

# Der Lügenbot und andere Münchhausen-Maschinen

Oliver Bendel

Hochschule für Wirtschaft FHNW, Bahnhofstrasse 6, CH-5210 Windisch  
oliver.bendel@fhnw.ch

„Das Lügen ist wahrlich ein verdammtes Laster. Sind wir doch Menschen und gesellige Wesen nur durch die Sprache. Würden wir die Tragweite und Scheußlichkeit dieses Lasters recht einsehen, wir würden es mit Feuer und Schwert verfolgen mit mehr Recht als andere Verbrechen.“ (Michel de Montaigne) Vielleicht sollten wir es auch bei Chatbots und Robotern verfolgen, die in der Wirtschaft eine nicht unerhebliche Rolle spielen? Denn lügen können sie u.U. auch. Oder muss man alles ganz anders sehen und ist das Lügen, ob bei Menschen oder Maschinen, eine neue Tugend?

Können Chatbots lügen? Sicherlich kann man den Beratern auf den Websites der Unternehmen in der Regel vertrauen. Ein Lügenbot passt eher in die Laboratorien der Wissenschaft als in die Wirtschaft, obwohl man das Gegenteil anzunehmen geneigt ist. Oder in die organisierte Kriminalität, wie die Fembots, die sich in virtuellen Räumen als Frauen ausgeben und auf Kreditkartennummern und Passwörter von leichtgläubigen Männern aus sind. Aber die Unternehmen und ihre Kunden sollten wachsam sein. Maschinen können anscheinend lügen, wenn sie sprechen oder schreiben können, denn lügen bedeutet, die Unwahrheit zu sagen. Sie können lügen, wenn sie die Wahrheit kennen, und sie können die Wahrheit sagen, wenn sie die Unwahrheit kennen, wobei die Bestimmung von Wahrheit und Unwahrheit eigene Abhandlungen bräuchte. Der Duden erklärt „lügen“ mit „bewusst und absichtsvoll die Unwahrheit sagen“. Ob die Maschinen dazu in der Lage sind, darüber lässt sich streiten. Bewusst im engeren Sinne werden sie in absehbarer Zeit gar nichts tun. Absichtsvoll können (teil-)autonome Systeme durchaus sein. Sie wissen um die Wahrheit und entscheiden sich dann aus bestimmten Gründen für eine Lüge. Dass die Absicht letzten Endes auf einen Programmierer oder Auftraggeber zurückzuführen ist, versteht sich von selbst. Allerdings vermögen sich selbstlernende Systeme durchaus von ihren Wurzeln zu entfernen. Können Chatbots und andere Maschinen also lügen? Wenn man einen weiten Begriff voraussetzt und auf das Sprechen und Schreiben abhebt, genauer auf Aussagen, die wahr oder falsch sind, durchaus.

Maschinen versuchen uns ständig etwas zu sagen. Ein Beispiel ist Wolfram Alpha, die Computational Knowledge Engine. Wenn wir diese fragen, wie das Wetter in Zürich ist, gibt sie uns eine Antwort. Sie formuliert keine vollständigen Sätze, liefert uns aber Angaben wie „cloudy“, „wind: SW at 5m/s“ und „humidity: 80%“. Wir verstehen die Antwortmaschine, und wir nehmen an, dass sie die Fakten kennt und die Wahrheit spricht. Tatsächlich versucht Wolfram Alpha, die Informationen aus vertrauenswürdigen Quellen zu erhalten. Diese Voraussetzung scheint so selbstverständlich zu sein, dass sie gerne in Vergessenheit gerät. Aber Stephen Wolfram hätte sich ebenso dazu entschließen können, unsichere Quellen hinzuzunehmen oder gar zu bevorzugen. Oder seine hundert Mitarbeiter händisch Fehlinformationen in die Datenbank eingeben zu lassen. Man mag einwenden, dass das nicht sinnvoll wäre, aber darum geht es nicht. Wollte die hoffentlich bestens informierte Engine nun lügen, könnte sie einfach die Angaben abändern. Sie könnte wider besseres Wissen „clear“, „wind: SW at 50m/s“ und „humidity: 20%“ sagen. Ein sonniges, windiges, trockenes Wetterchen, erdacht von einem Computergehirn.

Das Buch „Können Roboter lügen?“ von Raúl Rojas aus dem Jahre 2013 enthält einen Essay mit dem gleichen Titel. Der an der FU Berlin tätige Experte für Künstliche Intelligenz (KI) meint darin: „Dürfen Roboter also Menschen belügen? Nach dem [!] Asimovschen Roboter-Gesetzen wäre es ihnen vermutlich nicht gestattet: Nach Asimovs Zweiten [!] Gesetz sind Roboter den Menschen zum Gehorsam verpflichtet, d.h. sie müssten auf eine von Menschen gestellte Frage mit der Wahrheit herausrücken.“ Der Held von Asimovs Geschichte „Spiegelbild“ („Mirror Image“) von 1972 ist anderer Ansicht: „Für gewöhnlich lügt ein Robot nicht, aber er wird es tun, wenn die Befolgung der Drei Gesetze es erfordert. Er könnte lügen, um sein Dasein zu schützen, wozu er nach dem Dritten Gesetz berechtigt ist.“ Und weiter: „Er wird noch eher lügen, wenn das nötig ist, um in Übereinstimmung mit dem Zweiten Gesetz einen berechtigten Befehl auszuführen, den ihm ein Mensch erteilt hat. Er wird am ehesten lügen, wenn es

in Übereinstimmung mit dem Ersten Gesetz nötig ist, um ein Menschenleben zu schützen oder einen Menschen vor Schaden zu bewahren.“ Die Shortstory dreht sich um zwei Wissenschaftler, die sich eine bahnbrechende Entdeckung streitig machen. Einer von ihnen muss lügen. Man befragt die metallenen Privatdiener der beiden, die die Behauptungen ihrer Herren bestätigen. Auch von ihnen muss also einer die Unwahrheit sagen. Dem Helden gelingt es durch die Befragung der Roboter und durch seine Menschenkenntnis, den maschinellen und den menschlichen Lügner zu entlarven. In der Geschichte „Lügner“ („Liar“) von 1941 belügt ein Roboter die Menschen, um sie nicht zu verletzen (und das Erste Gesetz einzuhalten). Und in „Kleiner verlorener Robot“ („Little Lost Robot“) von 1947 tritt ein kleiner Lügner namens Nestor 10 auf. Rojas räumt ein, dass ein gewisser Spielraum bleibe: „Nach dem Ersten Gesetz müssen Roboter die Menschen beschützen. Es könnte dann sein, dass eine ‚weiße‘ Lüge Menschen vor sich selbst schützt.“ Der Professor gelangt am Ende des Essays zum Schluss: „Roboter kennen die Wahrheit nicht, deswegen können sie nicht lügen.“

Wie dargelegt wurde, kennen Maschinen die Wahrheit durchaus. Es wurde ein Mittel gezeigt, um diese zu verdrehen. Man kann Angaben abändern und dadurch Falschbehauptungen erzeugen. Ein Chatbot weiß in der Regel, wieviel Uhr es ist. Deshalb verabschiedet er sich am Morgen anders als am Abend. Er wäre dazu fähig, die Uhrzeit durchzusagen, eine korrekte und eine inkorrekte. Man kann weiterhin Aussagen negieren, etwa indem man ein „nicht“ einbaut oder ein „ein“ in ein „kein“ umwandelt. Ein Bot könnte Kenntnis vom Wetter in Zürich erlangen. Ein wahrheitsliebender Bot würde, wenn es regnet, sagen: „Es regnet.“ Ein Lügenbot würde sagen: „Es regnet nicht.“ Er bekäme zutreffende Aussagen aus verlässlichen Quellen und würde sie in ihr Gegenteil verkehren. Vielleicht hat das Touristenbüro ein Interesse daran, bei einer Flaute die Menschen in die Stadt zu locken. In seinem Beitrag „Können Computer lügen?“ von 2003 entwirft Rainer Hammwöhner den Heuristic Algorithmic Liar, kurz HAL, dessen Ziel es ist, „möglichst viele Zimmer zu möglichst hohen Preisen zu vermieten“, wobei das Akronym an den berühmten Computer in Stanley Kubricks „2001: A Space Odyssey“ von 1968 gemahnt, der die Astronauten auf ihrem Flug belügt. Nicht, dass solche Strategien auf Dauer aufgehen oder nicht dadurch durchkreuzt würden, dass andere Quellen zur Verfügung stehen, wie die Nachrichten im Fernsehen, ein Wetter- oder Bewertungsportal. Aber es existieren nicht selten gute Gründe, einen Bot zu einem virtuellen Münchhausen zu machen.

Zu den potenziellen Münchhausen-Maschinen zählen neben Robotern – zu nennen wäre z.B. noch der sprechende Teddy von Supertoy Robotics – und Chatbots auch

Sprachassistenten wie Siri, die Frau auf dem iPhone. Die Sprache ist, wie ausgeführt wurde, die entscheidende Prämisse. Denn wenn man lügt, sagt man eben die Unwahrheit, man sagt etwas, in langen Sätzen oder mit wenigen Worten, mit oder ohne Grafiken und Fotos. Den Kern, ob Antwortmaschine, Chatbot oder Roboter, bildet jeweils ein Rechner bzw. ein Programm, und schon Asimov betont in einem Vorwort zu einer Reihe von Geschichten: „Ein Roboter ist in gewisser Weise nicht mehr als ein mobiler Computer, und ein Computer ist umgekehrt nur ein stationärer Roboter.“ Moderne Lügenmaschinen, in welchem Gewand sie auch auftreten, in welchen Körper oder Avatar sie auch schlüpfen mögen, sind immer Computer. Der Computer erfährt in unserem obigen Beispiel die Wahrheit von einem anderen Computer. Er kann sie freilich auch selbst finden, indem er Berechnungen anstellt, Schlüsse zieht, Daten analysiert. Wenn er sie gefunden hat, kann er sie modifizieren und manipulieren. Und er kann einen Sachverhalt überprüfen, indem er weitere sichere Quellen konsultiert.

Wenn viele Menschen häufig lügen würden, würde das einerseits unser Vertrauen in unsere Mitmenschen schädigen und unser Zusammenleben erschweren. Kant hat das Gebot der Wahrhaftigkeit für fundamental gehalten und ihm, wenn man so will, das Gebot der Menschlichkeit geopfert. Andererseits wird auf der ganzen Welt gelogen, dass sich die Balken biegen, und gäbe es eine Statistik der Lügen, wären ihre Zahlen länger als die Nase von Pinocchio und der Zopf von Münchhausen. Es scheint also davon abzuhängen, wer lügt, wann man lügt, wie man lügt, wo man lügt, warum man lügt. Unser Zusammenleben scheint trotz der Lügen zu funktionieren, von gelegentlichen Streiten, Fehden und Kriegen einmal abgesehen. Wenn viele Maschinen häufig lügen würden, würde das einerseits unser Vertrauen in die Maschinen zerstören. Andererseits wird auf Plattformen gelogen, bis sich diese biegen, soziale Netzwerke behaupten, dass uns jemand vermisst, jemand zu uns passt, jemand nicht genügend aktiv ist. Es scheint wiederum auf das Subjekt, die Zeit, den Kontext etc. anzukommen. Ein lügendes soziales Netzwerk wird eher hingenommen als ein die Unwahrheit sagender Wetterbericht.

Wie ein Lügenbot aufgenommen würde, bei dem das Lügen zum Programm geworden ist, müsste man herausfinden. An Projekten dieser Art scheint es nicht zu mangeln. So schrieb das Wissensmagazin scinexx im Jahre 2010: „Roboter lügen nicht – oder doch? Amerikanische Forscher haben jetzt ihren Maschinenwesen die Kunst der Täuschung beigebracht. Mit Hilfe spezieller Algorithmen lernten die autonomen Roboter, sich zu verstecken und eine falsche Spur für ihren Verfolger zu legen. Und sie lernten auch zu erkennen, wann eine solche Täuschung für sie vorteilhaft und sinnvoll ist.“ (<http://www.scinexx.de/>)

wissen-aktuell-12248-2010-09-13.html) Die Roboter lügen also nicht, sondern sie täuschen, wie die Blattinsekten, die Rojas in seinem Essay erwähnt, diese Meister der Mimikry. Für das Blenden und Betrügen braucht es nicht zwangsläufig eine Sprache, für das Lügen schon. Tiere können wahrscheinlich nicht lügen, und wenn sie es können, dann allein in ihrer tierischen Sprache, mit ihren Lauten, mit ihrem Tröten, Singen und Pfeifen.

Arkin wird von scinexx mit den Worten zitiert: „Wir haben von Anfang an auch die ethische Bedeutung im Zusammenhang mit zur Täuschung fähigen Robotern bedacht ...“ Und weiter: „Wir verstehen, dass es hier vorteilhafte und negative Aspekte gibt. Wir wünschen uns daher eine Diskussion über die Zulässigkeit von täuschenden Robotern[,] um herauszufinden, ob und welche Regulationen oder Richtlinien die Entwicklung solcher Systeme beschränken sollten.“ Bei der Beschreibung der Vorteile wird auf lügende Maschinen eingegangen: „So könnte es für einen im Rettungsdienst eingesetzten Bergungsroboter wichtig sein, in Panik verfallene Katastrophenopfer durch eine Lüge zu beruhigen oder zur Kooperation zu bringen. Und dass dem Militär solche lügenden Maschinenhelfer äußerst willkommen wären, ist auch klar. Sie könnten den Feind in die Irre führen und damit sich selbst und menschliche Soldaten schützen.“ Weitere bekannte Projekte, beispielsweise der ETH Lausanne, beziehen sich wiederum eher auf das Täuschen als auf das Lügen. Für einen Lügenbot zum Zwecke der Forschung wäre durchaus noch Platz.

Damit sind wir bei der Ethik angelangt. Münchhausen-Maschinen werfen Fragen innerhalb der Menschenethik auf. Sie sind zudem interessant für die Maschinenethik. Diese hat die Moral von Maschinen zum Gegenstand, und mit ihrer Hilfe (und der Hilfe der KI) kann man an Maschinen arbeiten, die die Wahrheit sagen oder die Notlüge bemühen, wenn es die Situation erfordert. Vielleicht wird sogar der GOODBOT, der seit Sommer 2013 an der Hochschule für Wirtschaft FHNW entwickelt wird, ab und an zu einem Lügenbot. Es wäre gefährlich, lediglich eine Pflicht- oder Pflichtenethik in der Tradition von Kant und Co anzuwenden, z.B. die Regel „Du sollst nicht lügen“ befolgen zu lassen; hinzukommen sollte die Folgenethik, die nach dem fragt, was Wahrheit und Lüge bewirken. Diese Klassiker der normativen Modelle sind grundsätzlich durchaus implementierbar, genauso wie die Tugendethik, was verwundern mag. Die Tugend ist ja nichts, was man bei Maschinen vermutet, so wenig wie das Laster. Auf jeden Fall gilt, dass nur sprechende und schreibende Maschinen das Laster (oder die neue Tugend) des Lügens haben können. Und dass wir es in der Hand haben, wie moralisch sich Roboter, Chatbots und Sprachassistenten uns gegenüber verhalten.

### Asimovs Robotergesetze

1. Ein Robot darf keinen Menschen verletzen oder durch Untätigkeit zu Schaden kommen lassen.
2. Ein Robot muss Befehlen eines Menschen gehorchen, es sei denn, solche Befehle stehen im Widerspruch zum Ersten Gesetz.
3. Ein Robot muss seine eigene Existenz schützen, solange dieser Schutz nicht dem Ersten oder Zweiten Gesetz widerspricht.

Quelle: Asimov, Isaac. Alle Robotergeschichten. 4. Aufl. Bastei Lübbe, Köln 2012.

Oliver Bendel ist studierter Philosoph und promovierter Wirtschaftsinformatiker. Er lehrt und forscht als Professor für Wirtschaftsinformatik an der Hochschule für Wirtschaft in Brugg und Olten (Fachhochschule Nordwestschweiz FHNW), mit den Schwerpunkten Wissensmanagement, Social Media, Mobile Business, Informationsethik und Maschinenethik. Sein letztes Buch trägt den Titel „Die Rache der Nerds“ und enthält Gedanken und Geschichten zur Informationsethik. Weitere Informationen über [www.oliverbendel.net](http://www.oliverbendel.net), [www.informationsethik.net](http://www.informationsethik.net) und [www.maschinenethik.net](http://www.maschinenethik.net).

Der Artikel „Der Lügenbot und andere Münchhausen-Maschinen“ von Oliver Bendel wurde am 11. September 2013 über <http://cyberpress.de/wiki/Maschinenethik> veröffentlicht. Diese Website existiert nicht mehr. Deshalb wurde dieser Postprint erstellt.