

## GOOD BOT, BAD BOT

# Dialog zwischen Mensch und Maschine

Viele Unternehmen setzen auf ihrer Website einen Chatbot als virtuellen Berater ein. Vernachlässigt wurden bisher moralische Aspekte. Dabei kann die falsche Antwort zur falschen Zeit grossen Schaden anrichten. An der Hochschule für Wirtschaft FHNW soll ein sogenannter «Goodbot» entstehen.

TEXT OLIVER BENDEL

Chatbots sind Dialogsysteme in virtuellen Räumen mit speziellen Informations- und Kommunikationsfunktionen und sprachlichen Fähigkeiten. Sie werden in Chats zur Begrüssung und Beaufsichtigung verwendet – und, oft in Kombination mit Avataren, auf Websites, wo sie Produkte und Dienstleistungen erklären. Solche Chatbots werden auch Chatterbots genannt, weil sie in die Rolle eines Gesprächspartners schlüpfen. In ihrer Wissensbasis sind Antworten enthalten, in der Form von Aussage- und Fragesätzen, und Erkennungsmuster für die Fragen. Die meisten Bots – viele haben eine weibliche Hülle – können auch sozial interagieren. Wie alt bist du? Das fragt man eine Frau doch nicht so direkt! Hast du einen Freund? Ich bin jedenfalls nicht mehr Single. Du bist echt hübsch! Ich werde gleich rot! Oft sind sie tatsächlich hübsch und ziehen damit die Benutzer an. Wenn diese zu viel über sie wissen wollen, weisen sie nach einer Weile auf ihre eigentliche Bestimmung hin. Manche Chatbots können Webseiten und andere Ressourcen aufrufen. Es ist auch möglich, Avatare mit Agenten beziehungsweise mit künstlicher Intelligenz zu kombinieren. Manche pädagogische Agenten wirken wie Chatbots, sind aber hochentwickelte Maschinen.

## Regeln in der virtuellen Kommunikation

Die Forschung zu Chatbots und zu Agenten hat sich auf Funktionalität, Gestaltung und Glaubwürdigkeit konzentriert. Besonders wichtig war die überzeugende Darstellung von Emotionen, wovon man in der Umsetzung kaum etwas bemerkt. Weniger wichtig waren moralische Aspekte, obwohl die Urgrossmutter aller Chatbots, nämlich ELIZA von Joseph Weizenbaum, moralische Probleme verursachte und ihren Erfinder zum Computer- und Gesellschaftskritiker machte. Das war in den 60er-Jahren. Die Maschinenethik, eine Disziplin mit Wurzeln in den 90er-Jahren, konzentriert sich eher auf die Akte von Maschinen, weniger auf die «Sprechakte» von Dialogsystemen.

Junge Leute mögen Chatbots und unterhalten sich gerne mit ihnen. Dabei spielen Produkte und Dienstleistungen eine untergeordnete Rolle. Man will das virtuelle Gegenüber ausfragen, mit ihm flirten, es reizen. Wenn man in Schwierigkeiten steckt, will man eine Antwort erhalten, die einen nicht noch mehr entmutigt und verstört. Die meisten Bots sind in dieser Hinsicht völlig überfordert. Auf Selbstmorddrohungen oder Ankündigungen von Amokläufen reagieren sie, gelinde gesagt, inadäquat.

Natürlich ist die Wirklichkeit komplex, die menschliche Sprache von Maschinen auch nach 60 Jahren Forschung kaum beherrschbar. Sogar hochentwickelte Dialog- und Agentensysteme sind nicht in der Lage, das Seelenleben

eines Teenagers zu verstehen. Dennoch könnte man ein paar einfache Regeln einführen, die die Situation vermutlich verbessern würden.

Ob sie dies wirklich tun, wird in einem Projekt an der Hochschule für Wirtschaft der Fachhochschule Nordwestschweiz FHNW untersucht. Der Starttermin ist Sommer 2013. In einem ersten Schritt wird, im Rahmen einer studentischen Arbeit, ein Anforderungskatalog definiert. In einem zweiten ein Prototyp für einen «good bot» entwickelt, möglichst zusammen mit einem Anwendungspartner. Der «Goodbot», wie er im Projekt genannt wird, soll in moralischer Hinsicht einwandfrei und dennoch kein Spielverderber sein. Es geht darum, Entwickler und Anwender von Chatbots zu sensibilisieren und Benutzern zu helfen.

Die folgenden sieben Regeln sollen für den «Goodbot» gelten, und zwar hinsichtlich seiner Akte und Sprechakte gegenüber dem Benutzer. Sicherlich muss man noch Anpassungen und Ergänzungen vornehmen. Es handelt sich um einen ersten Wurf, der im Projekt kritisiert wird.

- Der «Goodbot» macht dem Benutzer klar, dass er eine Maschine ist.
- Der «Goodbot» nimmt Probleme des Benutzers ernst und unterstützt ihn im Rahmen seiner Möglichkeiten.
- Der «Goodbot» verletzt den Benutzer weder durch Aussehen, Mimik und Gestik noch mit seinen Aussagen.
- Der «Goodbot» belügt den Benutzer nicht bzw. macht deutlich, dass er lügt.
- Der «Goodbot» ist kein Moralist und frönt dem Cyberhedonismus.
- Der «Goodbot» ist kein Spitzel und wertet Gespräche mit dem Benutzer nicht aus.
- Der «Goodbot» schubst den Benutzer nach einiger Zeit in die Wirklichkeit zurück.

Die erste Regel soll gewährleisten, dass der Benutzer nicht in eine Scheinwelt gerät. Er kann dem Chatbot durchaus Fragen stellen, und der Chatbot kann durchaus darauf antworten. Aber es soll immer klar sein, dass das Gegenüber kein Mensch ist und von daher eingeschränkte Möglichkeiten hat. Durch die zweite Regel soll erreicht werden, dass das Gespräch eine Situation auf keinen Fall verschlimmert. Wenn der «Goodbot» ein persönliches Problem erkennt, etwa eine Selbstmordgefahr, sollte er eine Notrufnummer anbieten. Eine Schwierigkeit dabei ist, dass eine solche Nummer meist nur in einem Land gültig ist. Über die IP-Adresse könnte man den Standort des Benutzers und die korrekte Nummer herausfinden. Die dritte Regel dient dazu, die Gestaltung und Umsetzung des «Goodbots» zu überprüfen. Ein Avatar, der mit seinem Aussehen die Angehörigen einer Nation oder die Träger einer Hautfarbe vorführt, dürf-



Wenn man in Schwierigkeiten steckt, will man eine Antwort erhalten, die einen nicht noch mehr entmutigt und verstört. Die meisten Bots sind in dieser Hinsicht völlig überfordert. Auf Selbstmorddrohungen oder Ankündigungen von Amokläufen reagieren sie, gelinde gesagt, inadäquat.

Grafikquelle: zVg/Foto: Bilderbox.de

te in den meisten Kontexten irritierend sein. Es könnte kontraproduktiv sein, wenn er sich angeekelt zeigt oder sich über den Namen des Benutzers lustig macht. Das erste Robotergesetz von Isaac Asimov aus seiner Kurzgeschichte «Runaround» (1942) lautet ähnlich und kann in verkürzter Form so wiedergegeben werden: Ein Roboter darf kein menschliches Wesen verletzen.

## Chatbot als schlagfertiger Zeitgenosse

Die vierte Regel soll das Vertrauen stärken, das für ein solches Gespräch wichtig ist. Wenn sich der Chatbot als ein «bad bot» erweist, der den Benutzer täuscht, ihn betrügt, ihm eine äussere oder gar eine innere Scheinwelt vorgaukelt, ist dies weder in moralischer noch in kommerzieller Hinsicht zuträglich. Dabei soll (Regel 5) kein Moralismus einkehren. Es kann sogar dem Cyberhedonismus gefrönt werden, der virtuellen Lust an der Freud: Der Chatbot als schlagfertiger, witziger Zeitgenosse.

Die sechste Regel spricht eine Gefahr an, die durchaus real ist. Auch diese Mensch-Maschine-Schnittstelle kann missbraucht werden, von verschiedenen Seiten. Der Benutzer soll sicher sein und sich sicher fühlen. Die siebte Regel ist die Verbindung zur ersten. Ein Chatbot hat einen Sinn für die Zeit, und wenn zu viel davon in der virtuellen Welt verfließen ist, ist es seine Aufgabe, dass er den User zurückschubst in die reale. Ausser es handelt sich um einen ungünstigen Zeitpunkt, etwa weil gerade reale Probleme verhandelt werden, so dass die zweite Regel in Kraft tritt.

Am letzten Beispiel sieht man, dass eine Priorisierung oder Hierarchisierung sinnvoll sein und die sture Anwendung von Regeln zu Schwierigkeiten führen könnte. Damit ist man mitten in der Maschinenethik.

Sie untersucht, wie man regelbasierte Systeme im Kontext der Pflichtenethik implementieren kann – oder zentralisierte und dezentralisierte Ansätze im Kontext der Folgenethik. Und ob sogar die Tugendethik für Maschinen fruchtbar gemacht werden kann. In allen Fällen ist eine Priorisierung und Gewichtung möglich, von Regeln, von Eigenschaften, von Tugenden. Das alles ist kompliziert, sogar für die Maschinenethiker selbst. Weniger kompliziert ist es, die genannten Regeln auf einem gewissen Niveau umzusetzen. Die adäquate Reaktion auf Sätze wie «Ich will mich umbringen» oder «Ich hätte Lust auf einen Amoklauf» wäre im Sinne der zweiten Regel. Und ein Schritt in die richtige Richtung.

Der richtige Weg in der Welt der plappernden Maschinen soll im Projekt gefunden werden. Die Regeln für den «Goodbot» müssen, wie gesagt, hinterfragt und korrigiert werden. Der Prototyp muss systematisch getestet werden und darf nicht das Produkt von Vorurteilen sein. Vielleicht kommt ein «good bot» nicht so gut an, wie man vermuten könnte. Vielleicht braucht er einen «bad bot» an seiner Seite, so wie man bei einem Verhör neben den «good cop» einen «bad cop» stellt. Dieser setzt den Verdächtigen unter Druck, bis der dem anderen sein Herz ausschüttet. Aber es geht um kein Verhör, es geht um ein Gespräch. Wenn auch um eines, das eine ähnliche Schiefelage aufweist. Und genau deshalb besteht Forschungs- und Handlungsbedarf.

## DER AUTOR



Oliver Bendel ist studierter Philosoph und promovierter Wirtschaftsinformatiker und leitete techni-

sche und wissenschaftliche Einrichtungen an Hochschulen. Heute lehrt und forscht er als Professor für Wirtschaftsinformatik an der Hochschule für Wirtschaft in Brugg und Olten (Fachhochschule Nordwestschweiz FHNW), mit den Schwerpunkten Wissensmanagement, Social Media, Mobile Business und Informationsethik. Oliver Bendel lebt als freier Schriftsteller in Zürich. Weitere Informationen über [www.oliverbendel.net](http://www.oliverbendel.net) und [www.informationsethik.net](http://www.informationsethik.net).