

Zukunftsvisionen der KI: Wünschbare und realistische Wirkungen auf die Lebenswelt*

Prof. Dr. I. Wachsmuth

Mit den hier vorgetragenen Ausführungen berichte ich über die Tätigkeit derjenigen Arbeitsgruppe in unserem Diskursprojekt, die sich mit möglichen Auswirkungen zukünftiger KI-Systeme auf die Gesellschaft auseinandersetzt.¹ Ging es in einem Teil des Diskurses vor allem um die gegenwärtigen Entwicklungen der KI und um ihre kritische Reflexion, soll der Ausgangspunkt für diesen Beitrag die Frage sein: Was ist eigentlich für zukünftige Entwicklungen wünschbar, und sind nicht technische Innovationen nicht nur vom Standpunkt des Denkbaren, sondern auch des Wünschbaren bzw. des Gewünschten zu diskutieren? Die Vorhersage zukünftiger Entwicklungen und ihrer Auswirkungen ist schwierig, wenn nicht unmöglich. Wenn man sich aber überlegt, aus welchen Richtungen kommen denn Wünsche, was sind Leitvorstellungen und Motivationen der Beteiligten, dann kann man vielleicht eher erkennen, wohin die Entwicklungen laufen könnten.

Eine weitere Bemerkung will ich vorausschicken: Wir diskutieren hier zu einem Teil fiktive Szenarien, von denen man bei realistischer Einschätzung sagen möchte, "so kommt das alles sowieso nicht". Allerdings erkennt man doch in dem, was in den Medien über Künstliche Intelligenz geschrieben wird, daß ein großer Bedarf besteht, sich mit solchen Zukunftsvisionen – die ja in vielen Fällen mit bereits begonnenen oder geplanten Forschungs- und Entwicklungsarbeiten entstehen – auseinanderzusetzen. Selbst wenn wir also von einem Insiderstandpunkt zu erkennen glauben, daß manche der großen Entwürfe nicht oder nur mit großen Abstrichen erfolgreich sein werden, erscheint es mir sinnvoll zu fragen, wie wäre denn eigentlich eine Gesellschaft mit "intelligenten" Maschinen, um sich anhand der Implikationen des Denkbaren mit dem Wünschbaren zu befassen. Mein wiederholt vertretener Standpunkt ist dabei, daß Künstliche Intelligenz eine besondere Herausforderung für die Zukunftsgestaltung darstellt und daß "Zukunft" keine passive Angelegenheit sein sollte, in die man sich treiben

* Transkript des Vortrags.

¹ G. Görz, A. Kremer, H. Röpke, P. Schreiber, G. Strube, I. Wachsmuth, M. Wilker.



läßt, sondern daß die informationstechnische Zukunft mit KI aktiv zu gestalten ist. Denn die Richtungen, in die sich die Zukunft durch Künstliche Intelligenz entwickeln könnte, sind nicht völlig zufällig, sondern sie werden geleitet von mehr oder weniger expliziten Spekulationen der an der Technikentwicklung Beteiligten und ihrer Auftraggeber in Politik – auch Militär – und Wirtschaft.

Ich möchte nun zunächst einige Statements aus unserem Zukunftsdiskurs aufgreifen. Zuerst einmal hatten wir Schwierigkeiten, Künstliche Intelligenz von entwickelter Informationstechnik abzugrenzen. Und dies ist vielleicht auch schwer zu trennen, denn Techniken, die im Gebiet der KI entstanden sind, sind in vielen Fällen Ausgangspunkt informationstechnischer Innovationen gewesen (beispielsweise objektorientierte Programmiermethoden, die u.a. eine Basis heutiger Benutzungsoberflächen bilden). Wegen dieser grundsätzlichen Schwierigkeiten haben wir für unsere gegenwärtigen Überlegungen die folgende Arbeitshypothese an den Ausgang gestellt: Unter "entwickelter Künstlicher Intelligenz" wollen wir Systeme verstehen, die über kognitive Komponenten verfügen, die also beispielsweise Sensoren für optische und akustische Wahrnehmung haben, die solche Sensordaten verarbeiten können, die zu Schlußfolgerungen fähig sind, die heuristische Problemlösestrategien entwickeln und die derartige Fähigkeiten insgesamt in einem System integrieren. Wir beziehen uns also auf das, was man heute mit dem Stichwort "autonome Systeme" belegt, Systeme, die – mit Abstrichen – ähnlich wie der Mensch mit einer Außenwelt kommunizieren und interagieren.

Mit diesem Ausgangspunkt haben wir dann gestaffelte Entwicklungsstadien betrachtet – fiktive Stadien mit allen Vorbehalten über ihr tatsächliches Eintreten – und uns gefragt, angenommen, es käme so, was wären die Implikationen? In unserem Diskurs wollen wir also keine Aussagen darüber machen, was kommt, sondern die Frage diskutieren, "was wäre wenn?".

- Für das erste Stadium haben wir angenommen, daß Maschinen, die von Menschen entsprechend programmiert worden sind, intellektuelle Routinetätigkeiten übernehmen können. Dazu gehören einerseits klassische Expertensysteme, wie sie beispielsweise in der technischen Diagnostik bereits eingesetzt werden, des weiteren etwa maschinelle Systeme, die umfangreiche Meßdaten intelligent auswerten, indem sie beispielsweise Regularitäten abstrahieren; auch hierfür hat sich der Einsatz gegenwärtiger Techniken bereits als tragfähig und nützlich erwiesen.
- Für das zweite Stadium – für uns denkbar in den nächsten dreißig Jahren – galt folgender Gedanke: Wenn gegenwärtig projizierte Forschungsarbeiten erfolgreich sind, was wäre

dann als neue Qualität zu erwarten? Hierfür haben wir angenommen, daß Maschinen in größerem Umfang selbständig aus Beobachtungen, die sie vermöge ihrer Wahrnehmungskomponenten in ihrem Gegenstandsbereich anstellen, Wissen ansammeln; sie können auch unscharfes Wissen verwerten und eigene Fehler diagnostizieren. Dies ist sicherlich heute kaum auch nur ansatzweise möglich; es ist aber in Projektplänen derartiges vorgesehen.

- Für das dritte Stadium – über dreißig Jahre hinaus extrapoliert – haben wir spekulativere Zukunftsvisionen einbezogen, die annehmen, daß Maschinen nur noch teilweise determiniert sind, daß sie Situationen, mit denen sie konfrontiert sind, absichtsvoll verbessern und umgestalten können, indem sie aus sich selbst als Reaktion auf die Einwirkung der Umwelt kreatives Verhalten zeigen (etwa durch Strukturbildung und Selbstorganisation), daß sie Spontaneität, Zielstrebigkeit und Sensibilität besitzen können.

Ich betone noch einmal, wir haben diese hypothetischen Stadien einer entwickelten KI – die durchaus abgeleitet sind aus gegenwärtig diskutierten Zukunftsvorstellungen – als ein fiktives Szenario angenommen, und wir haben die damit verbundenen Entwicklungen nicht in Erwartung ihres wahrscheinlichen Eintretens als vielmehr hinsichtlich ihrer Implikationen und ihrer Wünschbarkeit betrachtet. Auch wenn die Möglichkeit der im dritten Stadium angesprochenen Entwicklungen mit Skepsis betrachtet wurde, darf nicht übersehen werden, daß derartige Zukunftsvisionen durchaus Einfluß auf Entscheidungsträger, etwa hinsichtlich der Finanzierung von Forschungs- und Entwicklungsprogrammen nehmen können. Der Punkt, um den es mir hier geht und der unsere spekulativen Überlegungen legitimiert, ist der folgende: Wenn aufgrund von projektierten Forschungs- und Entwicklungsarbeiten in breitem Umfang auf entsprechende Ziele hin gearbeitet wird, entstehen verwertete Teilergebnisse, und dann ist die Zukunft ein Stück weit anders, als wäre sie nicht von einer solchen Vision "gezogen" worden. Beispiele wie SDI haben gezeigt, daß bestimmte Leitvorstellungen kaum zurücknehmbare Auswirkungen haben, auch wenn die Visionen nicht im vollen Umfang erreichbar sind. Ihr Einfluß auf mögliche Zukunftsentwicklungen darf also nicht unterschätzt werden.

Die Fragen, mit denen wir uns in diesem Szenario auseinandergesetzt haben, haben sich zunächst einmal vornehmlich auf mögliche Veränderungen der Arbeitswelt bezogen; also etwa: Wird eine umfangreiche Neubewertung gegenwärtiger Qualifikationen eintreten? Welche Arbeit bleibt dem Menschen vorbehalten und wie wird sie dann bewertet? Verfällt beispielsweise ein Teil der Qualifikation des Ingenieurs? Werden sich grundsätzlich neue Strukturen in der Arbeitswelt entwickeln? Bilden wir noch zeitgemäß aus?

Aus unserem vorgelegten Bericht² möchte ich dazu die folgenden Überlegungen herausgreifen: Die technologische Weiterentwicklung kommt nicht von heute auf morgen über uns, pointiert ausgedrückt wird sie nicht über eine mehr oder weniger wehrlose Sozialgemeinschaft quasi "ausgegossen", sondern sie findet auf dem Umweg über die oft mühsame und langwierige Produktentwicklung Eingang in gesellschaftliche Wirkungszusammenhänge – nach Prinzipien der Markterschließung, Akzeptanz und der allmählichen Anpassung von Arbeitsorganisation und Arbeitsteilung. Soziale Systeme haben seit der Industriellen Revolution weitgehende Aufnahmefähigkeit für technologischen Fortschritt bewiesen. In vielen Fällen werden KI-Produkte als verbesserte Substitute bereits in Gebrauch befindlicher Artefakte in Erscheinung treten, wobei die Tatsache, daß ein Produkt nach seinem *Gebrauchswert* und kaum nach der ermöglichenden Technik bewertet wird, es fraglich erscheinen läßt, ob man darin verwandte KI-Techniken überhaupt als solche wahrnehmen wird.

Im weiteren Verlauf des Diskurses ist dann der Eindruck gewachsen, daß sich durch ein Vordringen von KI-Technologie in den Freizeitbereich ungleich umfangreichere Implikationen für die Alltags- und Lebenswelt ergeben könnten, und das soll im folgenden das Thema meiner Ausführungen sein. Um etwas konkreter auf das mögliche Vordringen von KI-Technologie in die Lebenswelt einzugehen, möchte ich einige Passagen aus einer Electronic Mail-Newsgroup³, die sich kürzlich mit diesem Thema befaßt hat, präsentieren. Dort ist unter der Frage "*AI in the year 2025*" – "Künstliche Intelligenz im Jahre 2025" – eine breitgestreute Umfrage überwiegend auf dem amerikanischen Kontinent gelaufen, die von ähnlichen Gedanken wie unser Vorgehen getragen war: "Was stellt Ihr Euch denn vor, das in dreißig Jahren ist?".

Von den zusammengestellten 35 Antworten, die auch Wünsche der Menschen erkennen lassen, will ich nicht alle und manche nur sehr kurz erwähnen. Da ist die Rede von Haushaltsrobotern, die Routine-Reinigungsarbeiten übernehmen, von städtischen Robotern, die den Müll einsammeln (doch die Frage, wohin damit, ist damit immer noch nicht gelöst), von selbstgesteuerten Automobilen und so banalen Wünschen wie "*no longer have to memorize phone numbers*". Mehrfach taucht das Thema "user interfaces" auf, daß also mit Hilfe der Künstlichen Intelligenz menschengerechte Benutzungsschnittstellen zu Computern geschaffen werden, daß man auch auf andere Weise als nur auf der Ebene von Fakten mit dem Computer kommunizieren kann, bis dahin, daß "*AI technologies will also be able to recognize and generate the emotional and stylistic content of such communication.*" Des weiteren wird erwähnt: "*voice recognition systems will be very good, widely implemented and will no longer be*

² siehe VDI Report 17 (1992), Abschnitt III (S. 138ff)

³ AI In the Year 2025 – Responses – comp.ai #6007

considered AI". Das haben wir schon öfter gehört, wenn etwas geklappt hat, dann ist es nicht mehr KI.

Andere Aussagen scheinen mir von größerer Tragweite zu sein: *"In the year 2025, the computer, the telephone network and television will have completely merged" [...] "the computer will become virtually (!) indistinguishable from our environment [...] And the principal key to this evolution in control over our environment is: AI!"* Es scheint bereits vom heutigen Standpunkt wahrscheinlich, daß gar nicht mehr der Computer als eine Hauptentwicklung dieses Jahrhunderts gewertet werden wird, sondern die mit dem Computer und der Kommunikationstechnologie verbundenen Netzwerke und die dadurch bedingten drastischen Einwirkungen auf gesellschaftliche Zusammenhänge. Von einem Computer, zu dem man sich zwecks Erledigung eines Auftrages begibt und wieder wegbegibt, sind wohl geringere Einflüsse auf die Lebenswelt zu erwarten, als von der medialen Vernetzung, die auch mit sich bringen wird, daß in weitem Umfang größere gesellschaftliche Gruppen weltweit miteinander in Kontakt und dadurch in neue, medial vermittelte Kommunikationsformen eintreten.

Ebenfalls bedeutsam erscheint mir die folgende Aussage: *"By the year 2025 decisionmakers of large corporations will trust their expert system over their human advisors"* – man scheint geneigt, eine Entwicklung zu akzeptieren oder mindestens für denkbar zu halten, bei der man sich mehr auf das umfangreiche Gespeicherte verläßt als auf aktuelle menschliche Perspektive und Einschätzung. Noch drastischer erscheint folgender Beitrag: *"Tapping the resources of the mind through artificial stimulation and embedded neuro-devices will be state of practice in the year 2025. Commercial offerings will include chemical, mechanical (micro machines), neural and biological approaches to enhancing decision making within the human brain – and not through external replication of the brain's processes as today"*. Möglicherweise ist diese Aussage ironisch gemeint, aber sie ist jedenfalls gegenwärtig im Denken einer Person, die sich über technologische Zukunft Gedanken macht; von einer solchen Vision ließen sich ganz andere Forschungsziele ableiten als von einer Sicht der technischen Hilfsmittel des intellektuellen Handelns, die außerhalb des menschlichen Körpers eingesetzt werden.

Jim Hendler, University of Maryland, der diese Umfrage gestartet hat, hat dann ein Statement von Freeman Dyson dazugesetzt: *"There is no illusion more dangerous than the belief that the progress of science is predictable"*. Dies paßt gut zu unserem Gedanken, daß wir keine Vorhersagen machen wollen, sondern stattdessen fiktive Szenarien diskutieren, um daran zu erörtern, was man für wünschenswert halten mag. Um zurückzukommen auf unser Thema, der möglichen Wirkungen von KI auf die Lebenswelt, sei noch eine weitere Aussage zitiert: *"The*

best AI will continue to be where the cash flow is greatest, in video-game entertainment where artificial lifeforms populate artificial reality."

Wir haben uns beispielsweise gefragt, ob neuartige Konsumgüter entstehen könnten, die die Lebenswelt drastisch verändern. Werden sie möglicherweise dazu führen, daß künstliche Welten neben einer realen Lebenswelt entstehen? Hier möchte ich nun auf unsere Überlegungen zur Verbindung von KI-Techniken mit der sog. "Virtual Reality"-Technik (auch "Artificial Reality" oder "Cyberspace") zu sprechen kommen.

Einige von Ihnen haben die Gelegenheit wahrgenommen, sich im Vorfeld des Symposiums bei der GMD erste Entwicklungsarbeiten zur technischen Nutzung von Virtual Reality demonstrieren zu lassen. Diese Technik ist ebenfalls von großer Attraktivität für die Entertainment-Industrie. Sie zielt auf eine möglichst unmittelbare Verbindung des Menschen mit seinen Aktivitäten und sensorischen Fähigkeiten mit der synthetischen Welt des Computers. Mit Hilfe der modernen Grafiktechnik soll sie eine neue qualitative Ebene des Erlebens ermöglichen; im Idealfall werden Betrachter der mit dem Rechner generierten Szenen "direkt" mit allen Sinnen in das Geschehen einbezogen, bis hin zur "Verbindung mit ihrem eigenen Intellekt" in einer künstlichen Welt. Andere Ziele, die im Zusammenhang mit der Technik der Breitbandkommunikationsnetze formuliert werden, gelten der Erschließung erweiterter Kommunikationsformen in einer virtuellen Gemeinschaft räumlich verteilter Benutzer ("Telepresence"). Bedingt durch die aktuellen Entwicklungen wurden auch schon Visionen geäußert wie: "Bonn und Berlin sind weit auseinander, man hat dann ein virtuelles Treffen in Helmstedt". Es sind jedoch auch künstliche Welten mit simulierten "Naturgesetzen" möglich, die sich radikal von unseren bisherigen Erfahrungsmöglichen unterscheiden. Welche Auswirkungen dies haben kann, ist bislang kaum untersucht worden, jedoch durchaus Gegenstand von Forschungsprogrammen⁴. Berichtet werden Effekte der Seekrankheit nach virtuellen Spaziergängen, und kolportiert wird, daß manche der Beteiligten – nach der Erfahrung des Durchdringens virtueller Wände – noch drei Tage später versucht haben, durch eine reale Wand zu gehen.

Die vieldiskutierten extremen Spekulationen von Hans Moravec⁵ möchte ich hier nicht erneut aufgreifen; unsere Auseinandersetzung damit hat in folgendem kurzen Absatz gemündet, mit dessen Zitat ich meine Ausführungen hier beschließe:

⁴ s. Computer Graphics Vol. 26 No

⁵ "Mind Children", Cambridge 1988

"Die sehr abgehobenen Spekulationen von Hans Moravec beziehen sich auf eine Übertragung des menschlichen Geistes auf Computer und somit die Abkopplung vom Körper. Hiermit werden Zukunftsszenarien entworfen, in denen die Fortexistenz menschlichen Kulturguts über die die Fortexistenz des Menschen als dominierende Spezies gestellt und der Mensch seiner Vorrangstellung beraubt wird. Es ist schwer, sich Menschen vorzustellen, die sich einem solchen Zukunftsentwurf anschließen wollen. Sich damit extrapolierend auseinanderzusetzen, fällt ebenfalls schwer, da solche Vorstellungen weit außerhalb vergleichbarer Erfahrungsmöglichkeiten des Menschen liegen. Das geeignete Genre dafür ist gegenwärtig am ehesten die Science Fiction"⁶.

Moravecs Entwürfe würden allerdings drastische Änderungen heutiger Vorstellungen vom Menschen beinhalten, und es ist allemal schwierig, sich über ihre Wünschbarkeit Gedanken zu machen.

⁶ siehe VDI Report 17 (1992), S. 167