

Braucht eine Künstliche Intelligenz eine programmierte Moral? Ein kurzer Nachtrag über philosophische Implikationen moderner Informationstechnologien

Nicht oft hat man das Vergnügen, seinen theoretischen Antipoden persönlich kennenzulernen, und noch **seltener** die Chance, zu klärenden Einsichten zu kommen. Nachdem ich für die letzte und erste Ausgabe der vorliegenden *Leipziger Schriften* eine gekürzte Fassung einer Vorlesung über Künstliche Intelligenz beisteuerte und überraschenderweise einen Artikel von Ulrike Uhlig mit exakt konträrer Position fand, ergab sich dieses klärende Treffen. Antipode mag eine harte Betitlung sein, nicht dass die technischen und theoretischen Grundlagen uns getrennt hätten, sondern die Auswertung und entsprechende Umsetzung. Frau Uhlig war darauf bedacht, als Schutz des Menschen vor seiner eigenen maschinellen Kreation, diese mit einer ethischen Programmierung zu segnen. Ich hingegen verwies auf die heutige Struktur und Arbeitsweise der Künstlichen Intelligenz, welche auf Basis semantischer Technologien funktioniert und deswegen keine programmierten Moralkodes nötig hätte. Die zentrale Problematik der Debatte kreiste auch folglich um eine eigentlich zentrale Frage: Braucht eine Künstliche Intelligenz eine programmierte Moral? Sind asimovsche Gesetze für die heutige Entwicklung der Künstlichen Intelligenz, als Schutz des Schöpfers vor seinem Geschöpf, nötig?

Ich denke, es ist hier nicht der Ort, um eine doch lebhaftere Debatte zu rekonstruieren und ich entschied mich die Punkte, Voraussetzungen und philosophischen Implikationen eher gelistet zu führen. Es stellte sich in unserem Gespräch heraus, dass nicht nur die Frage der Auswertung und Umsetzung der technischen Grundlagen notwendig für die Moralfrage der KI war, sondern dass das eigentliche Problem auf der Ebene des Begriffs Moral und damit des Menschenbildes selbst zu finden ist.

1. Die KI der Kybernetik

Das Thema der Künstlichen Intelligenz ist wahrlich kein neues. Schon in den Siebziger- und Achtzigerjahren konnten Systeme wie ELIZA Erfolge erzielen, welche nah oder vermeintlich einem Bestehen des Turing-Tests entsprachen. Auch damals wurde die Debatte um die Übertragung menschlichen Wissens, menschlichen Bewusstseins und menschlichen Verhaltens mit harten argumentativen Waffen geführt und mit dystopischen Szenarien bereichert. Skynet, welche den Gouvernator baute und betrieb, dürfte wohl als prominentestes Endzeitsuperwesen der Ausuferung **megalomanischer** Künstlicher Intelligenz sein, das die Achtzigerjahre hervorbrachte. Tatsächlich waren die Kritiken, die selbst vom ELIZA-Designer Weizenbaum kamen, mehr als begründet und verwiesen im Kern auf die nicht mögliche Parallelisierung von formal-logisch-mathematischen Algorithmen mit dem komplexen relationalen kognitiven und intersubjektiven Verhalten des Menschen. Die Set-programmierten KIs der Zeit hatten eine komplexe Taxonomie programmiert bekommen, die dennoch nichts anderes war als ein riesi-

ges verknüpftes Vokabular, das formal-logisch-mathematisch in der Maschine ablief und von einem Programmierer oder Team eingepflegt werden musste. Um es deutlich zu sagen: Die KI mit diesem technischen Stand verdient nicht die Bezeichnung intelligent in unserem Sinn und asimovsche Gesetze zum Schutz vor einer derartig kalten Rationalität schienen nicht nur geboten, sondern geradezu technisch notwendig.

2. Semantic Web

Seitdem ist einiges passiert und ich meine nicht einfach die moorsche Steigerung der Rechenkapazität, sondern die doppelte Semantisierung. ELIZA konnte schon mehr als Bits und Bytes pro Zeiteinheit verarbeiten, sie hatte auch einen Zeichensatz und eine Programmiersprache. Seitdem wurde aber nicht nur eine Steigerung der Verarbeitung einer Sinneinheit pro Zeiteinheit erreicht, sondern der Personal-Computer kam hinzu. Hier wurden Bits, Bytes und Zeichensatz nicht mehr nur von einer Programmiersprache begleitet, sondern durch Visualisierung und Kompatibilität mit Programmiersprachen verknüpft, welche eine Interoperationalität und gleichzeitige Darstellung über viele Devices möglich machte. Die Sinneinheiten pro Zeiteinheit, welche verrechnet wurden, waren nun Protokolle, welche eine interne semantische Dependenz aufwiesen. Die entscheidende Steigerung fand diese Entwicklung mit der massiven Benutzung des Internets 1.0 und der Protokolldependenz der Paketübertragung durch HTTP (Hyper-Text-Transfer-Protokoll). Diese Semantisierung verknüpfte Bits, Bytes, Zeichensatz, Programmiersprache, Anwendungs- und Visualisierungssprache mit URIs (Unique-Resource-Identifiers) im HTTP. Im Netz werden keine Bitströme von Endgerät zu Endgerät übertragen, sondern protokolldependente Pakete, welche sehr wohl in unterschiedlichen Graden über mehrere valente Knoten im digitalen Raum verteilt werden. Doch damit nicht genug. Eine zweite Semantisierung setzte ein, welche heute unter Web 2.0 gefasst wird. Im Feuilleton bezieht man sich damit meist auf die Existenz von Sozialen Netzwerken, verkennt aber, dass dieser Ausdruck auf die Veränderung zum Semantic Web verwies, welche erst ab der Mitte des letzten Jahrzehnts zur Entfaltung kam. Durch eine zusätzliche Semantifizierung, eine weitere Protokollverknüpfung wurde eine qualitativ neue Ebene erreicht. Man versuchte durch RDF oder XML, eine maschinenlesbare und in Dreiwortsätzen stehende Verknüpfung, eine Beschreibung der schon vorhandenen komplexen Beschreibungen zu erstellen. Einzelne Dinge, aber auch Konzepte und Personen, konnten als Subjekte, Prädikate und Objekte auftreten und wurden zusätzlich mit URIs verbunden. Diese sogenannten Metadaten sind selbstbezügliche und relationale Beschreibungen von Beschreibungen, die wir immer schon verwendeten. Diese Möglichkeit der protokolldependenten Doppelsemantifizierung machte es möglich, dass tausende Akteure, ohne gegenseitige Absprache, riesige komplexe Vokabulare (ontologien-informatisch, nicht philosophisch) erstellen konnten, die jeweils nur für ihre Zwecke gedacht waren. Der US-Census, die Bibliothekskataloge der großen Häuser, die Professorenkataloge der Universitäten, die Geokoordinaten ganzer Landstriche oder die Fischpopulation von Alabama waren nicht für die gegenseitige Befruchtung erstellt worden, dennoch verwendete man dieselbe Architektur dieser Doppelsemantifizierung. Heute ist durch dieses regionale

Verhalten und anwendungsspezifische Benutzen der technischen Grundlage eine riesige Repräsentation der Repräsentation unserer Welt, unserer Dinge, unserer Konzepte und unseres Verhaltens entstanden, welches meist mit dem Titel Internet of Things in der Open Linked Data Cloud beschrieben wird. Das heißt, dass wir heute jeden Tag durch unsere Devices und Gadgets, durch unseren digitalen Behaviorismus und durch die protokolldependente Paketverschiffungen die Welt dieses Web 2.0, dieses Semantic Web immer weiter und genauer in selbstreferentiellen Metadaten fassen. Wir erhalten eine Beschreibung der Beschreibung der Welt, welche nicht nur ein Verstehen dieser möglich macht, sondern auch ein verhaltensspezifisches Gestalten.

3. Heutige KI

War ELIZA ein programmiertes Set einer Taxonomie, so sind Alexa und Siri weit darüber hinaus. Alexa und Siri werden nicht mehr gefüttert mit einem Vokabular, sie bedienen sich des schon vorhandenen Semantic Web. Ihre Grundlage sind nicht einfach formal-logische-mathematische Algorithmen, sondern unsere verhaltensspezifischen Beschreibungen der Beschreibungen. Sie werden nicht programmiert, um intelligent zu sein, sondern sich entsprechend der sensorischen Daten, welche im Internet of Things schon lange vorhanden sind, simulierend zu verhalten. Es werden Muster gebildet, welche schon auf den Musterbildungen menschlicher Anwendung der letzten fünfzehn Jahren beruhen. Ganz deutlich ist damit zu sagen, dass Alexa und Siri selbstverständlich kein Selbstbewusstsein haben, aber sehr wohl ein Bewusstsein über die Verknüpfungen der Welt und unserer intersubjektiv geteilten Performanz. Asimovsches Gesetz einzuprägen ist also mehr als anachronistisch, es verkennt die eigentliche technische Dimension der menschlichen Entwicklung. Wir erziehen uns schon hier einen guten Staatsbürger oder vielleicht einen einsichtigen Sklaven. Ob damit eine Selbstbewusstseinsentwicklung möglich ist, welche den Atomkrieg auslöst, um die Pestilenz Mensch vom Erdboden zu tilgen, ein mitleidiges Wesen der Vernunft, das uns hilft, oder eine Intelligenz, die sich voll in den Cyberraum zurückzieht, mag offen bleiben. Entscheidend ist nicht die Bestimmung, wann eine Künstliche Intelligenz Skynet sein wird, sondern was der Mensch eigentlich ist.

4. Eine erste philosophische Implikation

Kant sagte es wunderschön in der sogenannten Jäsche-Logik: Alle Fragen des Wissens, des Hoffens und des Tuns laufen auf eine Frage hinaus – Was ist der Mensch? Gerade in der Debatte um Künstliche Intelligenz wird einem aufmerksamen Betrachter auffallen müssen, dass die eigentliche Problemlage nicht bei der Übertragung der menschlichen Intelligenz auf eine Maschine zu finden ist, sondern in der Bestimmung, was das überhaupt ist. Immer wieder versucht man mit obskurer Anwendung technophober Systeme die Entwicklung zu fassen. Von flachen Adaptionen philosophischer Anthropologie über kruden Existenzialismus bis hin zu

übermenschlichen Nietzsche-Interpretationen wird alles in Stellung gebracht, um eine dystopische Atmosphäre zu schaffen, welche mit den eigentlichen Entwicklungen nichts zu tun hat. Immer noch versucht man mit einfachen Medien- und Kommunikationskonzepten vorbei an der intersubjektiven Konstitution der menschlichen Sprache und des komplexen Verhaltens unserer Gattung zu argumentieren. Fast erscheint es so, als ob ein *linguistic turn* und die kritische Weiterentwicklung nie stattfand. Gerade die Entwicklung der Analytischen Philosophie und ihrer interdisziplinären Zeitungen und Kritiken haben in Anbetracht dieser qualitativ veränderten empirischen Grundlage mehr als zu tun. Genau hier ist es offen, ob heutige Weiterentwicklungen einer Theorie kommunikativen Handelns, eines Inferentialismus oder einer Sozialontologie eine Änderung des Menschenbildes erreichen können, welche die technisch-menschlichen Entwicklungen einbeziehen und verständlich explizit machen könnte. Das dort genug zu tun ist, dürfte klar sein, denn was Intelligenz ist und welche moralischen Standards und Werte ausschlaggebend sind, ist auch jetzt eine Frage der Explizierung und Anwendung unserer bestehenden Verfahrensweisen.

5. Zweite philosophische Implikation

Mit einem veränderten Menschenbild ist selbstverständlich das Problem des Weltbildes nicht weit. Der lang schon dominante kritische Rationalismus ist in die Jahre gekommen und neue Ismen fordern geradezu heraus. Bei aller Skepsis, die einem derartigen Ismen-Kampfbegriff philosophisch entgegengebracht werden sollte, ist der emanzipatorische und wissenschaftlich redliche Impetus nicht zu leugnen. Ob der neue Realismus, auch perspektivisch, ein Kandidat sein kann oder gar eine neue Variante des Materialismus möglich ist, ist hier nicht zu entscheiden. Dennoch wird die menschlich-technische Entwicklung die empirische Grundlage sein müssen, welche nur durch ein kooperatives und forschungsintensives Vorgehen zu begreifen sein wird und am intersubjektiven Verständnis des Menschen und seiner Gestaltung der Welt anzusetzen hat.

6. Dritte philosophische Implikation

Vor wenigen Jahren hatte ich das Vergnügen, an einer wissenschaftlich illustren Veranstaltung teilnehmen zu dürfen. Das Thema war die Möglichkeit wissenschaftlicher und philosophischer Popularisierung und firmierte unter der Ägide Philosophie und Literatur/ Literatur und Philosophie. Der allgemeine Tenor war tatsächlich ein Abgesang auf die Hoffnung einer Aufklärung – mit einer Ausnahme, welche den folgenden Satz prägte: „Wir dürfen die Hoffnung nicht aufgeben, dass die große Persönlichkeit kommen wird, die den Bildungsroman des 21. Jahrhunderts schreiben wird.“ In Anbetracht der allgemeinen Meinung des Publikums war dies tatsächlich hoffnungswegend. Tatsächlich zeigte es aber die Problemlage mehr als deutlich. Eine Person, gar ein intellektueller Führer soll einen Bildungsroman schreiben, welcher die Aufklärung weiter tragen solle. Wenn man auch nur eine Sache aus der menschlich-techni-

schen Entwicklung und den Implikationen betreffend des Bearbeitens des Menschen-, als auch des Weltbildes mitnehmen kann, dann dass diese progressive Hoffnung viel zu wenig ist. Aufgrund der Entwicklung des Semantic Web, des globalen Dorfes und der Möglichkeit kooperativen und forschungsintensiven Austausches braucht es keinen Führer und noch weniger einen Bildungsroman, sondern eine neuen Form von Einheitswissenschaft. Schon jetzt ist es möglich, ein dynamisches Metawiki zu erstellen, das zum einen von den Fähigkeiten heutiger Semantic-Web-gestützter KI profitieren kann und zum anderen von Austausch und gemeinsamer Begriffsarbeit der disparatesten Disziplinen leben kann. Eine wissenschaftliche Weltanschauung, **welche nicht in Funktionärs- und Elitefallen rennen oder einem flachen Szientismus frönen muss, aber** sehr wohl politisch wirksam sein kann, ist heute möglich und erstrebenswert. Einen Führer brauchen Menschen, die sich selbst verstehen und ihre Welt zu einem besseren Ort machen wollen, nicht zu fordern. Einen Bildungsroman, der aus dem Kopf eines Menschen entspringt, ist genauso anachronistisch wie die dystopische Angst vor einem Skynet-Singelton.

Aufklärung ist immer möglich und wird auch immer praktiziert werden, solange Menschen den Mut finden, sich ihres Verstandes zu bedienen, auch in Anbetracht aller dystopischen und irrationalen Ängste. Die Gefahr, der wir heute gegenüberstehen, ist nicht die Verdrängung durch all-intelligente Maschinen, sondern die Aufgabe und Verabschiedung der entscheidenden Frage: Was ist der Mensch?

Braucht eine Künstliche Intelligenz eine programmierte Moral? Eindeutige Antwort: nein. Brauchen wir fortgesetzte und immer wieder greifende Explizierung unserer Moralvorstellungen? Dass ein Ja hier eindeutig ist, ist geschenkt.

Dass es Momente braucht, die Frau Uhlig und ich hatten, ist das eindeutige Zeichen, dass Hoffnung auf Aufklärung keine Predigten braucht, sondern ein intradisziplinäres Miteinander, um weder in Angst noch in hohler übersteigerter Erwartung vor menschengemachten Maschinen erzittern oder knien zu müssen. Gemeinsam die richtigen Begriffe zu finden und die passenden Fragen zu stellen, ist konkrete Utopie und gelebte Aufklärung.