die Mehrdeutigkeit von „Denken“ **20**

Literaturhinweise **22**

Allgemeine Hinweise 22

Zum Unterrichtsmodul 1 22

Literaturverzeichnis **25**

Der allzu naheliegende Gedanke, daß Computer den Menschen im Denken ganz ähnlich seien und sich von ihnen nur durch einen Mangel an Gefühlen unterscheiden, leistet einer dualistischen und allzu vereinfachten Sicht der menschlichen Psyche Vorschub. Diesem Trend gedanklich entgegenzuwirken, ist eine der größten Herausforderungen im Zeitalter intelligenter Maschinen.

Sherry Turkle

Vorbemerkungen

Computer werden im Beruf und im privaten Alltag immer selbstverständlicher. Tips zur effektiven Nutzung haben selbst im Bereich der Geisteswissenschaften längst Grundsatzdiskussionen über das Für und Wider ersetzt. Auch in den Schulen werden Computer immer häufiger und in immer mehr Fächern als Werkzeug und Medium eingesetzt (und dafür vielleicht weniger als früher thematisiert und reflektiert). Neben dieser Normalität ist der Computer – ähnlich wie früher Uhr und Dampfmaschine – aber auch Symbol einer Epoche, mit ihm sind Hoffnungen, Erwartungen und Ängste sowie Veränderungen im Selbstverständnis der Menschen verbunden. Das gilt insbesondere für die Forschung und die Resultate im Bereich „Künstliche Intelligenz“ (KI).

Auch und gerade wenn man skeptisch gegenüber der KI-Propaganda ist, muß man zur Kenntnis nehmen, daß erstens Entwicklungen und Produkte aus diesem Bereich in allen Lebensbereichen eine zunehmend größere Rolle spielen werden, sei es als Expertensystem zur Bonitätsprüfung bei Kleinkrediten, als medizinisches Diagnosesystem, bei der automatischen Videoüberwachung oder bei der Entscheidung über das Abfeuern von Interkontinentalraketen, und daß zweitens die Propaganda selbst (auch ohne Realisierung der sehr hoch gesteckten Ziele) Auswirkungen auf den Glauben an technische Lösungen und auf unser Selbstbild hat. Neben der Kontroverse um die gesellschaftspolitischen und ökonomischen Chancen und Risiken der Informationstechnik (z. B. Datenschutz, Soft- und Hardwarezuverlässigkeit, Rationalisierung, (De-) Qualifizierungsprobleme) und den sich daraus ergebenden Konsequenzen für den Bildungsbereich tritt somit zunehmend auch die Frage in den Vordergrund, wieweit der Umgang mit Informationstechniken und die Resultate der KI-Forschung den Menschen – sein Wahrnehmen, Denken, Fühlen und Handeln, seine Kommunikation und sein Selbstverständnis – verändern.[1]

In diesem Zusammenhang wurden ein zunehmender „Imperialismus der instrumentellen

[1] Vgl. die Untersuchung von Sherry Turkle (1984) über den Einfluß des Computers auf das Denken über das Denken und das Selbstverständnis bei Kindern und bei KI-Forschern.

Vernunft“[2] und „das allmähliche Verschwinden der Wirklichkeit“[3] beklagt, die Befürchtung geäußert, „daß die Bedeutung des Denkens selbst Gefahr läuft, entstellt zu werden“[4] und „daß wir uns dem Computer annähern und dabei unsere Identität verlieren“[5]. Auf der anderen Seite wird der frühe und intensive Umgang von Kindern mit Computern gefordert, weil „sie Menschen helfen können, neue Beziehungen zu Wissen aufzubauen, die über alle traditionellen Grenzen hinweggehen, die heute die Geistes- von den Naturwissenschaften und die Kenntnis des eigenen Ichs von allen beiden trennen.“[6] Ohne die hier angedeutete Kontroverse endgültig entscheiden zu wollen, bin ich allerdings sicher, daß die sich daraus ergebenden pädagogischen und didaktischen Herausforderungen im traditionellen, getrennten Fachunterricht kaum angenommen werden können, denn die „Computer-Kultur braucht etwas, was die Arbeit am Computer von sich aus nicht hervorbringt, ja, zu vernachlässigen geneigt ist: die Fähigkeit zum philosophischen Zweifel, die Kraft zum moralischen Handeln, die Bereitschaft zu politischer Verantwortung.“[7]

Das sind große und wichtige Themen – mit gutem Grund zählt Klafki „die Gefahren und die Möglichkeiten der neuen technischen Steuerungs-, Informations- und Kommunikationsmedien“ zu den „epochaltypischen Schlüsselproblemen“[8] –, und die Unterrichtsmodule 1 und 2 können sicher nicht den Anspruch einlösen, sie vollständig zu behandeln. Das Ziel ist bescheidener: Sollen setzt, bei Menschen und Maschinen, Können voraus, und so ist es erstmal ganz der grundsätzlichen Frage gewidmet, ob Computer „wirklich“ denken können und was die Antworten darauf über den Menschen sagen. Für dieses Vorhaben ist ein fächerübergreifender Unterricht — Informatik *und* Philosophie — besonders geeignet, denn beide Fächer sind bei dieser Frage ergänzungsbedürftig, Arbeitsweise, Leistungen und Grenzen von Computern können im „reinen“ Philosophieunterricht nicht erfahren werden, andererseits umfaßt „Denken“ sehr viel mehr als Datenverarbeitung; der Kontrast und die wechselseitige Herausforderung beider Sichtweisen ist, im Rückbezug auf den Menschen, ein Beitrag zur „Selbstreflexion des Wissens“ (und zur Frage „Was ist der Mensch?“).[9] Fächerübergreifender Unterricht ist hier allerdings nicht nur der Sache angemessener, er fördert auch ein „Denken in größeren Zusammenhängen, über den Fachhorizont und die in ihm geltenden Problemdefinitionen hinaus“, wenn durch „einen ‚fremden‘ Blick, aus einer anderen Perspektive“, die fachspezifische „Konstruktion der Wirklichkeit in ihrer Differenz zu anderen als solche erkannt wird, ihre Grenzen deutlich bewußt, die Angewiesenheit des einen Spezialisten auf die anderen und, wie wir heute sagen würden, auch auf die Wahrnehmungen der Laien, insbesondere hinsichtlich der Relevanz, erfahrbar gemacht werden.“[10]

---

[2] Joseph Weizenbaum (1977), S. 337 ff.

[3] Hartmut von Hentig (1987b).

[4] Theodore Roszak (1986), S. 11.

[5] Walter Volpert (1985), S. 145.

[6] Seymour Papert (1982), S. 26 f.

[7] Hartmut von Hentig (1988), S. 100 und (1993), S. 68.

[8] Wolfgang Klafki (1991), S. 59

[9] Das gilt auch für den fachwissenschaftlichen Diskurs. In der KI spielten klassische Fragen und Probleme der Philosophie — etwa aus den Bereichen Logik und Erkenntnistheorie — von Anfang an explizit und implizit eine große Rolle. Auf der anderen Seite lenken beispielsweise die für viele Informatiker überraschenden Probleme maschinellen Sprachverstehens neuerdings den Blick auf Hermeneutik und Phänomenologie. Umgekehrt sind — neben ethischen Fragen — für die Philosophie die Ergebnisse der Informatik Anlaß, traditionelle Konzepte von Seele, Geist, Bewußtsein, Intentionalität, Rationalität usw. zu überdenken und die KI-Forschung im gewissen Sinne als „experimentelle Philosophie“ zu verstehen.

[10] Ludwig Huber (1993), S. 59, 65. Zur Begründung fächerübergreifenden Unterrichts auf der Sekundarstufe II vgl. auch Golecki (1994).

In den Schulen gibt es so gut wie keine Erfahrungen mit KI-Systemen, aber das Thema ist dort in anderer Weise angekommen, durch mehr oder weniger zutreffende Berichte in den Medien und vor allem durch moderne Mythen (z. B. *Science-Fiction*-Filme). Im Informatikunterricht habe ich immer wieder das große Interesse an diesem Thema erfahren und dann weiter festgestellt, daß die gegenwärtige und künftige Leistungsfähigkeit von Computern erheblich überschätzt wird. Eine gründlichere Behandlung dieser Frage im Fachunterricht stieß schnell an Grenzen, da die Voraussetzungen und der Raum zur Reflexion darüber fehlten, was genau die Computer leisten müßten, um als „intelligent“ zu gelten, und wie z. B. solche Begriffe wie „Denken“ und „Bewußtsein“ bestimmt werden können. Im Philosophieunterricht, etwa bei Fragen der Anthropologie und Erkenntnistheorie, machte ich die Erfahrung, daß auch bei den Schülerinnen und Schülern der Computer in Umkehrung der ursprünglichen Verhältnisse mehr und mehr Metapher und Modell für das menschliche Denken wird, nach dem „Elektronengehirn“ der fünfziger Jahre nun das „Computermodell des Geistes“. Auch hier war eine vertiefte Behandlung nicht möglich, da z. B. eine gemeinsame und zielgerichtete Erfahrung mit dem Programmieren von Computern fehlte („Begriffe ohne Anschauungen sind leer, Anschauungen ohne Begriffe blind“).

Daraus entstand dann der Plan, das Thema „Können Computer denken?“ in einer Kurskombination der Fächer Philosophie und Informatik zu behandeln. Nach und zwischen getrenntem Fachunterricht gab es zwei fächerübergreifende Phasen. In der ersten ging es um die Frage, was wir denn genau meinen, wenn wir nach der Intelligenz und Denkfähigkeit von Computern fragen, und anhand welcher Merkmale wir sie überprüfen können (Unterrichtsmodul 1). In der zweiten Phase wurde die Fragestellung der ersten Phase zugespitzt auf das Problem des *Sprachverstehens* als arttypische menschliche Intelligenzleistung und wesentlicher Beitrag zum Denken. Denn das maschinelle Sprachverstehen ist nicht nur wegen zahlreicher wichtiger Anwendungen (als Schnittstelle zwischen Mensch und Maschine, als Erklärungskomponente bei Expertensystemen, bei Auskunfts- und Dialogsystemen, bei der Datenbankrecherche, bei der automatischen Text- und Spracherfassung und bei der maschinellen Übersetzung technischer Dokumentationen und von Dokumenten innerhalb der EU) ein zentrales Problem der KI, sondern auch weil das ganze Projekt der Nachbildung der menschlichen Intelligenz mit der Lösung dieses Problems steht und fällt. Im Mittelpunkt steht dabei die Frage, ob sich auch die semantischen und pragmatischen Aspekte von Kommunikation und Sprachverstehen durch bloße Symbolverarbeitung nach syntaktischen Regeln einfangen lassen (Unterrichtsmodul 2).

Möglicherweise gibt es wichtigere Fragen als die, ob Computer denken können. Aber richtig verstanden und gewendet kann eine Reflexion über die Frage sehr wohl ein Beitrag sein zur Gestaltung von Arbeit und Gesellschaft durch die Entwicklung einer sozialen Technik (statt einer bloßen Sozialtechnologie) und zu einem durch empirische Erkenntnisse über intelligente Lebewesen und Systeme gestützten Bild von uns selbst als rationale und moralisch verantwortliche Personen, zur Selbstaufklärung und Selbstvergewisserung.

Nach Einschätzung der Schülerinnen und Schüler war die Bilanz des Unterrichts sehr positiv, insbesondere die Schülerinnen waren von diesem Anschauung, praktische Übung und gemeinsame Reflexion verbindenden und bei den Algorithmen an Sprachverarbeitung orientierten Ansatz sehr angetan. Ich persönlich kann mich diesem Urteil nur anschließen, auch mir hat der Unterricht viel Spaß gemacht.

Unterrichtsmodul 1

Die Frage stellen

Einführung

Es ist keineswegs ohne weiteres klar, was die Frage „Können Computer denken?“ genau fragt und angesichts welcher Computerleistungen man bereit ist, sie so oder so zu beantworten. Einen geradezu klassischen Vorschlag zu diesem Problem, auf den sich bis heute fast jede Diskussion bezieht, hat vor über 40 Jahren Alan Turing gemacht. Der nach ihm so genannte *Turing-Test*, die Adäquatheit seiner Umformulierung in „Sind Digitalrechner vorstellbar, welche sich beim Imitationsspiel bewähren?“ sowie seine Auseinandersetzung mit einer Reihe von *Apriori*-Einwänden gegen eine positive Antwort stehen deshalb im Mittelpunkt der in diesem Kapitel vorgestellten Unterrichtseinheit (Abschnitt 3). Vorher ist es sinnvoll, im Kurs eigene Erfahrungen beim „Dialog“ mit dem Computer zu machen (u. a. mit Joseph Weizenbaums ELIZA, Abschnitt 1) und eigene Antworten zu finden und zu dokumentieren (2). Die Frage nach der Denkfähigkeit von Computern führt zwangsläufig zu der, was wir beim *Menschen* unter „denken“ verstehen wollen (je nach Beleuchtung sehen wir beim Blick auf den Monitor mit etwas Anstrengung schemenhaft uns selbst). Der Text von Walter Schulz analysiert die unterschiedlichen Positionen und die Problematik einer eindeutigen Antwort (5). Der Exkurs über den Behaviorismus (4) ist dafür nicht unbedingt nötig, aber sehr anregend.

Für diese Unterrichtseinheit gibt es keine notwendigen Unterrichtsvoraussetzungen aus den Bereichen Philosophie und Informatik, sie eignet sich auch als Einstieg ohne weitere Vorkenntnisse. Möglicherweise ist zum Verständnis des Textes von Turing eine kurze Erläuterung zum Begriff „Digitalrechner“ und zu den Angaben über die Speicherkapazität nötig.

Der Zeitbedarf für das Unterrichtsmodul 1 beträgt ca. 15 Stunden.

Modularer Aufbau

Der hier und im folgenden Kapitel vorgestellte Unterricht zum Thema „Können Computer denken?“ wurde mehrmals (zuerst im Schuljahr 90/91) mit kleinen Variationen als

Kurskombination der Fächer Philosophie und Informatik in der Vorstufe erprobt (er ist aber mit entsprechenden Modifikationen auch für alle Semester der Studienstufe geeignet). Die schüleridentischen Lerngruppen in beiden Fächern für ein Jahr machten es leicht, den fächerübergreifenden Unterricht mit dem reinen Fachunterricht in Philosophie und Informatik abzustimmen und beides flexibel auf die insgesamt zur Verfügung stehenden Stunden zu verteilen. Mir ist bewußt, daß diese günstigen Bedingungen nicht überall gegeben sind. Deshalb habe ich für die fächerübergreifenden Phasen einen modularen[11] Aufbau gewählt: In weniger „privilegierten“ Situationen kann auch die Beschränkung auf eines der beiden Unterrichtsmodule sinnvoll sein.

Es können auch die einzelnen Module durch andere ersetzt werden, z. B. das im nächsten Heft vorgeschlagene Unterrichtsmodul 2 durch eins, in dem ein eigenes „Expertensystem“ (vgl. Teil 1, Abschnitt 3.3) programmiert wird und in dem über „Wissen“ und „Können“ nachgedacht wird. Weiter ist auch eine Zusammenarbeit mit weiteren Fächern möglich und wünschenswert, z. B. mit Psychologie (Unterrichtsmodul 1), Deutsch und Fremdsprachen (Unterrichtsmodul 2). Ich kann es mir auch als sehr reizvoll vorstellen, in einem dritten Modul in Zusammenarbeit mit der Biologie der Frage nachzugehen, ob sich das menschliche Denken unabhängig von unserer biologischen Herkunft und der speziellen „Architektur“ unseres Gehirns nachbilden läßt und ob konnektionistische Ansätze der KI dabei erfolgreicher sind als der Symbolverabeitungsansatz.[12] Damit hätte man dann den Computer beim Vergleich mit der Denkfähigkeit von uns Menschen auf drei unterschiedlichen Ebenen betrachtet: von außen als „black box“ im Modul 1, auf der Symbolverarbeitungsebene (Algorithmen und Datenstrukturen) im Modul 2, Architektur und Hardware in einem solchen dritten.

Verzahnung von fächerübergreifendem und Fachunterricht

Für die Anbindung an oder Einbettung in den Philosophieunterricht gibt es viele Möglichkeiten, z. B. als Ergänzung oder Vertiefung zu einem Kurs über die Frage „Was ist der Mensch?“ oder „Was kann ich wissen?“ Bei der Verzahnung von fächerübergreifendem und Fachunterricht habe ich gute Erfahrungen damit gemacht, vor dem Unterrichtsmodul 1 im Philosophieunterricht Fragen der Anthropologie, insbesondere nach dem Unterschied

[11] Ich stimme Joachim Metzner (1987, S. 145 ff.) darin zu, daß von der Elektronik über die Möbelindustrie bis zu den Video-Clips dieser Begriff „eine der wichtigsten industriellen Fertigungsideen der letzten Jahrzehnte“ bezeichnet, daß – wie auch hier bei mir – „der einst technische Begriff neuerdings auch zur bildungspolitischen Vokabel avanciert ist“ und daß Programmiersprachen wie auch das von mir eingesetzte LOGO „daraus ein strikt modulares Konzept entwickelt“ haben. Diese „Tendenzen sind älter als der Computer, und er hat sie nicht einmal wesentlich verstärkt. Er ist eher selbst ein Produkt des Modularisierungstrends“, denn der Computer „verkörpert bestimmte Gesetze menschlichen Handelns, ist eine Materialisierung ursprünglich geistiger Arbeit. In den Konstruktionsprinzipien der Technik (Hardware) und mehr noch im Bau der Programme (Software) stecken bestimmte Strukturen menschlichen Denkens und Verhaltens, und diese, nicht das Gerät, wirken auf den Benutzer ein.“ Auch bei dem von mir gewählten modularen Aufbau muß somit die Frage im Bewußtsein präsent bleiben, ob wir mit dieser Figur „eine Seite unserer allgemeinen Menschennatur“ verstärken, „diejenige, mit der wir uns die Welt in linearen Abläufen, logischen Verhältnissen, gleichförmigen und gleichbleibenden Regeln verfügbar zu machen suchen“ (Hartmut von Hentig (1993), S. 39).

[12] Vgl. etwa die Kontroverse zwischen John R. Searle und Paul M. Churchland & Patricia Smith Churchland im *Spektrum der Wissenschaft*, März 1990, S. 40–54. Vorschläge für die unterrichtliche Behandlung neuronaler Netze machen u. a. G. Rosenfeldt (1993) und W. Steup (1995, S. 188 ff. und 295 ff., 1995 f.).

von Mensch und Tier, zu behandeln. Die These von DESCARTES, der jeder möglichen Maschine eine eigene *Sprachfähigkeit* und *universelle* Handlungsmöglichkeiten abspricht[13] (*Von der Methode*, Fünfter Teil, 10.; s. o. Teil 2, Abschnitt 2.2), bildet dann einen guten Übergang zur Frage nach dem Unterschied von Mensch und Computer und damit zum Unterrichtsmodul 1. Auch in der Informatik gibt es eine Fülle von Themen, an die man anknüpfen oder mit denen man weitermachen kann (z. B. Chiffrieren, Programmieren eines eigenen ELIZA, Produktionssysteme, Parser und Compiler, Aspekte der theoretischen Informatik). Nach diesem Modul haben wir im Informatikunterricht einfache Programme in LOGO erstellt, u. a. zum Verschlüsseln von Texten, und den Algorithmusbegriff behandelt, es folgte als zweite fächerübergreifende Phase das Unterrichtsmodul 2. Um das Thema nicht zu Tode zu reiten, endete das Schuljahr mit „normalem“ Fachunterricht (in Philosophie das Leib-Seele-Problem mit den in der Fußnote 12 genannten Texten von SEARLE und den CHURCHLANDS als Überleitung, in Informatik Textverarbeitung unter besonderer Berücksichtigung von Korrekturhilfen).

1 ELIZA/DOCTOR

Unterrichtsskizze

Nach einer kurzen Einführung in das Thema und der Formulierung des Ziels, sowohl die informationstechnischen als auch die philosophischen Aspekte der Frage „Können Computer denken?“ zu behandeln, wurden den Schülerinnen und Schülern Disketten mit einer Version des ELIZA-Programmes gegeben. Sie hatten dann ausführlich Gelegenheit, zu zweit mit dem Programm zu „spielen“ und es zu erkunden unter der Fragestellung

- Was genau leistet das Programm?
- Ist das Programm „intelligent“?

Anschließend wurden die Beobachtungen und Einschätzungen im Plenum zusammengefaßt. Die Schülerinnen und Schüler bemerkten – z. T. durch „unsinnige“ oder immer wieder gleiche Eingaben – recht schnell, daß das Programm „antwortet“, indem es entweder

- einen eingegebenen Satz nach ganz bestimmten Mustern zu einer Frage umformt,
- oder einen floskelhaften Satz ausgibt,
- oder einen früher eingegebenen Satz wieder aufnimmt,
- und dabei gewisse Variationen vornimmt.

Die Frage nach der „Intelligenz“ des Programmes wurde einstimmig verneint, sehr viel schwerer fiel den Kursteilnehmern allerdings eine genaue Begründung dafür. Ein Einwand

[13] Vgl. dazu auch das Unterrichtsbeispiel von Matthias TICHY (1996), der diesen Gedanken vertieft und in einer fächerübergreifenden Unterrichtseinheit der Fächer Philosophie und Informatik mit DESCARTES' Methodenlehre eine Brücke zur Informatik schlägt.

gegen die Intelligenz von ELIZA war, daß das Programm nicht lernfähig sei. Daraufhin wurde den Schülerinnen und Schülern ein weiteres Programm zum „Spielen“ gegeben. Dieses Programm kann „Tiere raten“, indem es Fragen stellt („Hat das Tier vier Beine?“, „Ist das Tier ein Haustier?“), die die Benutzer mit „ja“ oder „nein“ beantworten müssen. Danach „rät“ das Programm das Tier und fragt im Falle einer falschen Lösung nach einer mit „ja“ oder „nein“ zu beantwortenden Frage, mit der sich das gedachte Tier von dem fälschlich genannten unterscheiden läßt. Auf diese Weise „lernt“ das Programm, weitere Tiere richtig zu raten (die Kursteilnehmer fanden schnell heraus, daß sie dem Programm auch Phantasietiere oder Gegenstände des täglichen Bedarfs beibringen konnten).

Danach wurde kurz über den Aufbau und die Entstehungs- und Wirkungsgeschichte des Programmes ELIZA/DOCTOR informiert und dazu die Seiten 17–20 aus Weizenbaum (1977) vorgelesen und besprochen (auch im Hinblick auf Möglichkeiten und Sinn einer computergestützten Psychotherapie).

Kommentar

Das Programm ELIZA hat – trotz der Verwendung einiger KI-Techniken – nur wenig mit KI und kaum etwas mit Sprachverstehen (im Sinne des Unterrichtsmoduls 2) zu tun, dennoch taucht es – hauptsächlich wegen seiner Wirkungsgeschichte – in der Diskussion über KI immer wieder auf. Es wurde zum Test und zur Demonstration einer interaktiven Arbeit mehrerer Nutzer gleichzeitig (statt der Abgabe eines Stapels Lochkarten und der späteren Sichtung des Resultates) mit einer Großrechenanlage geschrieben (und wohl auch, um Besucher und Geldgeber zu beeindrucken). Es sucht vom Benutzer eingegebene Sätze nach vorgegebenen Schlüsselworten ab und produziert je nach dem, welche es vorfindet, bestimmte Antworten. Die Schlüsselworte, ihre Hierarchie sowie die von ihnen abhängigen Antworten sind in einem „Script“ zusammengefaßt. Eines dieser Scripts und die berühmteste Version von ELIZA ist DOCTOR.

Einige Beispiele für die Regeln[14] von DOCTOR:

- Taucht im eingegebenen Satz das Wort „same“, „like“ (mit einer Form von „to be“ irgendwo vorher) oder „alike“ auf, so gebe den Satz „IN WHAT WAY“ oder „WHAT RESEMBLANCE DO YOU SEE“ oder … oder … aus.
- Hat ein Satz die Form „blah-blah$_1$ you are blah-blah$_2$“, so gebe „WHAT MAKES YOU THINK I AM blah-blah$_2$“ oder „DOES IT PLEASE YOU TO BELIEVE I AM blah-blah$_2$“ oder „DO YOU SOMETIMES WISH YOU WERE blah-blah$_2$“ oder … aus (wobei „blah-blah$_i$“ für irgendwelche Worte steht).

[14] Wegen der einfacheren Grammatik lassen sich im Englischen solche Regeln viel leichter formulieren, das Ergebnis sieht eleganter aus und wirkt überzeugender. Das Beispiel von Weizenbaum (1977, S. 15 f.) ist die deutsche Übersetzung (also nicht Ergebnis einer deutschsprachigen Version) des in seiner Erstveröffentlichung zu ELIZA (1966, S. 36 f.) angegebenen Beispiels (er bezeichnet dort diese „Konversation“ als „typisch“; ich bezweifle das!). Eine deutschsprachige Version, die bei genau dieser Eingabe genau diese Ausgabe produziert, wäre nur mit sehr großem Aufwand (oder *ad hoc*) möglich. Abgesehen von der Doppeldeutigkeit der Worte „gleich“ und „wie“ bereitet die Wortstellung dabei große Probleme (z. B. Verb in Nebensätzen am Ende). Die Stellung von „NOCH“ in der achten, das Weglassen von „jedenfalls“ in der sechsten und der Wechsel von „besonders“ zu „SEHR“ in der elften Computerausgabe würden beispielsweise die vergleichsweise einfachen Möglichkeiten zur Mustererkennung von ELIZA sprengen (ganz abgesehen von noch freieren Übersetzungen wie bei Stephanie Sand (1986), S. 81 f.).

- Findet sich im Satz keines der vorgegebenen Schlüsselworte oder Muster, gib „PLEASE GO ON“, „THAT'S VERY INTERESTING“, „I SEE“ oder ... aus.

Die verschiedenen Ausgabemöglichkeiten werden der Reihe nach abgearbeitet, bevor sie sich wiederholen. Alternativ zur letzten Regel können – von einem Zufallsmechanismus gesteuert – auch Sätze wie „DOES THAT HAVE ANYTHING TO DO WITH THE FACT THAT YOUR blah-blah$_2$“ ausgegeben werden, wobei auf zwischengespeicherte Benutzereingaben der Form „blah-blah$_1$ my blah-blah$_2$“ zurückgegriffen wird. Das alles ist clever ausgedacht, aber kein Beispiel für *Verstehen*, sondern nur für die Illusion davon, d. h. eher ein Taschenspielertrick (und Joseph Weizenbaum hat nie etwas anderes behauptet).

Die Parodie eines psychotherapeutischen Erstgespräches durch DOCTOR wurde nach Weizenbaums eigener Aussage vor allem deshalb gewählt, weil es auf diese Weise möglich ist, so gut wie kein *explizites* Wissen „über die Welt“ im Programm zu repräsentieren und dennoch glaubhaft zu bleiben: die Frage „Was bedeutet für sie xyz?“ wird in diesem Kontext für sehr bedeutungsvoll und nicht für die einer oder eines absolut Ahnungslosen gehalten (und er weist darauf hin, daß Glaubwürdigkeit und Bedeutung *unser* Beitrag ist, nicht der der Programmierer). Unter den vielen Varianten schien ihm die *klientenzentrierte Gesprächspsychotherapie* nach Carl R. Rogers besonders geeignet, da es zu deren Prinzipien gehört, die Ratsuchenden als absolut gleichberechtigt anzusehen (daher auch die Bezeichnung „Klient“ statt Patient), möglichst einfühlsam ihren Standpunkt einzunehmen und das deutlich zu zeigen, keine Ratschläge, sondern Hilfe zur Selbstexploration zu geben, kurz also möglichst viele der Äußerungen widerzuspiegeln und zurückzugeben (*dieses* und anderes Wissen von J. Weizenbaum ist allerdings *implizit* in ELIZA repräsentiert).

Obwohl Weizenbaum das alles offen darlegte und die erste Publikation zu ELIZA ausdrücklich schrieb, um dem Programm die „magische Aura“ zu nehmen, mußte er danach drei für ihn schockierende Erfahrungen machen. Erstens veröffentlichte Kenneth Colby, ein Psychiater, der mit Weizenbaum bei der Entwicklung von DOCTOR zusammengearbeitet hatte, in einer nervenärztlichen Fachzeitschrift einen Artikel, in dem das Programm als Schritt zu einer computergestützten Psychotherapie zur Behebung des Mangels an menschlichen Therapeuten vorgestellt wurde.[15] Weiter erlebte er, daß Menschen, die gut über das Programm Bescheid wußten, eine emotionale Beziehung zu DOCTOR entwickelten und um „intime“, unprotokollierte Sitzungen baten. Drittens wurde das Programm fälschlich als bedeutender Beitrag zum maschinellen Sprachverstehen angesehen. Das alles führte dazu, daß er sich zu einem der entschiedensten Kritiker der KI entwickelte und später u. a. forderte, daß – unabhängig von der technischen Realisierbarkeit – in zwei Bereichen auf keinen Fall Computer eingesetzt werden sollten: das Ersetzen menschlicher Funktionen, die mit „gegenseitigem Respekt, Verständnis und Liebe“ zusammenhängen, und das automatische Erkennen und Verstehen gesprochener Sprache (1977, S. 350 ff.).

Die Schülerinnen und Schüler sollten auf jeden Fall Gelegenheit haben, ausführlich dieses Programm zu erproben. Die dabei gemachten Erfahrungen, die anfängliche Verblüffung, das allmähliche „Demaskieren“ bis hin zu als „Echo“ zurückgegebenen Tippfehlern sind durch keine gedruckten Beispiele zu ersetzen. Um eine reine Benutzerperspektive zu gewähr-

[15] Inzwischen wird von seinem Sohn Peter Colby in den USA eine Therapiesoftware „Overcoming Depression“ als angeblich „probates Mittel gegen leichtere bis mittlere Depression“ angeboten (vgl. dazu das Gespräch mit J. Weizenbaum in *c't*, 1/1993, S. 42 ff. und *c't*, 8/1993, S. 80 ff.).

leisten, empfiehlt es sich, eine kompilierte bzw. *stand alone* Version zu verwenden.[16] Im Rahmen dieser Unterrichtseinheit ist m. E. diese „behavioristische“ Perspektive völlig ausreichend, eine genauere Analyse des Programmes über das von den Schülerinnen und Schülern selbst Entdeckte und einige wenige allgemeine Erläuterungen (s. o.) hinaus ist nicht nötig. (Das schließt eine *spätere* Behandlung, z. B. als Programmierprojekt, nicht aus.) Auf jeden Fall sollten auch die Wirkungsgeschichte sowie die Problematik einer psychiatrischen oder psychotherapeutischen Exploration und Therapie mittels Computer angesprochen werden.

Das Tiere „erratende“ Programm ist ein beliebtes Programm zur Demonstration der Datenstruktur binäre Bäume und steht in der einen oder anderen Form sicher jeder Fachkonferenz Informatik zur Verfügung. Im Kern geht es um einen binären Baum, in dessen Knoten Fragen gespeichert sind, die mit „ja“ oder „nein“ beantwortet werden müssen („Ist das Tier ein Säugetier?“). Je nach Antwort wird zum nächsten Knoten gegangen. In den Endknoten wird das Tier „geraten“ („Ist es eine Katze?“) und beim Mißerfolg wird das richtige Tier und eine spezifizierende Frage gefordert. Die Frage und die Antwort mit dem neuen Tier werden dann an den Baum angefügt. Auf diese Weise ist das Programm „lernfähig“.

2 Fragebogen

Unterrichtsskizze

Die Diskussion über die „Intelligenz“ der beiden Programme führte zu der Frage, was denn ein Computer leisten müßte, um die Kursteilnehmer von seiner Intelligenz zu überzeugen (und die führte zu der weiteren, wie denn Intelligenz bei Lebewesen zu definieren und festzustellen sei).

Daraufhin erhielten die Schülerinnen und Schüler den Auftrag, sich in Kleingruppen Aufgaben auszudenken, deren Bewältigung durch einen Computer für sie untrügliches Zeichen seiner Intelligenz wäre. Ihre Antworten – z. B. den Schachweltmeister besiegen, ein Mittel gegen AIDS finden, einen guten Witz von einem schlechten unterscheiden können, in der Medizin die richtige Diagnose stellen und eine Therapie vorschlagen, ein Gedicht gut übersetzen, ein Gedicht interpretieren – wurden gesammelt und in Form eines Fragebogens dazu, ob und, wenn ja, wann sie solche Leistungen bei Computern erwarten (schon heute, in ca. 10 Jahren, in 50 Jahren, niemals), zurückgegeben und anschließend ausgewertet.[17]

Das Ergebnis der Umfrage (mit ihren kontroversen, aber auch Tendenzen aufzeigenden Antworten) gab neuen Anlaß zur Diskussion mit offenen Fragen: muß möglicherweise unterschieden werden, ob die Computer speziell für diese Aufgabe programmiert wurden, ist es dann eine *eigene* Intelligenz, sind im anderen Falle die Aufgaben „fair“ (auch wir

[16] Z. B. das der Arbeit von Wilmar Steup (1995) beiliegende Programm. Vgl. dazu auch die Handreichungen von Jörg Beleites (1989) und Renate Seidler & Lorenz Hilty (1994, Abschnitt 4). Fur eigene Versuche gibt – in einer Art LISP – Weizenbaum (1966) hinreichend Auskunft.
Ein ähnliches Angebot mit deutlich erweiterter Domne bietet `www.charlie.de`.

[17] Die Fragebogenaktion läßt sich gut als Beispiel einer Anwendung konventioneller Datenverarbeitung durchführen: Bei und nach der Diskussion die Beispiele der Kursteilnehmer in ein vorbereitetes Leerdokument tippen, gleich ausdrucken, ausfüllen und auswerten.

können teilweise die Aufgaben nicht oder nur mit spezieller Ausbildung lösen), sind sie tatsächlich ein guter Test für eine universelle, menschenähnliche Intelligenz?

Kommentar

Der Fragebogen hat die wichtige Funktion, die nur schwer zu beantwortende Frage, was denn Programmen wie ELIZA genau fehle, um ihnen Intelligenz oder Denkfähigkeit zusprechen zu können (d. h. indirekt die Frage nach einer allgemeinen Definition), an Beispielen zu konkretisieren. Das Nachdenken über eigene Kriterien und die sich daraus ergebenden offenen Fragen ist eine gute Vorbereitung für die Diskussion des im folgenden Unterrichtsabschnitt vorgestellten Turing-Tests (und sollte daher ebenso wie das eigene Erkunden von ELIZA unbedingt vorher erfolgen). Weiter ergibt sich aus der zusätzlichen Frage, wann die Kursteilnehmer diese Leistungen von Computern erwarten, die Möglichkeit, das Vorwissen sowie die Hoffnungen und Ängste in Bezug auf KI zu erkunden. Zum Abschluß der gesamten Unterrichtssequenz (nach Modul 2) wurde den Schülerinnen und Schülern der gleiche Fragebogen wieder vorgelegt und sie erhielten nach dem Ausfüllen ihre erste Antwort zum Vergleich. Das erwies sich für die Reflexion über den Unterricht und den Lernprozeß als sehr wertvoll.

Die Auswertung des Fragebogens zeigte deutlich die Tendenz, den Unterschied zwischen Mensch und Computer hauptsächlich im affektiven, emotionalen, subjektiven Bereich zu sehen, die „Vorstellung vom vernunftbegabten Tier weicht einer paradoxen Deutung des Menschen als emotionale Maschine“[18]. Die mit diesem Dualismus von Emotionalität und Rationalität, mit ihrer Sicht als völlig getrennte Bereiche menschlichen Vermögens, verbundene tendenzielle Überschätzung der Formalisierbarkeit letzterer[19] ist auch eine gefährliche Quelle für ein blindes Vertrauen in Expertensysteme und andere KI-Produkte. Allerdings erscheint mir – wie im letzten Absatz angedeutet – eine Diskussion darüber erst sinnvoll, wenn diese Problematik im weiteren Unterricht vertieft und erfahrbar wurde.

Ich selbst habe das Beispiel „Kreide holen“ zum Fragebogen beigetragen. Damit sollte auch ein Akzent gegen eine – auch durch den Turing-Test nahegelegte – bloß „logozentrische“ Sicht der KI gesetzt und die spätere Behandlung der Frage, wieweit eine Intelligenz ohne „Sensoren“ und „Effektoren“ und ohne *eigene* Kenntnis der Außenwelt vorstellbar ist, vorbereitet werden. Auch hier ist ein Unterschätzen der Problematik deutlich. Zur Erfüllung des Auftrages ist neben einer entwickelten Sensorik viel Wissen (z. B. über das Gebäude und den eigenen Standpunkt) nötig und es werden recht komplexe Entscheidungen getroffen. (Was mache ich, wenn die Hausmeisterloge geschlossen ist? Ist es nicht günstiger, auf dem gleichen Flur eine leere Klasse zu suchen oder höflich irgendwo anzuklopfen? Was genau ist gemeint, wenn eine Sportlehrerin diese Bitte beim Nachzeichnen von Linien auf dem Sportplatz äußert?)

Falls Zeit und genügend Interesse vorhanden sind, kann auch etwas genauer auf die Problematik einer allgemeinen Definition von „Intelligenz“ – vom „IQ“ über die Definition

[18] Sherry Turkle (1993), S. 72. Damit besteht die Gefahr, daß „das Kognitive möglicherweise zum rein logischen Prozeß, das Gefühlsmäßige zum Triebhaften degradiert“ wird (a. a. O. S. 73). Vgl. auch Turkle (1984), S. 71.

[19] Beispielsweise hielten es ca. 80 % der Kursteilnehmer für möglich, daß Computer heute oder in spätestens 10 Jahren ein Gedicht gut übersetzen können, aber mehr als 50 % bezweifelten, daß sie es jemals interpretieren können.

„Intelligenz ist die allgemeine Fähigkeit eines Individuums, sein Denken bewußt auf neue Forderungen einzustellen; sie ist die allgemeine geistige Anpassungsfähigkeit an neue Aufgaben und Bedingungen des Lebens“ von William STERN bis zur eher kybernetischen Sicht „Intelligenz ist Aufnahme, Speichern und Verarbeiten von Informationen“ – und eines dabei eventuell jeweils vorhandenen menschlichen „Chauvinismus“ gegenüber Tieren (das lohnt insbesondere dann, wenn schon Anthropologie vorher Thema im Philosophieunterricht war) oder Maschinen eingegangen werden (vgl. die Ausführungen im Abschnitt 1 von Teil 1 und die Literaturhinweise dazu).

3 Der Turing-Test

Unterrichtsskizze

Es wurden einige Informationen zur Person und Biographie von Alan TURING gegeben und dann eine gekürzte Fassung (TURING (1987b), Abschnitte 1–3 und 6 (gekürzt)) seines Aufsatzes *Rechenmaschinen und Intelligenz* abschnittweise gelesen und besprochen.

Die neue Frage (Abschnitte 1–3)

Nach der Klärung verschiedener Detailfragen (Was spricht nach TURING gegen die ursprüngliche Frage? Wie genau geht das Imitationsspiel? Welche Vorteile hat die neue Fragestellung? Welche anderen Maschinen wären denkbar?) stand hier im Mittelpunkt des Unterrichts eine Diskussion darüber, ob die Frage „Wird es eines Tages Digitalrechner geben, die erfolgreich am Imitationsspiel teilnehmen können?“ eine adäquate Ersetzung von „Können Computer denken?“ ist, ob es also stimmt, daß durch das Imitationsspiel festgestellt werden kann, inwieweit Computer menschenähnliche intellektuelle Fähigkeiten haben und ob es wirklich geeignet ist, „fast jeden Bereich menschlichen Bemühens einzubringen“ (und somit ein besserer Test ist als die Beispiele des Fragebogens). In diesem Zusammenhang wurden die von den Schülerinnen und Schülern bei der Erstellung des Fragebogens genannten Beispiele mit den von TURING angegebenen Musterfragen und -antworten verglichen und fiktiv weitere Dialoge skizziert.

Antworten, Einwände und Repliken (Abschnitt 6)

Zu Beginn wurde TURINGS eigene Antwort und Prognose diskutiert. Dann wurde ausführlich über das *Bewußtseinsargument* gesprochen. Es entbrannte auch hier die im Rahmen der KI heftig geführte Debatte, ob Computer zum „echten“ Denken oder nur zur Simulation von Denken fähig seien, ob man zwischen beiden überhaupt einen sinnvollen Unterschied machen kann und ob – falls ja – der Turing-Test dafür ein geeignetes Instrument ist. Weiter wurde die Frage aufgeworfen, wie wir denn bei unseren Mitmenschen sicher sein können, daß Denken, Humor, Trauer echt und nicht bloß simuliert sind, ob das mehr als eine „höfliche Übereinkunft“ ist und wie man dabei den Problemen des Solipsismus entgehen kann.

Bei der Besprechung der *Argumente, die verschiedene Unfähigkeiten betreffen*, legten die Schülerinnen und Schüler besonderes Gewicht auf die Rolle der Emotionen beim menschli-

chen Denken und die Möglichkeit emotionaler Reaktionen bei weiter entwickelten Computern (dabei wurde das von TURING angesprochene Problem, induktiv von den vorhandenen Computern auszugehen, ein wenig durch die „Erfahrung“ diverser Science-Fiction-Filme ausgeglichen). Der Einwand, daß Maschinen keine Fehler machen können, war für die Kursteilnehmer von geringerer Bedeutung (und nur sehr wenige hatten bemerkt, daß das Additionsergebnis bei den Musterantworten falsch ist), da die vermeintliche Fehlerfreiheit von Computern eher als ihre Stärke angesehen wird. Hier mußte die von TURING aufgeworfene Fragestellung erst in ihren Konsequenzen verdeutlicht werden, ob solche Fehler nur „absichtlich“ zur Täuschung der Menschen gemacht werden (und somit besonders clever sind) oder ob man nicht bei einer mit komplexer, menschenähnlicher Denkfähigkeit bzw. Intelligenz ausgestatteten Maschine, die auch die Fähigkeit zu induktiven Schlüssen hat, gerade deswegen mit Fehlern rechnen muß.

Nach einer kurzen Information über Charles BABBAGE, Lady Ada LOVELACE und die „Analytische Maschine“ wurde der *Einwand von Lady Lovelace* in seinen verschiedenen Fassungen (ein Computer kann (nur) das, wozu er programmiert wurde; uns nie überraschen; nie etwas wirklich Neues) besprochen.

Zur Illustration der Antwort TURINGS (es ist ein Trugschluß, daß mit bestimmten Tatsachen auch alle Folgen dieser Tatsachen unmittelbar offenbar sind) wurde ein kleines Experiment veranstaltet: Es wurden kurz die (allgemein bekannten) Konstellationen von Sonne, Mond und Erde bei einer Sonnen- bzw. Mondfinsternis ins Gedächtnis gerufen und dann Kopien eines Kalenders (Monatsübersicht) mit den dabei üblichen Symbolen für die Mondphasen (ohne auf diese besonders hinzuweisen) verteilt mit der Frage „Wenn man nur wüßte, daß in diesem Monat auf jeden Fall eine Sonnen- bzw. Mondfinsternis stattfindet, könnte man dann auch sagen, wann genau sie stattfindet?“. Dabei wurde deutlich, daß mit der Kenntnis der notwendigen Konstellation keineswegs auch die zwingende Schlußfolgerung, daß eine Sonnenfinsternis nur bei Neumond, eine Mondfinsternis nur bei Vollmond stattfinden kann, gegeben ist. Daran schloß sich eine Diskussion darüber an, inwieweit das Explizieren impliziten Wissens überraschen kann bzw. etwas wirklich Neues ist.

Von den Kursteilnehmern wurde als weitere Variation der Einwand vorgebracht, Computer könnten nie *kreativ* sein. Das führte zu der (sehr schwierigen) Debatte, was genau denn mit „Kreativität“ gemeint ist und ob man diese Fähigkeit programmieren kann. Am anderen Ende des Spektrums der Antworten stand die Einschätzung, daß auch wir – sei es durch Evolution und Gene, sei es durch Erziehung und Sozialisation (wie auch von TURING in seiner ersten Replik zu diesem Einwand angemerkt) – nur „programmiert“ sind und wir auch mit unseren kreativen Leistungen diesen Rahmen nicht verlassen. Diese Problematik, das Problem der Intentionalität (weitere Variation des Einwandes: Computer können nicht etwas *selber* wollen) sowie die oben im Zusammenhang mit dem Bewußtseinsargument aufgeworfenen Fragen wurden im hieran anschließenden psychologischen Exkurs vertieft.

Kommentar

Es ist sicher keine Übertreibung, daß Alan TURING mit seinem Aufsatz ein großer Wurf gelungen ist. Vor mehr als 40 Jahren warf er Fragen und stellte er Thesen auf, die bis heute im Mittelpunkt der KI-Diskussion stehen (und das in einer Klarheit und Verständlichkeit, die diesen Text besonders geeignet für den Unterricht machen). So bestechend das von ihm erdachte Imitationsspiel gerade in der Hinsicht ist, daß es offen dafür ist, welche

Kriterien man für „Denken“ bzw. „Intelligenz“[20] im einzelnen oder als einzelner hat, so wenig ist damit die Frage nach der *Adäquatheit* des Turing-Testes beantwortet (diesem Problem entgeht man auch nicht, wenn man den Turing-Test als *operationale Definition* maschineller Intelligenz auffaßt; der bei dieser Wortwahl wohl beabsichtigte Vergleich mit dem Menschen bedarf auch dann einer Rechtfertigung). Durch seine Beschränkung auf beobachtbares verbales Verhalten ist der Turing-Test sicher keine *notwendige* Bedingung dafür, daß ein Computer denken kann (ihn könnte das Spiel so langweilen, daß er es durch schlechte Antworten schnell beenden will; er könnte aus bestimmten Gründen seine wahre Intelligenz verheimlichen wollen). Schwerer zu klären ist die Frage, ob er eine *hinreichende* Bedingung ist. Könnten nicht – wie das Beispiel ELIZA zeigte – schon relativ simple Programme die menschlichen Schiedsrichter täuschen? Handelt es sich also um „eigene“ oder nur um vom Programmierer verliehene Intelligenz? Reichen schriftliche Antworten zur Beurteilung oder müßte man auch untersuchen, wie sich der Computer im „wirklichen Leben“, in Handlungen und Aktionen bewährt? Deckt der verbale, sprachliche Bereich alle relevanten Facetten ab? Kann man überhaupt „von außen“, ohne Kenntnis der internen Mechanismen und Vorgänge zu einem angemessenen Urteil kommen (das führt zu den grundsätzlichen Einwänden gegen den Behaviorismus, vgl. den folgenden Exkurs)?

Meine Antwort: Ich halte diese Einwände für so gewichtig, daß man wohl nicht mit Sicherheit sagen kann, ein Computer, der den Turing-Test besteht, habe eine mit Menschen vergleichbare Denkfähigkeit (letztlich wird es auch wohl eher um einen Akt der *Anerkennung* als um einen strengen Beweis gehen); allerdings wäre ich von einem Computer, der längere Zeit mit verschiedenen Juroren und über beliebige Themen erfolgreich am Imitationsspiel teilnehmen würde, ausgesprochen beeindruckt. Ein solcher Computer müßte nicht nur über viel „Weltwissen“ verfügen, das in keinem Lexikon steht, er müßte wohl auch eigene kausale Verbindungen (Sensoren und evtl. auch Effektoren) zur „wirklichen“ Außenwelt haben, um bei so simplen Fragen wie z. B. „Wie war denn das Wetter heute morgen, als du aufgestanden bist?“ nicht durch ausweichende Antworten aufzufallen; eine im jeden Detail „vorprogrammierte Intelligenz“ reicht sicher nicht (vgl. dazu die Abschnitte 5 und 6 im Teil 2).

Der genauen Regeln des Imitationsspiels sind beim ersten Lesen etwas schwer verständlich (das einleitende Spiel mit einer Frau und einem Mann mag ein wenig verwirren[21]). Hier empfiehlt sich auf jeden Fall eine Tafelskizze: ein Fragesteller (C) ist über Datenleitungen mit zwei Terminals X und Y verbunden und soll erraten, ob am Terminal X bzw. Y der Computer (A) oder der Mensch (B) „sitzt“ (in neueren Darstellungen wird häufig ein Terminal weggelassen). Zum besseren Verständnis der auf das Spiel folgenden Argumentation sollten Turings eigene These, die Einwände und die Repliken jeweils deutlich getrennt und kurz schriftlich festgehalten werden.

Turing hat sicher recht mit seinem Hinweis, daß eine genauere Betrachtung der Frage „Können Maschinen denken?“ neben einer Explikation von „denken“ auch einer von „Maschine“ bedarf. Das Problem wurde im Unterricht kurz angesprochen (Abschnitt 3 des Textes), nicht aber vertieft (Abschnitte 4 und 5 des ungekürzten Textes). Eine genauere Behandlung des Begriffes „Digitalrechner“ (z. B. Von-Neumann-Architektur, Church-

[20] Zwar ist im Text von „Denken“ die Rede, der Originaltitel des Aufsatzes bezieht sich aber auf „Intelligence“.

[21] Dazu, daß es sich hierbei um mehr als ein „exzentrisches Vorspiel“ handelt, und zu einer Analyse der möglichen Motive Turings für diesen Einstieg vgl. Werner Sesink (1993, S. 40 ff.) und – mit Hinweis auf die Unterscheidung von *sex* und *gender* – insbesondere Bettina Heintz (1993, S. 264 ff.).

Turing-These, Universalität) halte ich im Rahmen dieser einführenden Unterrichtseinheit weder für nötig noch für sinnvoll (das mag sich anders darstellen, wenn man den Aufsatz beispielsweise in oder nach einem späteren Kurs mit Themen aus der theoretischen Informatik liest). Weiter fiel der Abschnitt 7 („Lernende Maschinen") der Zeitschere zum Opfer. Das geschah mit etwas größerem Bedauern, denn viele der hier angesprochenen Aspekte können für den Unterricht durchaus lohnend sein. So ist beispielsweise ein Vergleich der von TURING vorgetragenen Gedanken mit der *tabula rasa* von John LOCKE (vgl. Teil 2, Abschnitt 2.3) oder mit bestimmten, speziell behavioristisch orientierten Lerntheorien ausgesprochen reizvoll.

Auch bei der Behandlung der in Abschnitt 6 angeführten Einwände gegen TURINGS eigene Position habe ich gekürzt. Die ausgewählten Einwände sind solche, die sowohl bis heute in der Debatte um die KI von großer Bedeutung als auch ohne weiteres Spezialwissen zu formulieren sind. Letzteres gilt z. B. nicht für den *mathematischen Einwand* (unter Berufung auf den Unvollständigkeitssatz von Kurt GÖDEL) und das *Argument der Stetigkeit innerhalb des Nervensystems*. Weiter halte ich auch das *Argument aus dem informellen Charakter des Verhaltens* für durchaus interessant (einige Aspekte dieses Problems wurden im anschließenden Exkurs und – soweit es das unter „Inferenzsysteme" im Abschnitt 3.2 von Teil 1 dargestellte „Rahmenproblem" betrifft – im Unterrichtsmodul 2 behandelt).

Das *Bewußtseinsargument*[22] ist eng verbunden mit einigen *Argumenten, die verschiedene Unfähigkeiten betreffen*. Zu den in der Fachdiskussion angeführten Unfähigkeiten gehören u. a. die, Selbstbewußtsein oder die Vorstellung von einem Ich zu entwickeln, beim Gebrauch der Sprache auch im vollen Sinne die Bedeutung der Worte zu verstehen, Intentionalität oder auch das mit dem „immer schon in der Welt sein" verbundene intuitive Situationsverständnis zu haben (vgl. dazu z. B. den im Abschnitt 5 behandelten Text von Walter SCHULZ oder die Arbeiten von Gotthard GÜNTHER, die Debatte um das Gedankenexperiment „Chinesisch-Zimmer" von John SEARLE (vgl. den Abschnitt 6 von Teil 2) und die Kritik der Brüder DREYFUS oder von Terry WINOGRAD und Fernando FLORES (vgl. den dritten Teilabschnitt zur Formalisierung der Semantik im Anhang vom Unterrichtsmodul 2). TURING mag diese Einwände geahnt haben, einen angemessenen Raum haben diese erst viel später ausformulierten Argumente in seinem Text selbstverständlich nicht. Es kann daher bei der Lektüre nicht das Ziel sein, sie an dieser Stelle in ihrer Breite und Tiefe zu behandeln (das geschah z. T. an den genannten Stellen später). Im Kern geht es bei all diesen Fragen um das Problem, ob und wodurch sich „echtes" Denken von bloß simuliertem unterscheiden läßt („rechnet" ein Taschenrechner „wirklich" oder „simuliert" er das nur?). Wie schon gesagt halte ich TURINGS Repliken gegen die Einwände nicht für unbedingt schlüssig, wohl aber müssen sich alle Opponenten die Frage gefallen

[22] Die feine Ironie des Textes von Alan TURING erschließt sich an dieser Stelle erst durch einige Hintergrundinformationen. Sir Geoffery JEFFERSON, ein Gehirnchirurg, hielt als Reaktion auf Norbert WIENERS *Cybernetics* einen als *Lister Oration* bekannt gewordenen Vortrag mit dem Titel *The Mind of Mechanical Man*. Die von TURING zitierten Zeilen enden mit den Worten: „Ich stelle mich eher mit dem Humanisten Shakespeare in eine Reihe als mit den Mechanisten, indem ich Hamlets Zeilen in Erinnerung rufe: ‚What a piece of work is man! How noble in reason! How infinite in faculty'." Dazu Andrew HODGES (1989, S. 467) dann weiter: „Shakespeare wurde in diesen Diskussionen oft als Beweis für die exquisiten menschlichen Empfindsamkeiten des Redners zur Schau gestellt. Jefferson hatte jedoch selbst eine Menge zur Verbesserung des ‚piece of work' beigetragen, nicht nur indem er die zerbrochenen Schädel zweier Weltkriege zusammenflickte, sondern auch als Exponent der Frontallobotomie [operativer Eingriff im vorderen Hirnbereich, der eine gewisse Zeit zur ‚Heilung' von psychischen Störungen in Mode war und körperliche, geistige und seelische Krüppel erzeugte (R. G.)] in den späten dreißiger Jahren."

lassen, ob die Begriffe „Bewußtsein“, „Selbstbewußtsein“, „Bedeutung“, „Intentionalität“ oder „Intuition“ hinreichend klar sind, um mit ihnen *Apriori*-Einwände gegen die KI zu stützen (vgl. die Abschnitte 6 und 7 im Teil 2 sowie den Abschnitt zur Formalisierung der Semantik im Anhang vom Unterrichtsmodul 2), bzw. wie es auf intersubjektiv und wissenschaftlich befriedigende Weise möglich ist, sie beim Menschen zu erforschen. Das war ein wesentlicher Grund dafür, diesen Fragen im anschließenden Exkurs über den Behaviorismus nachzugehen.

Der *Einwand von Lady Lovelace*, ein Computer könne ausführen, was immer wir ihm zu befehlen wissen, ist streng genommen so gar kein Einwand gegen die KI, er wird es erst durch das Hinzufügen eines im Original nicht vorhandenen „nur“. In dieser eingeschränkten und wörtlichen Form wird der Einwand fast nur von Leuten ohne Programmiererfahrung erhoben. Alle anderen wissen, daß er so nicht stimmt.[23] Computer machen häufig nicht das, was die Programmierer wollten, und es gehört zu den gesicherten Erkenntnissen der Informatik, daß das Verhalten von Programmen bei allen möglichen Eingaben weder immer theoretisch vorhergesagt noch ausreichend vorher getestet werden kann (und das gilt insbesondere für alle komplexeren Programme, die sich nicht mehr auf das Befehlswissen und -wollen nur einzelner Programmierer stützen können). Und sicher hat Turing recht, daß – wie auch das im Unterricht durchgeführte Experiment mit der Bestimmung des Zeitpunkts einer Sonnen- oder Mondfinsternis zeigt – mit den Prämissen nicht auch schon alle Konklusionen (und mit den Programmbefehlen nicht alle Konsequenzen) offenbar werden und somit auch das Explizieren impliziten Wissens Neuigkeitswert haben und überraschen kann. Subtiler ist allerdings die von Turing in seinem Abschnitt 7 aufgeworfene Frage, ob nicht möglicherweise immer komplexere Computerprogramme irgendwann eine „kritische Masse“ überschreiten und eine Art *eigenen* Verstand als emergentes Phänomen entwickeln könnten.

4 Psychologischer Exkurs: Behaviorismus

Einführung

Die Psychologie hat sich erst relativ spät als eigenständige wissenschaftliche Disziplin etabliert (neben anderen Gründen haben sicherlich der mit der Aufbruchsstimmung des mechanistischen Weltbildes eng verbundene strenge Dualismus von Geist und Materie und der Glaube an die absolute Sonderstellung des Menschen zu dieser Verzögerung beigetragen). Zu den Anfängen werden die sehr einflußreichen Versuche von Wilhelm Wundt im ausgehenden 19. Jahrhundert gezählt, die in der Auseinandersetzung zwischen Rationalismus und Empirismus entstandenen Konzepte über mentale Repräsentationen und Gesetze der Assoziation (vgl. den Abschnitt 2.3 im Teil 2) durch *Introspektion* unter kontrollierten

[23] Die weitergehende Ansicht, daß auch *wir* „in gewisser Weise nur ein Programm“ sind (Schank & Childers (1986), S. 69), ist in der KI-Forschung recht verbreitet. Ebenso deutlich wird die Irrelevanz des „Einwandes von Lady Lovelace“ von Seiten der Soziobiologie behauptet: „Es gibt ein bekanntes Schlagwort, ... das besagt, man könne aus Computern nicht mehr herausholen, als man hineingetan habe. Eine andere Version lautet: Computer tun exakt das, was man ihnen sagt, und sind daher niemals schöpferisch. Dieses Schlagwort ist nur in einem schrecklich belanglosen Sinne richtig, in dem Sinne nämlich, daß Shakespeare niemals etwas anderes geschrieben hat als das, was ihm sein Grundschullehrer beigebracht hat – Wörter.“ (Richard Dawkins (1987), S. 84 f.) Wir sind gespannt, ob seine computergestützte Evolution von aus Pixeln gebildeten „Insekten“ auch noch einen Winternachtstraum hervorbringt.

Bedingungen zu untersuchen. Nach großen Erfolgen insbesondere in Deutschland wurde dieser Forschungsansatz wegen widersprüchlicher, kaum intersubjektiv nachprüfbarer Ergebnisse zunehmend kritisiert und es wurden „objektive“, streng „naturwissenschaftliche“ Methoden für die Psychologie verlangt.

Besonders radikal war dabei das von John B. Watson begründete Forschungsprogramm des Behaviorismus. Seine These war, „daß das Bewußtsein weder ein definierbarer noch ein nützlicher Begriff ist“ und „daß der Glaube an die Existenz eines Bewußtseins aus den alten Zeiten des Aberglaubens und der Magie herrührt“; dagegen setzte er das Programm, „Gegenstand und Methode zu präzisieren“, „das Problem der Psychologie neu zu formulieren und dabei alle mittelalterlichen Vorstellungen über Bord zu werfen“, aus dem „wissenschaftlichen Vokabular alle subjektiven Termini wie Empfindung, Wahrnehmung, Vorstellung, Wunsch, Absicht und sogar Denken und Gefühl, soweit diese subjektiv definiert“ sind, zu streichen.[24] In den Mittelpunkt psychologischer Untersuchungen stellte er das Studium objektiv beobachtbaren *Verhaltens* von Lebewesen als Reaktion auf (nach dieser Theorie das Individuum fast vollständig determinierende) Umweltreize mit dem Ziel der genaueren Vorhersage und Kontrolle dieses Verhaltens.

Der behavioristische Ansatz hat in den USA die Psychologie vier Jahrzehnte lang bis in die späten fünfziger Jahre hinein weitgehend beherrscht (mit großen Auswirkungen auch auf die Pädagogik und die Lehrpläne), mit der Folge, daß die Erforschung „mentaler“ Strukturen und Prozesse einem wissenschaftlichen Tabu unterlagen. Nun benennt Watsons „Streichliste“ zwar so ziemlich genau das, was für Laien und Wissenschaftler im Bereich der Psychologie des Menschen interessant ist und was ihre ursprüngliche Neugier weckt, aber es dauerte erstaunlich lange, bis – zum einen durch sprachwissenschaftliche Arbeiten (u. a. von Noam Chomsky, vgl. im Unterrichtsmodul 2 den Abschnitt 5), zum anderen und noch wirkungsvoller durch den „Informationsverarbeitungsansatz“ und die KI-Forschung – die „kognitive Wende“ eingeleitet wurde. So verstanden sich die KI-Pioniere Allen Newell und Herbert Simon auch als Wegbereiter einer psychologischen Forschung, die, mit einem ebenso explizit „naturwissenschaftlichen“ Anspruch wie Watson, gegen sein Verdikt die *black box* des Mentalen öffnen und Theorien über Denkprozesse und Intelligenzleistungen experimentell mittels Computerprogrammen überprüfen will (und die Frage bleibt, ob der beim Problem der kognitiven Adäquatheit einer solchen Modellierung manchmal zu hörende Hinweis auf die gleiche Leistungsfähigkeit, auf den gleichen *output*, nicht auch behavioristisch ist). In letzter Zeit wurde innerhalb der KI die Debatte um den Behaviorismus durch eine Diskussion darüber neu entflammt, inwieweit die neue Richtung *Konnektionismus* „neobehavioristisch“ (mit den dafür geltenden Unzulänglichkeiten und Einwänden) sei. Kurz: der Behaviorismus gehört zur (Vor-) Geschichte der KI, und vieles versteht man besser, wenn man den Ansatz und den Erfolg der KI als spezifisch amerikanische Antwort auf, als „Befreiungsschlag“ gegen die „Diktatur“ des Behaviorismus ansieht – und manchmal auch als Fortsetzung mit anderen Mitteln.

Auch Alan Turing ist – bei der Behandlung des „Bewußtseinsargumentes“, des „Argumentes von Lady Lovelace“, im Abschnitt 7 über lernende Maschinen – in seiner Argumentation behavioristisch beeinflußt (und es gibt interessante Ähnlichkeiten zur Polemik von Gilbert Ryle gegen das „Dogma vom Gespenst in der Maschine“, vgl. Abschnitt 2.6 vom Teil 2). Allerdings kann ein Computerwissenschaftler nicht dabei stehen bleiben, das „Verhalten“ von Computern zu beobachten, er muß sie auch bauen und sich dabei sehr

[24] John B. Watson zitiert nach Anderson (1988), S. 20.

genau mit den *inneren* Strukturen und den ablaufenden Prozessen beschäftigen. Insofern haben seine Behandlung der Fragestellung, der (über die behavioristische Gleichbehandlung von Mensch und Tier hinausgehende) Vergleich von Mensch und Computer und seine Arbeiten zur Berechenbarkeit ironischerweise einen wichtigen Beitrag zur Überwindung des Behaviorismus geliefert.

Unterrichtsskizze

Ausgangspunkt war – im Anschluß an das Imitationsspiel und das dabei von TURING diskutierte „Bewußtseinsargument“ – die Frage, wie wir denn bei unseren Mitmenschen feststellen können, daß Denkfähigkeit, Verständnis, Trauer, Liebe etc. „echt“ und nicht bloß simuliert sind, und ob die analoge Übertragung eigenen Erlebens auf andere mehr ist als eine „höfliche Übereinkunft“.

Anschließend wurden kurz die wesentlichen Thesen und methodischen Forderungen des Behaviorismus vorgestellt (u. a. mit einem Text aus GARDNER (1989), S. 122 f.) und die Teilnehmer wurden aufgefordert, Beispiele aus dem einführenden Gespräch (z. B. einen Satz verstehen, traurig sein, verliebt sein) „behavioristisch“, als beobachtbares Verhalten, als Reaktionen auf Reize, zu beschreiben. Diese ungewohnte Sicht auf Vertrautes bot Anlaß und Material, über die Stärken und Schwächen des Behaviorismus zu diskutieren. Nach diesen Vorbereitungen war es möglich, einen Vergleich zwischen dem behavioristischen Ansatz und unserem Umgang mit ELIZA, mit dem im Unterricht erstellten Fragebogen und mit dem Turing-Test zu ziehen und unser Vorgehen zu bewerten.

Den Abschluß des Exkurses bildete ein kurzer Blick auf die behavioristische Lerntheorie (mit Textauszügen aus HELBIG (1979), S. 11, 86 und NEEL (1983), S. 150 f., 227) und ihre Auswirkungen sowie ein erneutes Aufgreifen der im Zusammenhang mit dem „Einwand von Lady Lovelace“ aufgetauchten Kontroverse, ob wir wirklich Neues denken und schaffen können oder im gewissen Sinne auch nur – sei es durch Umwelt oder Gene – „programmiert“ sind. Das konnte selbstverständlich nicht abschließend geklärt werden, aber es gab Anlaß, über unser Selbstverständnis von Menschenwürde, Autonomie und Willensfreiheit nachzudenken.

Kommentar

Es ist in der Tat nicht leicht, bei konsequenter Anwendung des „Bewußtseinsargumentes“ den Abgründen des Solipsismus[25] zu entgehen, und es ist in diesem Zusammenhang eine gute Übung, genau zu überlegen, was uns jeweils veranlaßt zu glauben, daß ein Verständnis des Unterrichtsstoffes, eine psychische Erkrankung, ein Orgasmus echt und nicht bloß simuliert sind.[26] Danach ist der radikale Ansatz WATSONS besser verständlich (und es ist wohl nötig, ihn zunächst zu stärken): die Forderung nach intersubjektiver Überprüfbarkeit, nach anerkannten „naturwissenschaftlichen“ Methoden, nach reproduzierbaren Ergebnissen. Auch sollte man bei einer Betrachtung der hohen „Kosten“ für die Dominanz des

[25] Auf einer Brücke stehen zwei Philosophen und schauen ins Wasser. „Schau, wie glücklich die Fische sind!“ „Woher willst du das wissen, du bist doch kein Fisch?“ „Woher willst du das wissen, du bist doch nicht ich?“

[26] Eine Behavioristin zur anderen: „Hat es mir gefallen?“

Behaviorismus über so lange Zeit (eingeschränkte Methoden, Verzicht auf viele wichtige Fragestellungen) nicht ganz außer Acht lassen, daß er eine große Herausforderung für ein neues Nachdenken über die tatsächlichen oder vermeintlichen Unterschiede zwischen Mensch und Tier oder zwischen Geist und Materie darstellte.

Auf der anderen Seite machen Umformulierungen wie von „Er war wütend und beleidigt, weil ich ihm eine Ohrfeige gab“ in „Nach einem Schlag von der und der Kraft auf seine linke Wange wurde die Haut stärker durchblutet, in den Augen war Tränenfluß zu beobachten, Blutdruck, Puls und Adrenalinspiegel stiegen an, er machte eine Körperdrehung zur senkrechten Achse um 180 Grad, hob das rechte Bein und senkte es wieder kräftig auf den Fußboden ...“ oder von „Sie ging los, um zur Feier dieses besonderen Ereignisses Kaviar und Champagner zu holen“ in „Nach neun Stunden ohne Nahrungszufuhr und den und den Aktivitäten zog sie einen Mantel an ...“ deutlich, daß der Behaviorismus viele Fragen, die wir über die menschliche Psyche haben, weder beantworten will noch kann. Als Argumente für den grundsätzlichen Einwand gegen den Behaviorismus, daß sich Erforschung und Erklärung mentaler Zustände und Prozesse nicht ohne Verlust durch die Beobachtung von Verhalten ersetzen lassen, kann man u. a. anführen:

- Es kann mentale Zustände geben, mit denen kein beobachtbares Verhalten korrespondiert (Erstaunen ohne Heben der Augenbrauen wegen einer Lähmung, bewußtes und aktives Verschweigen einer Information).
- Es kann umgekehrt beobachtbares Verhalten geben, das nicht mit den vermuteten mentalen Zuständen korrespondiert (z. B. bei perfekten Schauspielern).
- Das Verhalten hängt nicht nur von bestimmten Außenreizen ab, sondern es wird auch jeweils anders durch bestimmte mentale Zustände und Operationen veranlaßt.[27]
- Somit wird es unmöglich, angemessen zwischen *Verhalten* und (von Motiven, Wünschen, Überzeugungen, Absichten, Zwecken, kurz: Intentionen geleitetem) *Handeln* zu unterscheiden, und die Behavioristen würden letzteres zumindest *für sich selbst* in Anspruch nehmen, wenn sie ihre Ratten durch ein neues Labyrinth schicken.

Zur behavioristischen Lerntheorie ist in den letzten Jahrzehnten in der pädagogischen Diskussion so viel gesagt worden, daß hier auf weitere Einzelheiten verzichtet werden kann (und sie ist wohl gerade aus diesem Grunde den Schülerinnen und Schülern heute kaum bekannt). Dabei lohnt es sich – neben der Herausforderung an das Selbstverständnis des Menschen – auch ein wenig den aktuellen Zeugnissen dieser Theorie im Alltag und in der Schule nachzuspüren. Ein Thema kann dabei auch das Erkenntnis- und Verwertungsinteresse von Wissenschaft sein (wenn es hauptsächlich darauf ankommt, *daß* etwas bestimmtes gelernt wird oder *daß* ein Kaufhaus viel Umsatz hat, ist es möglicherweise auch legitim, zu erforschen, durch welche Hintergrundmusik es am besten erreicht wird), ein weiteres die Verläßlichkeit ihrer Zeugnisse (vgl. dazu die unterschiedlichen Versionen der traurigen

[27] „Suppose a behaviorist analyzes someone's wanting an ice cream cone as his having a set of behavioral dispositions such as the disposition to grasp an ice cream cone if one is ‚presented‘ to him. But someone who wants an ice cream cone will be disposed to grasp it only if he *knows* it is an ice cream cone (and not in general if he thinks it is a tube of axle grease being offered to him as a joke) and only if he does not *believe* that taking an ice cream cone would conflict with other *desires* of more importance to him (for example, the desire to avoid an obligation to return the favor).“ (Ned Block (1981), S. 11 f.)

Geschichte vom „kleinen Albert“ und der weißen Ratte bei Helbig (1979), S. 86, und Neel (1983), S. 150 f.).

Die Kritik am Behaviorismus floß auch ein in die Zwischenbilanz am Ende dieses Unterrichtsmoduls. Das Erkunden von ELIZA, die Sichtweise des Fragebogens und der Turing-Test *sind* behavioristisch. Das hilft uns dabei, über den genauen Sinn der Frage „Können Computer denken?“ nachzudenken und sie gegebenenfalls neu zu stellen, nicht aber dabei, sie befriedigend zu beantworten. Es wurde deutlich, daß ein Vergleich des „Outputs“ nicht genügt, sondern zur genaueren Klärung neben der Frage nach unseren Zwecken und Zielen bei der Konstruktion und dem Einsatz von Computern (vgl. den Text von Walter Schulz im folgenden Unterrichtselement) auch die nach einem Vergleich der „inneren“ Vorgänge bei Computern und Menschen gestellt und beantwortet werden muß (vgl. die Abschnitte 4 und 6 vom Teil 2).

5 Die Mehrdeutigkeit von „Denken“

Unterrichtsskizze

Der Text „Die Mehrdeutigkeit der Frage nach der Denkfähigkeit des Computers“ von Walter Schulz (1972, S. 222–226, etwas gekürzt) wurde von den Schülerinnen und Schülern unter der Fragestellung

1. Stellen Sie kurz die drei im Text genannten Ansätze zu der Frage, ob Computer denken können, dar. Welche Einwände könnte jeweils ein Vertreter dieser Position gegen die anderen erheben?
2. Erläutern Sie, inwiefern nach Schulz die Vertreter der genannten Ansätze „Denken“ in unterschiedlichen Bedeutungen verwenden. In welchem Verhältnis stehen bei beiden Bedeutungen jeweils Mensch und Computer?
3. Welcher der drei Ansätze entspricht am ehesten Ihrer eigenen Position?

schriftlich bearbeitet. Die Ergebnisse wurden anschließend verglichen, und es entstand eine lebhafte Diskussion über die Antworten zur dritten Frage. Eine Zwischenbilanz, welche Aspekte der Frage „Können Computer denken?“ geklärt und welche offen geblieben sind sowie ob und inwiefern sich die Fragestellung verändert hat, bildete den Abschluß dieses Unterrichtsmoduls.

Kommentar

Der Text von Walter Schulz über die Mehrdeutigkeit von „Denken“ stellt klar verschiedene Antworten zur Denkfähigkeit von Computern und damit verbundene Konzepte von „Denken“ gegenüber, er eignet sich daher gut für einen zusammenfassenden Abschluß dieser Unterrichtseinheit (auch als Klausur). Seine Unterscheidung in „Denken$_1$:= kraft Selbstbewußtsein auf sich, sein Handeln und dessen Ergebnisse reflektieren“ (wobei die Konstruktion und der Einsatz von Computern einen Teil der „dialektischen Bewegung der

Selbstvermittlung darstellt“) und „Denken$_2$:= Hervorbringen bestimmten, objektiv von außen an Hand der Leistung zu beobachtenden Verhaltens“ orientiert sich einerseits an einer bestimmten philosophischen Tradition, andererseits an einem behavioristischen (Geist als „black box“) bzw. an einem bloß auf Leistung angelegten, technischen Verständnis der KI („weiche KI“), und er bewegt sich damit durchaus im Rahmen des Turing-Testes (dort werden nicht Denkprozesse verglichen sondern nur ihre Ergebnisse). Offen muß an dieser Stelle bleiben, ob seine Argumentation auch gültig ist für neuere Entwicklungen wie den kognitivistischen „Physikalisches-Symbol-System-Ansatz“ (vgl. Abschnitt 1 vom Teil 2) oder der Konnektionismus. Diese Richtungen der KI bestehen darauf, daß Computer später einmal nicht nur Denken *simulieren*, sondern daß der dann bei ihnen ablaufende Prozeß Denken *ist* (weil z. B. auch unser Denken nichts anderes sei als regelgeleitete Symbolmanipulation). Ohne dieses Problem hier klären zu können, sollte diese dritte Definition „Denken$_3$:= das, was – *funktional* gesehen – das Gehirn macht (unabhängig davon, *ob* ein Gehirn es macht)“ als weitere Position angedeutet werden („harte KI“ und Funktionalismus, vgl. den Abschnitt 4 vom Teil 2).

Der Hinweis von W. Schulz, daß bei „Denken$_2$“ der Prozeß der maschinellen Verarbeitung gleichsam als Mittelstück von „Denken$_1$“ für sich gesetzt wird, hat eine interessante Parallele zum für den Informatikunterricht grundlegenden EVA-Prinzip (Eingabe, Verarbeitung, Ausgabe). Bei der schematischen Darstellung dieses Prinzips wird gerne weggelassen, daß es nur sinnvoll zu denken ist, wenn an den Ein- und Ausgabegeräten *Menschen* vorhanden sind (und auch die Verwendung desselben Verbs für die Ein- und Aus„gabe“ verwischt Unterschiede). Da das natürlich bei der Entwicklung von Computern und Programmen durchaus bedacht wird, wäre das angemessene Bild für das EVA-Prinzip ein Mensch mit einer „Denkblase“, in der Eingabe-, Verarbeitungs- und Ausgabegeräte samt der an ihnen arbeitenden Menschen zu sehen sind. Auf einer weiteren Reflexionsstufe habe nun ich das Bild von Menschen im Kopf, die dieses Bild im Kopf haben ...

Literaturhinweise

Allgemeine Hinweise

Nach der Pionierarbeit von Sherry TURKLE (1984) über den Einfluß des Computers auf das Selbstverständnis und das Denken über das Denken bei Kindern und in der KI-Forschung Tätigen wurden auch in Deutschland eine Reihe von empirischen Untersuchungen durchgeführt: Hermann ROSEMANN (1986), Jörg PFLÜGER & Robert SCHURZ (1987), Dieter SINHART-PALLIN (1990), Peter NOLLER & Gerd PAUL (1991) und Christel SCHACHTNER (1993). Die pädagogischen Perspektiven und Konsequenzen sind zentrales Thema bei Stefan M. GERGELY (1986), Peter FAULSTICH & Hannelore FAULSTICH-WIELAND (1988), im Studienbrief des DIFF (1989; einer der Autoren, Hans D. HINKELBEIN (1988), hat seine Gedanken zur Auswirkung der KI auf das Selbstverständnis des Menschen auch noch etwas ausführlicher dargestellt) und, aus anthroposophischer Sicht, bei Ernst SCHUBERTH (1990). Notwendige Schritte für eine Technikfolgenabschätzung im Bildungsbereich hat Ekkehard MARTENS (1988a) skizziert.

Der Verbindung von Informatik und Philosophie hat auch Rüdeger BAUMANN (1996) einen Abschnitt gewidmet. Henk GOORHUIS (1990) plädiert dafür, daß die Ansätze und Probleme der KI zum zentralen Gegenstand einer informations- und kommunikationstechnischen Allgemeinbildung werden. Die Herausforderungen der KI für die Didaktik der Philosophie erörtern Ekkehard MARTENS (1988b) und Klaus-Dieter SCHELLSCHMIDT (1988). Kommentierte Textsammlungen für den Philosophieunterricht mit vielen grundlegenden Texten zum Thema „Künstliche Intelligenz“ haben & Ekkehard MARTENS (1986) und Bruno HELLER (1991) herausgegeben. Über ihre Unterrichtserfahrungen zu diesem Thema berichten Friedhelm KAPTEINA (1987), Reinhard MENGES (1987) und Matthias TICHY (1996), viele Hinweise und Materialien dazu auch in der Handreichung des IPTS (1991). Renate SEIDLER & Lorenz HILTY (1994) haben das Thema KI als Rahmen und roten Faden für einen ausführlichen und detaillierten Einführungskurs in die Informatik (mit LISP) gewählt.

Zum Unterrichtsmodul 1

Andere Akzente zur Behandlung von **ELIZA** im Informatikunterricht setzen Jörg BELEITES (1989) — die Schülerinnen und Schüler entwickeln ein eigenes ELIZA-Programm (vgl. auch Renate SEIDLER & Lorenz HILTY (1994), S. 76–85) — und Gertrud EFFE-STUMPF, Gabriele GLÄSSING & Annette HABIGSBERG (1988) — hier wird in einem Computerkurs für Frauen der Begriff „einfühlsames Verstehen“ in den Mittelpunkt einer Beschäftigung mit ELIZA gestellt. Aufbau und Wirkung von ELIZA thematisieren auch Rüdeger BAUMANN (1995) und Wilmar STEUP (1995; das auf Diskette beigefügte Programm läßt sich auch in der Art des Turing-Teste einsetzen). Der theoretische Bezugsrahmen der von ihm entwickelten Gesprächstherapie läßt sich z. B. bei Carl R. ROGERS (1976) finden. Sherry TURKLE (1984, S. 41 ff.) beschreibt die beiden extremen Haltungen beim längeren Umgang mit ELIZA („austricksen“ bzw. „Illusion aufrechterhalten“), Joseph WEIZENBAUM selbst hat das Programm (1966) und seine Reaktion auf die Reaktionen (1977, 1984) ausführlich dargestellt. Pamela MCCORDUCK (1987, S. 237 ff., 287 ff.) zeichnet die Entstehungsgeschichte von ELIZA sowie den Streit zwischen WEIZENBAUM und COLBY nach

und versucht dabei, auch COLBYS Standpunkt zu stärken (kurz: wer hilft hat recht). Allen, denen ein Einsatz von Computern in der Psychotherapie als völlig abwegig erscheint, sei der Artikel von Christopher JOYCE (1989) empfohlen: in der Tat, wenn es wie z. B. im *Minnesota Multiphasic Personality Inventory* gilt, ein (diagnose- und therapierelevantes) Persönlichkeitsprofil dadurch zu erstellen, daß in einem standardisiertem Verfahren 566 als „richtig“ oder „falsch“ zu bewertende Aussagen ausgewertet werden, ist ein Computer eher besser (dort finden sich auch weitere Beispiele für „computerunterstützte Psychotherapie“). Arno BAMMÉ, Günter FEUERSTEIN, Renate GENTH, Eggert HOLLING, Renate KAHLE & Peter KEMPIN (1983, S. 34 f.) stellen dann im Rahmen ihrer Untersuchung über das Maschinenhafte im Menschen auch konsequent die Frage, ob nicht weniger die Therapeuten die Fähigkeiten solcher Programme überschätzen, sondern *wir* die Tätigkeit der Therapeuten, ob also nicht ein wesentlicher Teil der (insgesamt ohnehin nicht völlig nachprüfbaren) Therapieerfolge darauf beruht, daß die „Klienten“ durch Techniken zum Reden gebracht werden und sich dann „ungehemmt narzißtisch verausgaben“ können (und vielleicht liegt auch eine Stimme aus meinem Bekanntenkreis nicht so falsch, die vermutet, daß bei der Propagierung von therapeutischen Erstgesprächen am Computer durch den *Psychiater* COLBY eine nicht eingestandene Verachtung für die Tätigkeit von *Psychologen* wie ROGERS zum Vorschein kommt).

Eine umfassende und äußerst lesenswerte Biographie Alan TURINGS mit vielen erhellenden Hinweisen zur Problemgeschichte seines Werkes hat Andrew HODGES (1989) geschrieben. Vollständige deutsche Fassungen von Alan TURINGS Aufsatz *Computing Machinery and Intelligence* sind im *Kursbuch 8* (1967) und in anderer Übersetzung in der von Bernhard DOTZLER & Friedrich KITTLER herausgegebenen Auswahl seiner Schriften (1987b, dort auch Angaben zu Leben und Werk) erschienen. Für den Unterricht wählte ich die letztere Fassung aus diesem sehr empfehlenswerten, mit einer Diskette versehenen Band (das *Kursbuch 8* ist wegen einer Reihe weiterer Aufsätze zur Philosophie der Mathematik und Informatik auch sehr empfehlenswert). Gekürzte, ein wenig anders als im Material 3 geschnittene Fassungen finden sich bei Matthias TICHY & Ekkehard MARTENS (1986) und bei Douglas R. HOFSTADTER & Daniel C. DENNET (1986)[28]. In diesen beiden kommentierten Sammlungen wurde auch eine gekürzte bzw. vollständige Fassung von HOFSTADTERS *Ein Kaffeehaus-Gespräch über den Turing-Test* abgedruckt, in dem viele Aspekte des **Turing-Test**es reflektiert werden (Erstdruck in: *Scientific American 244*, 1981, deutsch: *Spektrum der Wissenschaft*, Juli 1981, Nachdruck auch in HOFSTADTER (1988) und in KURZWEIL (1993), S. 80–93). Ausführliche Analysen zur Adäquatheit des Imitationsspiels lieferten u. a. James H. MOOR (1976), der die These vertritt, daß man den Turing-Test weder als operationale Definition noch als notwendige oder hinreichende Bedingung, wohl aber als Quelle guter induktiver Evidenz maschineller Intelligenz verstehen kann, und Ned BLOCK (1981), der den Turing-Test (und ELIZA) zum Ausgangspunkt seiner Zurückweisung einer behavioristischen Auffassung von Intelligenz wählt; dagegen setzt er die These, daß von zwei Systemen mit völlig identischem Output bei einem beliebigen sprachlichen Input eins intelligent, das andere ohne jede Intelligenz sein kann. Daniel C. DENNETT (1993) hingegen hält wegen des nötigen „Weltwissens“ den Turing-Test in uneingeschränkter, allgemeiner Form für hinreichend aussagekräftig. Auch Oswald WIENER (1984, 1990) verteidigt ausdrücklich den Aussagewert des Turing-Testes und hält

[28] Leider gibt es zumindest in der mir vorliegenden zweiten Auflage einen interessanten Druckfehler: im Musterdialog (S. 60) wird statt des von TURING absichtlich falsch gewählten Additionsergebnisses 105 621 fälschlich das rechnerisch korrekte Ergebnis 105 721 als Antwort gegeben. Computerdenken?

die Prognose TURINGS auch für vereinbar mit den Ergebnissen einer (seiner Ansicht nach viel zu wenig genutzten) gründlichen Introspektion geistiger Prozesse. Der *Einwand von Lady Lovelace* ist Ausgangspunkt zu einer umfassenden Untersuchung des Phänomens Kreativität von Margaret BODEN (1992); ihre These: Konzepte und Ergebnisse aus der KI-Forschung können viel beitragen zur Klärung der Kreativität beim *Menschen*. Mit der „Paradoxie der Mechanisierung von Kreativität" beschäftigt sich auch HOFSTADTER (1988, S. 565 ff., vgl. auch S. 239 ff. und 1985 *passim*). *Der mathematische Einwand* (unter Hinweis auf GÖDELS Theorem) ist immer wieder vorgebracht — jüngst wieder von Roger PENROSE (1991) — und ebenso häufig bestritten worden; eine neuere Diskussion und Bewertung dieses Argumentes unternimmt Rosemarie RHEINWALD (1991). Eine sehr amüsante Umkehrung des Turing-Testes (Wie lange kann man einen ausgefuchsten Fachmann glauben machen, daß er mit einem intelligenten Programm und nicht mit Menschen kommuniziert?) schildert HOFSTADTER (1988, S. 551 ff., auch in KURZWEIL (1993), S. 94–101). Über die Ergebnisse eines (sehr eingeschränkten) realen Turing-Testes im Rahmen eines Wettbewerbes für Programme des Bostoner Computer Museums — sozusagen *The State of the Art* — berichtet *Der Spiegel* 47/1991 (S. 329 ff.), 19/1997 (S. 200 f.) und *c't* 7/1993 (S. 64 ff.)

Der Text zu der **Mehrdeutigkeit von „Denken"** wurde dem großartigen, viele Fragen unserer durch die Wissenschaften bestimmten Welt in praktischer Absicht reflektierenden Buch von Walter SCHULZ (1972, S. 222–226, etwas gekürzt)[29] entnommen. Im Umfeld dieser Seiten — ein Kapitel zu den philosophischen Fragen der Kybernetik und ihrem zentralen Begriff der Information — findet sich auch eine Auseinandersetzung mit der Position von Gotthard GÜNTHER (1963a, 1980a), der einen ähnlichen, aber im Detail doch anderen dialektischen Ansatz vertritt.

Eine systematische Darstellung (jeweils Basiseinheit, Arbeitsfeld, Methoden, Grundsätze, Verständnis bestimmter Schlüsselbegriffe) und Bewertung der vielfältigen psychologischen Theorien bis zur Mitte der sechziger Jahre bietet Ann F. NEEL (1983), dabei werden auch die verschiedenen Spielarten des **Behaviorismus** vorgestellt. Eine kurze Geschichte der Psychologie und die Wende vom Behaviorismus zur kognitiven Psychologie ist neben vielem anderen (es handelt sich um eine Darstellung der Entwicklung der Kognitionswissenschaft und der Beiträge von Philosophie, Psychologie, KI, Linguistik, Anthropologie und Neurowissenschaft dazu) im äußerst lesenswerten Buch von Howard GARDNER (1989) zu finden. Das Programm der kognitiven Psychologie ist eine am „Informationsverarbeitungsansatz" orientierte Klärung der mentalen Strukturen und Prozesse auf einer eigenständigen, mittleren Ebene zwischen der Physiologie und dem Verhalten (so wie eine höhere Programmiersprache zwischen Prozessorebene und Output steht). Weit mehr als nur eine Einführung in dieses Programm hat mit einer Fülle von Themen und Beispielen John R. ANDERSON (1988) vorgelegt.

[29] Eine ein wenig anders geschnittene Version haben auch M. TICHY & E. MARTENS (1986) in ihre Sammlung aufgenommen; Ausschnitte mit einer viel detaillierter als von mir aufbereiteten Fragestellung im Rahmen eines Einführungskurses in die philosophische Anthropologie auch bei Gerburg BÖHRS, Margrit GEISSLER & Walter HIRSCH (1992), S. 30–33, bzw. als Beitrag von Margit GEISSLER in IPTS (1991), S. 115–138.

Literaturverzeichnis

Anderson, J. R. (1988). *Kognitive Psychologie: eine Einführung.* Heidelberg: Spektrum der Wissenschaft (Original: Cognitive Psychologie and Its Implicationes, New York und Oxford 1985: Freeman)

Bammé, A., G. Feuerstein, R. Genth, E. Holling, R. Kahle & P. Kempin (1983). *Maschinen-Menschen, Mensch-Maschinen: Grundrisse einer sozialen Beziehung.* Reinbek bei Hamburg: Rowohlt

Baumann, R. (1995). Sprachverarbeitung als Mustervergleich oder: Wider die „Computerfreaks"! *LOG IN*, 3/95, S. 30–36, 5-6/95, S. 40–43

Baumann, R. (1996). *Didaktik der Informatik.* Stuttgart: Klett (2., vollst. neu bearb. Aufl.)

Beleites, J. (1989). *Eliza: Ein Mensch-Maschine-Dialog als Einstieg in die Informatik.* (Handreichung zum Informatikunterricht, hrsg. von der Freien und Hansestadt Hamburg)

Block, N. (1981). Psychologism and Behaviorism. *The Philosophical Review,* XC, S. 5–43

Boden, M. A. (1992). *Die Flügel des Geistes: Kreativität und künstliche Intelligenz.* München: Artemis & Winkler (Original: The Creative Mind. Myths & Mechanisms, London 1990: Weidenfeld and Nicolsen)

Böhrs, G., M. Geissler & W. Hirsch [Hrsg.] (1992). *Vom Menschen.* Heft 1: *Der Mensch ist eingelassen in seine Welt.* Stuttgart: Ernst Klett Schulbuchverlag (Reihe Philosophie — Ethik — Werte und Normen)

Dawkins, R. (1987). *Der blinde Uhrmacher: ein neues Plädoyer für den Darwinismus.* München: Kindler (Original: The Blind Watchmaker, 1986)

Dennett, D. C. (1993). Können Maschinen denken? In: R. Kurzweil, *KI: das Zeitalter der Künstlichen Intelligenz* (S. 48–61). München, Wien: Hanser

DIFF [Hrsg.] (1989). *Lehren und Lernen mit dem Computer. 2. Computer als Herausforderung an Pädagogik und Gesellschaft.* Tübingen: Deutsches Institut für Fernstudien an der Universität Tübingen

Effe-Stumpf, G., G. Gläßing & A. Habigsberg (1988). *Weibliche Wege zu Naturwissenschaft und Computer?* Bielefeld (Arbeitsmaterialien aus dem Bielefelder Oberstufen-Kolleg (AMBOS) Band 27)

Faulstich, P. & H. Faulstich-Wieland (1988). *Computer-Kultur: Erwartungen — Ängste — Handlungsspielräume.* München: Lexika

Gardner, H. (1989). *Dem Denken auf der Spur: der Weg der Kognitionswissenschaft.* Stuttgart: Klett-Cotta (Original: The Mind's New Science. A History of the Cognitive Revolution, New York 1985: Basic Books)

Gergely, S. M. (1986). *Wie der Computer den Menschen und das Lernen verändert: ein kritischer Ratgeber für Eltern, Lehrer und Schüler.* München, Zürich: Piper

Golecki, R. (1994). *Fächerübergreifender Unterricht auf der Sekundarstufe II — warum, wozu, woran, wie, wodurch?* Compuskript, Hamburg

Goorhuis, H. (1990). Warum gehört das Thema „Künstliche Intelligenz" in die Allgemeinbildung? In: Cyranek, G., H. J. Forneck & H. Goorhuis (Hrsg.), *Beiträge zur Didaktik der Informatik* (S. 103–125). Aarau, Frankfurt am Main, Salzburg: Diesterweg, Sauerländer

Günther, G. (1963a). *Das Bewußtsein der Maschinen.* Krefeld: Agis

Günther, G. (1980a). Schöpfung, Reflexion und Geschichte. In: ders., *Beiträge zur Grundlegung einer operationsfähigen Dialektik,* Band 3 (S. 14–56). Hamburg: Meiner

Heintz, B. (1993). *Die Herrschaft der Regel: zur Grundlagengeschichte des Computers.* Frankfurt/Main, New York: Campus

Heller, B. (1991). *Kybernetik und Computer: Texte zur philosophischen Auseinandersetzung mit neuen Technologien.* Stuttgart: Klett (Kletts Arbeitstexte Philosophie, mit Kommentarband)

Hentig, H. v. (1987). *Das allmähliche Verschwinden der Wirklichkeit.* München, Wien: Hanser (3., erweiterte Auflage)

Hentig, H. v. (1988). Das Ende des Gesprächs? In: Gauger, H.-M. & H. Heckmann (Hrsg.), *Wir sprechen anders: warum Computer nicht sprechen können* (S. 81–101). Frankfurt am Main: Fischer

Hentig, H. v. (1993). *Die Schule neu denken: eine Übung in praktischer Vernunft.* München, Wien: Hanser

Hinkelbein, H. D. (1988). *„Künstliche Intelligenz“ und das Selbstverständnis des Menschen.* Tübingen: Deutsches Institut für Fernstudien an der Universität Tübingen (Arbeitsberichte Medienforschung 2)

Hodges, A. (1989). *Alan Turing, Enigma.* Berlin: Kammerer & Unverzagt (Original: Alan Turing: The Enigma, London 1983: Burnett Books)

Hofstadter, D. R. (1985). *Gödel, Escher, Bach: ein Endloses Geflochtenes Band.* Stuttgart: Klett-Cotta (Original: Gödel, Escher, Bach: an Eternal Golden Braid. New York 1979: Basic Books)

Hofstadter, D. R. (1988). *Metamagicum: Fragen nach der Essenz von Geist und Struktur.* Stuttgart: Klett-Cotta (Original: Metamagical Themas: Questing for the Essence of Mind and Patterns. New York 1985: Basic Books)

Hofstadter, D. R. & D. C. Dennett [Hrsg.] (1986). *Einsicht ins Ich: Fantasien und Reflexionen über Selbst und Seele.* Stuttgart: Klett-Cotta (Zweite, unv. Auflage, Original: The Mind's I, New York 1981: Basic Books)

Huber, L. (1993). Nur allgemeine Studierfähigkeit? Oder doch Allgemeine Bildung? Zur Wiederaufnahme der Diskussion über „Hochschulreife“ und die Ziele der Oberstufe. In: *Abitur — Hochschulreife — Studierfähigkeit. Zur Grundlegung eines neuen Maturitätskataloges* (S. 53–72). Loccumer Protokolle 56/93

IPTS [Hrsg.] (1991). *Unterrichtsfach: Philosophie. Der Philosophieunterricht und die Computertechnologie. Hinweise und Materialien.* (Hrsg.: Landesinstitut Schleswig-Holstein für Praxis und Theorie der Schule (IPTS), 2575/90)

Joyce, C. (1989). Der Computer als Co-Therapeut. In: Psychologie heute (Hrsg.), *Wie uns der Computer verändert* (S. 113–127). Weinheim, Basel: Beltz

Kapteina, F. (1987). Computer-Denken. *Zeitschrift für Didaktik der Philosophie*, 4/1987, S. 224–230

Klafki, W. (1991). Grundzüge eines neuen Allgemeinbildungskonzepts. Im Zentrum: Epochaltypische Schlüsselprobleme. In: ders., *Neue Studien zur Bildungstheorie und Didaktik* (S. 43–81). Weinheim: Beltz (2. Auflage)

Kurzweil, R. (1993). *KI: das Zeitalter der Künstlichen Intelligenz.* München, Wien: Hanser (Original: The Age of Intelligent Machines, Massachusetts Institute of Technology, 1990, 1992)

Martens, E. (1988a). Computer als Modell menschlichen Denkens? Erfordernisse der Technikfolgenabschätzung im Bildungsbereich. In: Philipp, P. A., Daimler-Benz-Aktiengesellschaft, *Technikfolgenabschätzung und Technikbewertung: Konzeption, Anwendungsfälle, Perspektiven* (S. 61–71). Düsseldorf: VDI ((Report/Daimler-Benz-AG, 10, 9. Daimler-Benz-Seminar Berlin der Forschungsgruppe Berlin, 19./20. November 1987)

Martens, E. (1988b). Rehabilitierung der Angewandten Philosophie? Zum Beispiel Computer-Ethik. *Zeitschrift für Didaktik der Philosophie*, 4/1988, S. 204–210

McCorduck, P. (1987). *Denkmaschinen: Die Geschichte der künstlichen Intelligenz.* Haar: Markt u. Technik (Original: Machines who think, San Francisco 1979: Freeman)

Menges, R. (1987). „Können Computer denken?“ Erfahrungen aus einem Unterrichtsprojekt zur „informations- und kommunikationstechnologischen Bildung“ in einer Jahrgangsstufe 11. *Zeitschrift für Didaktik der Philosophie*, 4/1987, S. 215–224

Metzner, J. (1987). Der Computer — eine sozialpädagogische Herausforderung? In: J. Fritz (Hrsg.), *Computer in der Jugendarbeit* (S. 144–153). Mainz: Matthias-Grünewald-Verlag

Moor, J. H. (1976). An Analysis of the Turing Test. *Philosophical Studies*, 30, S. 249–257

Neel, A. F. (1983). *Handbuch der psychologischen Theorien.* Frankfurt a. M.: Fischer (Taschenbuchausgabe der dtsch. Ausgabe von 1974, Original: Theories of Psychology: A Handbook, Cambridge, MA 1969: Schenkman)

Noller, P. & G. Paul (1991). *Jugendliche Computerfans: Selbstbilder und Lebensentwürfe — Eine empirische Untersuchung.* Frankfurt/Main, New York: Campus (Studienreihe des Instituts für Sozialforschung Frankfurt am Main)

Papert, S. (1982). *Mindstorms: Kinder, Computer und neues Lernen.* Basel, Boston, Stuttgart: Birkhäuser (Original: Mindstorms. Children, Computer, and Powerful Ideas, New York 1980: Basic Books)

Penrose, R. (1991). *Computerdenken: Die Debatte um Künstliche Intelligenz, Bewußtsein und die Gesetze der Physik.* Heidelberg: Spektrum der Wissenschaft (Original: The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics, 1989, Oxford University Press)

Pflüger, J. & R. Schurz (1987). *Der maschinelle Charakter: sozialpsychologische Aspekte des Umgangs mit Computern.* Opladen: Westdeutscher Verlag

Rheinwald, R. (1991). Menschen, Maschinen und Gödels Theorem. *Erkenntnis*, 34, S. 1–21

Rogers, C. R. (1976). *Die klientenzentrierte Gesprächspsychotherapie.* München: Kindler (Taschenbuchausgabe. Original: Client-Centered Therapy, Boston 1951)

Rosemann, H. (1986). *Computer: Faszination und Ängste bei Kindern und Jugendlichen.* Frankfurt am Main: Fischer

Rosenfeldt, G. (1993). *Neuronale Netze – Eine Einführung für den Biologie- und Informatikunterricht.* Handreichung hrsg. von der Freien und Hansestadt Hamburg, Amt für Schule

Roszak, T. (1986). *Der Verlust des Denkens: über die Mythen des Computer-Zeitalters.* München: Droemer Knaur (Original: The Cult of Information, 1986)

Sand, S. (1986). *Künstliche Intelligenz: Geschichten über Menschen und denkende Maschinen.* München: Heyne

Schachtner, C. (1993). *Geistmaschine: Faszination und Provokation am Computer.* Frankfurt am Main: Suhrkamp

Schank, R. C. & P. G. Childers (1986). *Die Zukunft der Künstlichen Intelligenz: Chancen und Risiken.* Köln: DuMont (Original: The Cognitive Computer: on Laguage, Learning, and Artificial Intelligence. Reading, MA 1984: Addison Wesley)

Schellschmidt, K.-D. (1988). Von existierenden Menschen und laufenden Programmen: ein philosophischer Blick auf die „Künstliche Intelligenz“. *Philosophie. Beiträge zur Unterrichtspraxis*, Heft 22, S. 58–69

Schuberth, E. (1990). *Erziehung in einer Computergesellschaft: Datentechnik und die werdende Intelligenz des Menschen.* Stuttgart: Verlag Freies Geistesleben (Menschenkunde und Erziehung, 49, Schriften der Pädagogischen Forschungsstelle beim Bund der Freien Waldorfschulen)

Schulz, W. (1972). *Philosophie in der veränderten Welt.* Pfullingen: Neske

Seidler, R. & L. Hilty (1994). *Elemente der künstlichen Intelligenz: funktionale Programmierung und Problemlösen.* (Handreichung zum Informatikunterricht, hrsg. von der Freien und Hansestadt Hamburg, Amt für Schule, Referat S 13/12)

Sesink, W. (1993). *Menschliche und künstliche Intelligenz: der kleine Unterschied.* Stuttgart: Klett-Cotta

Sinhart-Pallin, D. (1990). *Die technik-zentrierte Persönlichkeit: Sozialisationseffekte mit Computern, eine empirische Studie.* Weinheim: Deutscher Studien Verlag

Steup, W. (1995). *Mensch und neue Maschinensysteme: Künstliche Intelligenz und Sprachverarbeitung in der politischen Bildung.* Schwalbach/Taunus: Wochenschau-Verlag

Steup, W. (1995 f.). „Dämonen“ in der Tabellenkalkulation. Simulationen zur Künstlichen Intelligenz – neuronale Konzepte am Beispiel mustererkennender Fenster-Perceptrons im Unterricht der Sekundarstufen I und II. *LOG IN,* 15 (1995), Heft 1, S. 62–72. Neuronale Konzepte der Künstlichen Intelligenz. Tabellenkalkulation als Hilfsmittel zur Reflexion des Mensch-Maschine-Verhältnisses. Teil 2: Laterale Inhibition und Hopfield-Netz in den Sekundarstufen I bzw. II. *LOG IN,* 16 (1996), Heft 1, S. 40–52.

Steup, W. (1996). „Ist der Mensch eine Maschine?“ *Computer und Unterricht,* H. 21, S. 28–34

Tichy, M. & E. Martens (1986). *Computer-Denken.* Hannover: Schroedel (Forum Philosophie)

Tichy, M. (1996). Wer hat Angst vor dem Computer? Fächerübergreifender Unterricht Philosophie – Informatik. *Zeitschrift für Didaktik der Philosophie und Ethik*, 2/1996, S. 115–123

Turing, A. M. (1967). Kann eine Maschine denken? *Kursbuch 8,* S. 106–138 (Original: Computing Machinery and Intelligence, Mind 59, 1950)

Turing, A. M. (1987b). Rechenmaschinen und Intelligenz. In: B. Dotzler & F. Kittler (Hrsg.), *Alan Turing: Intelligence Service* (S. 147–182). Berlin: Brinkmann & Bose (Original: Computing Machinery and Intelligence, Mind 59, 1950)

Turkle, S. (1984). *Die Wunschmaschine: Vom Entstehen der Computerkultur.* Reinbek: Rowohlt (Original: The Second Self: Computers and the Human Spirit, New York 1984: Simon and Schuster)

Turkle, S. (1993). Aufwachsen im Zeitalter intelligenter Maschinen: Rekonstruktion des Psychologischen und Neubetrachtungen des Menschlichen. In: R. Kurzweil, *KI: das Zeitalter der Künstlichen Intelligenz* (S. 68–73). München, Wien: Hanser

Volpert, W. (1985). *Zauberlehrlinge: die gefährliche Liebe zum Computer.* Weinheim und Basel: Beltz

Weizenbaum, J. (1966). ELIZA — A Computer Program For the Study of Natural Language Communication Between Man And Machine. *Communications of the Association for Computing Machinery*, 9, S. 36–45

Weizenbaum, J. (1977). *Die Macht der Computer und die Ohnmacht der Vernunft.* Frankfurt a. M.: Suhrkamp (Original: Computer Power and Human Reason. From Judgement to Calculation, 1976, Freeman a. Co.)

Wiener, O. (1984). Turings Test: Vom dialektischen zum binären Denken. *Kursbuch 75,* S. 12–37

Wiener, O. (1990). *Probleme der Künstlichen Intelligenz.* Berlin: Merve