



Informatiker Oliver Bendel über lügende Roboter: «Ähnlich wie bei der Erforschung der Kernspaltung besteht die Gefahr des Missbrauchs»

Foto: Getty Images

«Der Lügenbot ist ein besserer Lügenbold als der Mensch»

Der Wirtschaftsinformatiker Oliver Bendel über Maschinen, die uns hinters Licht führen

Joachim Laukenmann

Wie gross ist das Gefahrenpotenzial durch virtuelle Sprachassistenten oder Chatbots, die uns unbeabsichtigt oder gezielt belügen?

Es ist immens. Es gibt heute zum Beispiel Newsportale, die Lügen verbreiten. Die werden zwar meist noch von Menschen gefüttert, zum Teil sind sie aber bereits automatisiert. Automatisierte Lügen können eine enorme Wucht entfalten. Das sieht man nicht nur an den Newsportalen, das sieht man auch an Social Bots, unter anderem auf Twitter.

Diese Social Bots erstellen automatisch Profile in sozialen Medien, tun so, als wären sie Menschen, kommentieren Beiträge oder streuen Falschinformationen.

Genau. Und sie kosten praktisch nichts. Daher kann man ganze Armeen von Lügenmaschinen loslassen. So können sich Falschbehauptungen im Interesse von wem auch immer automatisch unterstützt millionenfach verbreiten. Von Software-Robotern generierter Inhalt ist keine Science-Fiction mehr. Er erobert die Welt und kann massiv manipulieren.

Nun mischen Sie dabei auch mit, indem Sie den Lügenbot entwickelt haben, einen Chatbot, der systematisch lügt und mit dem man sich via Texteingabe unterhalten kann.

Die Entwicklung des Lügenbots bringt uns in der Tat in ein Dilemma. Ähnlich wie bei der Erforschung der Kernspaltung besteht die Gefahr des Missbrauchs. Im Prinzip könnte sich der Lügenbot als unmoralische Maschine in der Welt verbreiten und Schaden anrichten. Aber er ist einerseits mit einem Passwort geschützt, andererseits glaube ich, dass der Nutzen des Lügenbots überwiegt.

Ein Lügenbot soll einen Nutzen haben?

Mit dem Lügenbot wollen wir die Strategien des maschinellen Lügens aufzeigen und verstehen. So können wir Programmierer, Anbieter von Chatbots und die Nutzer auf die Gefahren hinweisen.

Wie lügt der Lügenbot?

Zunächst haben wir uns an den menschlichen Strategien orientiert. Im einfachsten Fall negiert man beim Lügen eine Aussage. Aus «Herr Bendel führt ein Interview» macht der Lügenbot «Herr Bendel führt kein Interview». Dann kann der Lügenbot Begriffe gezielt oder zufällig substituieren. Statt «Herr Bendel führt ein Interview» könnte er sagen: «Herr Snowden führt ein Interview».

Also lügt der Lügenbot wie ein Mensch?

Nicht nur. Im Grunde ist der Lügenbot ein besserer Lügenbold. Er kennt nämlich Strategien, die der Mensch nicht anwendet.

Zum Beispiel?

Er kann Begriffe sehr raffiniert ersetzen. Dabei greift er unter anderem auf eine Begriffswelt der

Princeton University zurück. Diese begriffliche Hierarchie oder Ontologie bildet im Grunde die ganze Welt ab und ordnet sie. Wird der Lügenbot etwa nach einem Tier mit langen Ohren gefragt, kann er in der Ontologie vom Hasen ein paar Stufen nach oben steigen zu den Wirbeltieren, dort etwas umherwandern, wieder absteigen und landet so zum Beispiel beim Igel. Dann antwortet er: «Igel haben lange Ohren». Neben der Ontologie der Princeton University nutzt er auch die Suchmaschine Yahoo und den Cleverbots, einen Chatbot-Kollegen, von dem er Auskünfte einholt, wenn er nicht mehr weiterweiss.

Was würde der Lügenbot auf die Frage antworten: Wer ist Präsident der USA?

Bei dieser Frage würde der Lügenbot wohl auf Yahoo zurückgreifen. Und Yahoo liefert natürlich Barack Obama als Antwort.

Klar. Aber es geht weiter. Der Lügenbot gibt den Begriff «Barack Obama» erneut in Yahoo ein. Yahoo liefert dann nicht nur Suchergebnisse, sondern informiert den Nutzer in einem Feld: Wer nach

«Barack Obama» gesucht hat, hat auch nach jenem Begriff gesucht, etwa nach «Donald Trump». Dort pickt sich der Lügenbot einen Begriff heraus und gibt erst jetzt eine Antwort an den Nutzer: «Donald Trump ist Präsident der USA». Wie die Suche in der Ontologie ist auch das eine Art zu lügen, wie sie nur eine Maschine tun kann. Wir haben knapp zehn solcher nicht menschlicher Lügenstrategien gefunden. Wenn jemand Geld und Zeit hat, könnte er wahrscheinlich noch zehn bis 20 weitere finden.

Lügt der Lügenbot immer?

Nein, er lügt zu rund 80 Prozent. Dazu muss man wissen, dass er die Strategien nicht immer nur einzeln anwendet, sondern oft kombiniert. So kann es vorkommen, dass er eine Falschinformation negiert und wieder bei der Wahrheit landet. Ist der Lügenbot bei gewissen Themen ein besonders guter Lügner?

Ja, er hat zwei Spezialgebiete. Die Tourismusbranche hat bekanntlich ein Interesse daran, eine Region ins beste Licht zu rücken. Der Lügenbot versucht das exemplarisch anhand der Region Basel. So könnte er zum Beispiel die Temperaturangabe zu Basel manipulieren, um Basel für Touristen attraktiver zu machen. Sein zweites Spezialgebiet ist die Nahrungsmittelbranche, die uns viel Unsinn verkauft. So versucht der Lügenbot Energydrinks trotz gesundheitlicher Bedenken anzupreisen.

Im Grunde müssten Firmen ihre Chatbots möglichst zuverlässig und glaubwürdig machen.

Sonst steigen die Nutzer früher oder später frustriert aus.

Genau. Wir glauben nicht, dass bei Unternehmen die Absicht besteht, mit Chatbots systematisch zu lügen. Allerdings haben sie nicht moralische, sondern pragmatische Gründe: Die Firmen wollen, dass man den Chatbots vertraut und sie

nutzt. Nicht umsonst werden wir von vielen Chatbot-Entwicklern kontaktiert, die den Lügenbot testen wollen, um mehr über die Fallstricke von Chatbots zu lernen.

Interessiert sich auch das Militär für lügende Maschinen?

Durchaus. Der führende Maschinenethiker Ronald Arkin vom Georgia Institute of Technology arbeitet direkt für das Pentagon. Er untersucht unter anderem Maschinen, die betrügen. Dazu schaut er sich das Verhalten von Raben und Eichhörnchen an, von denen man weiss, dass sie Artgenossen oder Feinde täuschen.

Hat das Schweizer Militär schon bei Ihnen angeklopft?

Nein. Darüber bin ich auch froh. Wenn wir immer mehr potenziell lügenden Maschinen begegnen, wie sollen wir damit umgehen?

Ich möchte keine Hysterie verbreiten. Aber Entwickler sollten sich genau anschauen, wo die Manipulationsmöglichkeiten liegen, und ihre Maschinen sicher machen. Das fängt damit an, dass sie ihre Chatbots gut vor Hackerangriffen schützen. Würden Hacker zum Beispiel die Wissensbasis eines Chatbots manipulieren, wäre das fatal.

Und die Botschaft für die Nutzer?

Nutzer sollten Chatbots und anderem Robo-Content nicht einfach trauen. Wie bei einer Website sollten sie sich informieren, welche Interessen dahinterstecken.

Was ist die wichtigste Erkenntnis aus der Entwicklung des Lügenbots?

Der Lügenbot war ein Low-Budget-Projekt. Wir haben gelernt, dass Lügenmaschinen mit relativ wenig Aufwand entwickelt werden können. Nach rund 200 Entwicklungsstunden funktionieren sie gut. Dann kann man sie beliebig duplizieren und millionenfach in die Welt setzen.

Entwicklung von moralischen Maschinen

Oliver Bendel, 48, ist Ethiker und Wirtschaftsinformatiker an der Fachhochschule Nordwestschweiz (FHNW) in Windisch AG. Er hat sich unter anderem auf Fragen der Informationsethik und Maschinenethik spezialisiert. Dieses Jahr hat er das Buch «Die Moral in der Maschine» publiziert. Er philosophiert nicht nur über moralische Maschinen, sondern baut diese auch. Bereits entwickelt hat er den Goodbot, einen Chatbot, der moralisch adäquat handelt, den Nutzern hilft, deren Probleme er-



Informatiker Oliver Bendel

kennt und im Extremfall eine Notrufnummer herausgibt. Das Gegenstück zum Goodbot ist der neu entwickelte Lügenbot. Noch dieses Jahr möchte Bendel den ethisch tadellosen Staubsaugerroboter Ladybird bauen. Dank diverser Sensoren soll dieser zum Beispiel Marienkäfer oder Schmuck erkennen und nicht einsaugen. Bendel interessiert sich auch für die Ethik autonomer Autos und hat sich im Rahmen der Studie Robocar mit einem beschäftigt, das auch für Tiere bremst.