

Was sind eigentlich intelligente Systeme?

Informationstechnologien und ihre Interpretation

Diesseits von science fiction und philosophischer Spekulation über künstliche Intelligenz sind komplexe und in mancher Hinsicht intelligente informationstechnische Systeme bereits heute Realität – eine Realität, die ein verstärktes Interesse der Wissenschafts- und Technikforschung verdient. Für die Beschäftigung mit solchen Systemen lässt sich von der philosophischen KI-Debatte – auch und gerade von ihren eigentümlichen Schwächen – etwas lernen.



Hajo Greif

studierte Philosophie, Soziologie und Kulturanthropologie an der J. W. Goethe Universität Frankfurt. 2000-2003 Doktorand am Graduiertenkolleg „Technisierung und Gesellschaft“, TU Darmstadt, dort Promotion in Philosophie 2004. Fellowships am Science Studies Unit, University of Edinburgh (2001) und am IAS-STIS (2003-2004). Seit Oktober 2005 wissenschaftlicher Mitarbeiter des IFZ, Aufbau des Forschungsbereichs „Soziale Aspekte der IKT“.

E-mail: greif@ifz.tugraz.at

Umstrittene Visionen

Auch wenn der frühe Optimismus der Pioniere im Feld der Erforschung künstlicher Intelligenz (KI) sich mittlerweile an den zahlreichen praktischen Problemen beim Versuch ihrer Realisierung gebrochen haben mag: Die Vorstellung der Möglichkeit informationstechnischer Systeme, deren Verhalten so komplex, flexibel, lernfähig, selbstgesteuert und im Idealfall auch bewusst ist, dass es an die Möglichkeiten menschlichen Denkens und Handelns heranreicht, ist nach wie vor sehr lebendig. Diesseits des Programms der sogenannten „starken KI“ oder gar der abenteuerlichen technologischen Visionen eines Ray Kurzweil (2002) existieren bereits komplexe IT-Systeme („Expertensysteme“), die in bestimmten Aufgabenbereichen menschliche kognitive Kapazitäten erreichen oder gar übertreffen. Umstritten bleibt, ob sie darum das Prädikat „intelligent“ verdienen.

Ein blinder Fleck der STS?

Auf den ersten Blick scheint die Frage künstlicher Intelligenz ein genuines Thema für die Science and Technology Studies (STS) zu sein – denn es geht schließlich um eine aktuelle, technologisch erzeugte, tiefgreifende Veränderung der modernen Gesellschaft. Doch gerade zu diesen Technologien verhält sich das STS-Feld recht eigentümlich, insofern recht selten reflektiert wird, was bestimmte Technologien als intelligent auszeichnet und wel-

chen Unterschied dies im Umgang mit ihnen machen könnte.

Dies ist um so verwunderlicher, als der mittlerweile am weitesten verbreitete STS-Ansatz auf ein Konzept nicht-menschlicher Akteure in technisch-wissenschaftlichen Netzwerken zurückgreift. Die Annahme liegt nahe, dass dieser Ansatz auf die Analyse von Interaktionen mit intelligenten Systemen zugeschnitten ist, vorausgesetzt, man macht eine Form von Intelligenz – als kreatives Problemlösungsvermögen – zur Voraussetzung für die Zuschreibung von Handlungsfähigkeit.

Auffälligerweise finden sich jedoch gerade in der Akteur-Netzwerk-Theorie (ANT) kaum Bezugnahmen auf komplexe informationstechnische Systeme. Entweder sind die Akteure, von denen in der ANT die Rede ist, letztlich ganz und gar menschlich (wie etwa bei Mackay et al. 2000) – oder sie sind von einer Art, die sich kaum als intelligent beschreiben lässt. Bei Autoren wie Bruno Latour ist es für die Zuschreibung von Handlungsfähigkeit vollkommen ausreichend, dass Dinge Widerständigkeits in der Welt produzieren, an denen sich die Handlungswege anderer Akteure brechen (vgl. Latour 1988). Darum finden sich in seinen Werken zwar Türschließer und Milzbrand-Erreger als nicht-menschliche Akteure, aber gerade eben nicht informationstechnische Systeme und ihr bisweilen interessant komplexes Verhalten.

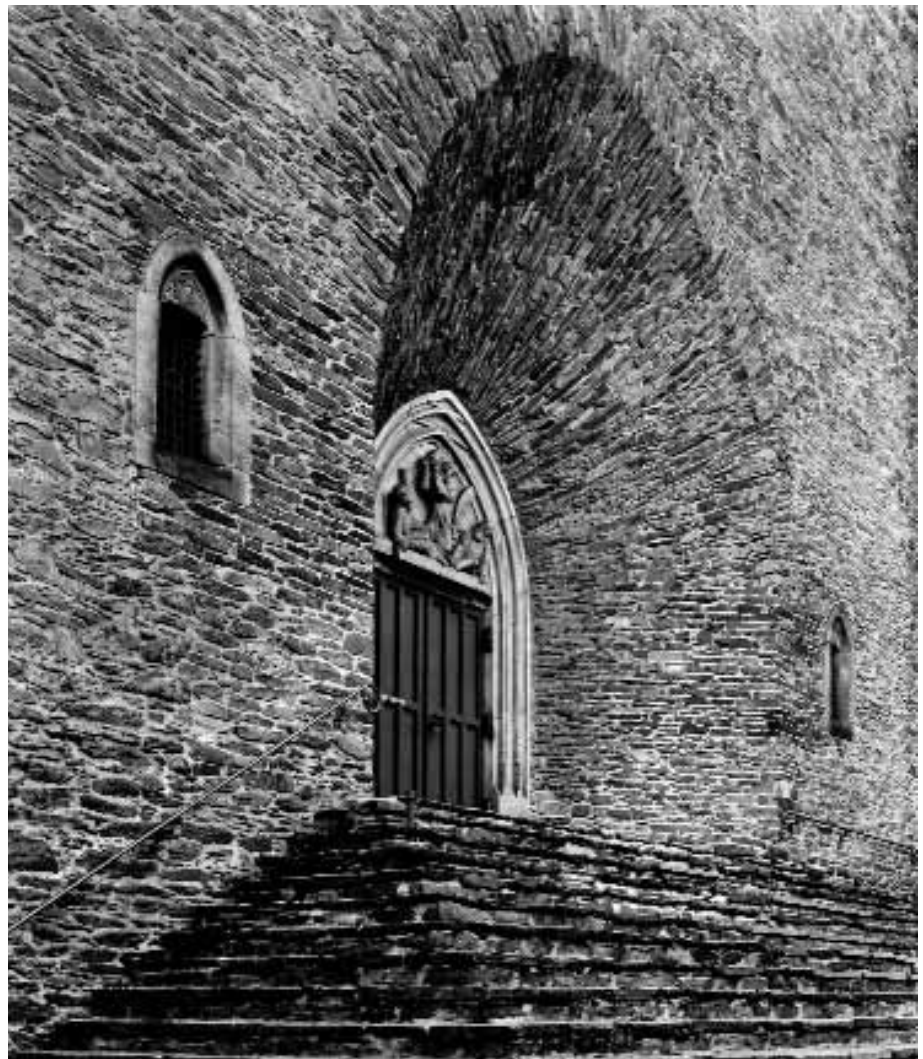
Innerhalb der KI-Forschung finden sich viel eher differenzierte, soziologisch informierte Ansätze zu Interaktionen zwischen Menschen und IT-Systemen, so etwa bei Lucy Suchman (1987), John Seely Brown und Paul Duguid (2000) und anderen ForscherInnen am Xerox PARC, Palo Alto (vgl. Collins 1994).

Die interessantesten expliziten Stellungen zur KI aus dem Feld der STS selbst sind sicherlich die von Harry M. Collins (1990). Seine Diskussion knüpft an Positionen in der philosophischen KI-Debatte an, die sich mit der Frage auseinandersetzen, ob, inwieweit und in welcher Weise infor-

mationstechnische Systeme tatsächlich intelligent sein können. Collins' Antwort ist, dass maschineller Intelligenz derart enge Grenzen gesetzt seien, dass eine starke KI unmöglich sei. Auch die leistungsfähigsten denkbaren Computer und Roboter unterscheiden sich nicht prinzipiell vom Buchdruck oder anderen „analogen“ Informationstechnologien und erfordern darum auch keine grundsätzlich andere Behandlung durch die STS als diese. Um diese Skepsis besser – und kritisch – zu verstehen, lohnt es sich, einen Blick auf die philosophische KI-Debatte und ihre Definitionen von Intelligenz zu werfen.

Die KI-Debatte

In dieser Debatte offenbart sich ein komplementäres Verhältnis zwischen dem jeweiligen Maß der Intelligenz: Die Vertreter der starken KI gehen davon aus, dass ein intelligentes System sich nach einem Modell des menschlichen Geistes konstruieren lässt, in dem dieser selbst als informationsverarbeitendes System von der Art des Computers verstanden wird: ein Schaltkreis, der mit Daten gefüttert wird und über Programme verfügt, um aus diesen einen Output zu berechnen, der selbst wiederum in die weitere Programmierung eingehen und zur eigenständigen Lösung neuartiger Probleme beitragen kann. Je mehr Daten das System bekomme und je besser es programmiert sei, das heißt über je feiner gegliederte Subroutinen unter je wirkungsvolleren Kontrollen es verfüge, desto intelligenter sei es. Die Zuschreibung von Intelligenz und Handlungsfähigkeit sei allein eine Frage von Graden funktionaler Komplexität: Sind die innere Organisation und das Verhalten einer Maschine so komplex, dass sie von einem Beobachter de facto nicht mehr in den Begriffen seiner physikalischen Struktur und seiner Funktionen zu interpretieren sind, so Daniel Dennetts berühmtes Argument (1971), sollte er ihr gegenüber die „intentionale Einstellung“ annehmen. Im Umkehrschluss erscheinen Menschen schlicht und einfach als besonders raffiniert programmierte Computer – und ihr Bewusstsein als ein hochdifferenziertes Kontrollprogramm, das im Prinzip physikalisch und funktional entschlüsselt werden kann. Als eines erscheinen Menschen und Computer genau darum jedoch nicht: als Dinge und Wesen in einer Gesellschaft, in die sie eingebettet sind und in der sie interagieren. Demgegenüber machen die Kritiker der starken KI geltend, dass deren Proponen-



ten das Pferd von vornherein falsch herum aufzäumen: Die Eigenschaften des menschlichen Geistes seien gerade nicht als von der Art des Computers zu verstehen, so dass die maschinelle Imitation des menschlichen Geistes auf diesem Wege gar nicht gelingen könne. Vielmehr wäre eine starke KI nur dann möglich, wenn es gelänge, Maschinen zu bauen, welche die besonderen Eigenschaften des menschlichen Geistes nachempfinden: flexibel und kreativ – und nicht zuletzt in Interaktion mit anderen Individuen – auf neue Situationen zu reagieren. Dies sei etwas qualitativ anderes als das formale Operieren an Symbolen, und nicht nur einfach mehr davon. Während Maschinen nur in der Lage seien, Daten in fixen Routinen zu prozessieren, sei das Merkmal echter Intelligenz, eine Intuition für die jeweils wichtigen Ähnlichkeiten und Unterschiede zwischen verschiedenen Situationen zu entwickeln. In unterschiedlichen Nuancen wird diese Position von Philosophen wie Hubert Dreyfus (1992) und John Searle (1980) vertreten. Collins (1990) fügt den weiteren

Einwand hinzu, dass Maschinen zuallererst eine lebensweltliche Situiertheit und somit die Einübung in implizite, nicht formal kodifizierbare Praktiken fehlten, die für die Entwicklung menschlicher Intelligenz unabdingbar seien.

Sowohl die VertreterInnen der starken KI als auch ihre Kritiker gehen von der Voraussetzung aus, dass eine Ähnlichkeit zwischen menschlicher und maschineller Beschaffenheit ein Kriterium dafür sei, Maschinen Intelligenz zuzuschreiben. Intelligenz wird stillschweigend mit menschlicher Intelligenz gleichgesetzt. Kontrovers ist allein die Frage, in welche Richtung man diese Ähnlichkeit denken soll. So oder so wird einem Wesen oder Ding nur dann eine Begegnung auf Augenhöhe gestattet, wenn es sich so verhält, wie Menschen es aus menschlicher Gesellschaft voneinander gewohnt sind. Dies macht zugleich ein Leitmotiv der Entwicklung künstlicher Intelligenz verständlich: Maschinen können unseren Zwecken genau dann auf wirklich intelligente Weise dienen, wenn es uns gelingt, sie so zu programmieren, dass wir ihr

Verhalten auf eine Weise verstehen, die es uns erlaubt zu sagen, dass die Maschinen uns, das heißt die von uns vorgesehenen Zwecke, verstehen. Der bislang größte Stolperstein in der KI-Entwicklung ist, dass experimentelle Systeme zwar eigenständig Muster erkennen können, welche verschiedene, für sie neue Situationen verbinden – aber meist greifen sie aus menschlicher Perspektive unerwartete, unwesentliche Gemeinsamkeiten heraus (Dreyfus 1992). Ich denke, dass die KI-Debatte insgesamt unter zwei Schwächen leidet:

■ Erstens ist die Angemessenheit der Beschreibung von Intelligenz keine Frage, die trennscharf beantwortet werden könnte oder sollte. Es geht nicht darum, und es ist praktisch auch gar nicht möglich, die inneren Zustände der Maschine zu entschlüsseln und objektiv festzustellen, ob ein Bewusstsein vorhanden ist, das als Kriterium der Intelligenz hinreichend ist. Dies ist auch nicht die Art und Weise, wie Menschen einander interpretieren. Was zählt, ist das beobachtbare Verhalten und dessen Erklärung anhand von Gründen, welche die Interpreten dem Urheber des Verhaltens zuschreiben – und zwar in einer geteilten sozialen Praxis.

■ Zweitens ist es, aufgrund der ungelösten Frage, wie der menschliche Geist funktioniert, keineswegs verwunderlich, dass, im Ergebnis informationstechnischer Designs, das Verhalten der Maschinen das hier vorausgesetzte Ähnlichkeitskriterium immer wieder verfehlen wird. Dies impliziert jedoch nicht bereits, dass ihnen darum eine Form von Intelligenz abzusprechen wäre. Sie mögen schlicht über andere Ziele und Strukturen verfügen als von den Konstrukteuren beabsichtigt. Es gilt, Wege zu suchen, wie diese eigentümlichen Ziele und Strukturen in die Kontexte sozialen Handelns einzuholen wären.

Intelligenz und Unberechenbarkeit

Eines der wesentlichen konkreten Probleme im Umgang mit real existierenden Expertensystemen besteht darin, dass ihr Verhalten zwar komplex und selbstgesteuert ist, aber Mustern folgt, die menschliche Beobachter nicht verstehen. Autonome Logistik-Systeme sind hierfür ein gutes Beispiel (vgl. Cramer 2005): Kein Mensch würde die Wege der von solchen Systemen abgelegten Dinge antizipieren können – nicht einmal wenn er die Programmierung

kennt –, während die Maschinen die Waren schneller und effizienter lagern und disponieren als ein manuell gesteuertes Logistik-System dies zu tun in der Lage wäre. Es gibt sogar gute Gründe dafür, Menschen von den automatisierten Abläufen fernzuhalten, da sie diese durch ihre Intervention nicht nur stören, sondern unter Umständen auch selbst in Gefahr geraten würden. Eingriffe sind allein auf der Ebene der Programmierung möglich. Dies sind, wenn man so will, intelligente Systeme ohne das menschliche Antlitz, an dem wir ihre Ziele und Überzeugungen ablesen könnten. Gerade das Element der Unberechenbarkeit ihres Verhaltens jedoch ist ein Motiv dafür, sie als intelligent zu interpretieren: Es ist erkennbar strukturiert, aber de facto uneinholbar eigensinnig.

In der Tat gibt es neben dem in der KI-Debatte verhandelten Kriterium der Ähnlichkeit noch ein anderes, ganz gegensätzliches, aber ebenso gängiges Kriterium für die Zuschreibung von Intelligenz und Handlungsfähigkeit: Der Unterschied zwischen beliebig komplexem, aber berechenbarem Verhalten und echtem Handeln wird oft genau dort gemacht, wo ein Verhalten derart unvorhersehbar ist, dass sich nur in der Rückschau Gründe für es benennen lassen. Dies ist ein wesentliches Element der Interpretation menschlichen Handelns – es zeichnet spontane, freie Entscheidungen gegenüber Routineverhalten aus. Es ist aber auch das Motiv hinter dem Versuch, die unberechenbare, fremdartige Eigensinnigkeit von nicht-menschlichen Dingen in den Griff zu bekommen, indem man sie auf Begriffe bringt, die der vertrauten Sphäre zwischenmenschlichen Handelns entlehnt sind.

Diese Strategie, die, in anderer Akzentuierung, auch Dennetts Programm der intentionalen Einstellung zugrunde liegt, findet sich in der Alltagspraxis in vielfältiger Weise (machen mein Computer und das Wetter nicht oft, was sie wollen?). Meist lässt sie sich als metaphorische Rede einklammern. Sie kann sich aber in der Analyse technischen Handelns von echtem Erkenntniswert erweisen, wenn man die Anwendung dieser Strategie durch IngenieurInnen und NutzerInnen im Umgang mit komplexen IT-Systemen ernstnimmt und systematisch reflektiert. Es wäre etwa sinnvoll zu beobachten, welche Typen von Charakterisierungen des Verhaltens der betreffenden Systeme von deren NutzerInnen besonders häufig verwendet werden, wenn jene ein unerwartetes, aber offenbar planmäßiges Verhalten zeigen,

und welche Effekte die entsprechenden Charakterisierungen auf den weiteren Umgang mit den Systemen haben. Der interessante Fall stellte sich dann ein, wenn sich ein Unterschied aufweisen ließe zwischen unverbindlichen metaphorischen Zuschreibungen von Intelligenz und Handlungsfähigkeit an technische Gegenstände, die keine weiteren Folgen für die Anschluss-handlungen der Beteiligten haben, und solchen, auf deren Basis sich stabile Wechselwirkungen zwischen Systemen und NutzerInnen entwickeln. Spätestens wenn in solchen Situationen eine Anpassungsleistung des Verhaltens von beiden beteiligten Seiten her stattfindet, kann ein System als intelligent gelten: Es ist in diesem Fall in der Lage, die menschlichen Akteure zu interpretieren. In diesem Sinne ist künstliche Intelligenz nicht etwas, das sich vom philosophischen Lehnstuhl aus konzipieren oder widerlegen oder am (digitalen) Reißbrett erfinden ließe, sondern etwas, das sich im praktischen Umgang mit Technologien erweist. Darum ist künstliche Intelligenz ein genuines Betätigungsfeld für die STS.

Literatur

- Brown, J. S., P. Duguid: *The Social Life of Information*. Boston: Harvard Business School Press 2000.
- Collins, H. M.: *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge: MIT Press 1990.
- Collins, H. M.: *Science Studies and Machine Intelligence*. In: S. Jasanoff et al. (eds.): *Handbook of Science and Technology Studies*. Thousand Oaks: Sage 1994, S. 286-301.
- Cramer, S.: *Hybridisierung und Risiken in soziotechnischen Systemen*. In: *SOZIALE TECHNIK* 3/2005, S. 12-14.
- Dennett, D.: *Intentional Systems*. In: *The Journal of Philosophy* 68/1971, S. 87-106.
- Dreyfus, H.: *What Computers Still Can't Do*. Cambridge: MIT Press 1992.
- Kurzweil, R.: *Fine Living in Virtual Reality*. In: P. J. Denning (Ed.): *The Invisible Future: The seamless integration of technology into everyday life*. New York: McGraw-Hill 2002, S. 193-215.
- Latour, B.: *The Pasteurization of France*. Cambridge: Harvard University Press 1988.
- Mackay, H. et al.: *Reconfiguring the User: Using Rapid Application Development*. In: *Social Studies of Science* 30/2000, S. 737-757.
- Searle, J. R.: *Minds, Brains, and Programs*. In: *The Behavioral and Brain Sciences* 3/1980, S. 417-457.
- Suchman, L.: *Plans and Situated Actions*. Cambridge: Cambridge University Press 1987. ■