WIE UNS KÜSTLICHE INTELLIGENZ MANIPULIERT

GESCHRIEBEN VON
MAXIMILIAN TRÄNKNER (MT093)
BENJAMIN SCHMIEDEL (BS095)

Der Einfluss künstlicher Intelligenz auf die Politik

Die Wahlsieger von Mexiko (2012), Brasilien (2014), den Vereinigten Staaten (2016) und Großbritannien (2016) aus diesem Jahrzehnt, verbindet eine Sache: Sie vertrauten in ihrem Wahlkampf auf den Einsatz von Künstlicher Intelligenz [1].

Diese extremen politischen Ereignisse zeigen was passieren kann, wenn die heutige Technik in die falschen Hände gerät. Mittels künstlicher Intelligenz werden falsche Nachrichten im Netz verbreitet und ein Meinungsbild geschaffen, was bis zur Manipulation eines Wahlergebnisses führen kann.

Wie es gelingt, mit künstlicher Intelligenz unsere politische Meinung zu beeinflussen, ist Thema dieses Artikels.

Fake News

Als Fake News werden Falschmeldungen bezeichnet, die mit manipulativer Absicht unter vielen Menschen verbreitet werden, um Meinungsmache zu betreiben [2]. Dies umfasst Deepfakes, gefälschte Bilder und Videos und Schlagzeilen. Zum einen kann KI helfen, genau diese Fälschungen zu erkennen. Zum anderen wird sie jedoch verwendet, um falsche Schlagzeilen zu erstellen und über automatisierte Bots an viele Menschen zu verbreiten und anschließend durch ganze Bot-Netzwerke zu verifizieren.

2015 taucht ein Video von Finanzminister Varoufakis aus Griechenland auf, in dem er Deutschland den Mittelfinger zeigt (Siehe Bild 1).

Bild 1: Varoufakis Mittelfinger (Quelle: Zdf_neo)

Varoufakis selbst bestreitet unter anderem in der Sendung von Günther Jauch, den Finger gezeigt zu haben und bringt eine vorsätzliche Fälschung des Videos ins Spiel. In der Diskussion um die Echtheit des Videos veröffentlicht Jan Böhmermann in seiner Sendung "Neo Magazin Royale" einen Beitrag, in dem er zeigt, wie sein Team aus der Hand des Finanzministers durch Videobearbeitung einen Mittelfinger erstellt hatte. Dies war eine weitere Bestätigung der These, das Video sei eine Fälschung [3].
Damit wurde eine große Verwirrung über die Wahrheit geschaffen. Niemand war sich mehr sicher, ob die Jauch-Redaktion mit ihrer Aussage, das Video sei echt, im Recht sei oder Varoufakis ehrlich war und das Video gefälscht wurde. Zu guter Letzt existierte noch Böhmermanns Version von der Fälschung des Videos durch die ZDF neo - Redaktion.
Nach einiger Zeit deckt Böhmermann jedoch auf, dass die Bearbeitung des Videos ebenfalls nur ein Fake war und das Mittelfinger-Video echt ist.
Damit schuf er sogar unter Medienprofis große Verwirrung und zeigte auf, wie einfach Manipulation in unserer Zeit ist.

Deep Fakes

Ian Goodfellow, Mitarbeiter von Google Brain warnte 2018 in einer Fachpublikation des Massachusetts Institute of Technology: „Es ist ein historischer Glücksfall, dass wir uns bisher auf Videos als Beweis für Tatsachen verlassen konnten" [4].
Worauf Goodfellow damit anspielt, ist die steigende Gefahr vor gefälschten Medieninhalten, die mit wachsender Leistungsfähigkeit der Technik einhergeht. Blickt man in die Zukunft, dann ist es wahrscheinlich, dass die Menschheit in ein Zeitalter verfallen wird, in dem nicht mehr abzugrenzen sein wird, ob die dargestellten Medien der Wahrheit entsprechen oder nicht.
KI-Forscher sprechen aus diesem Grund von einer "Informationsapokalypse", in der Fake-News zu Deepfake-News werden.

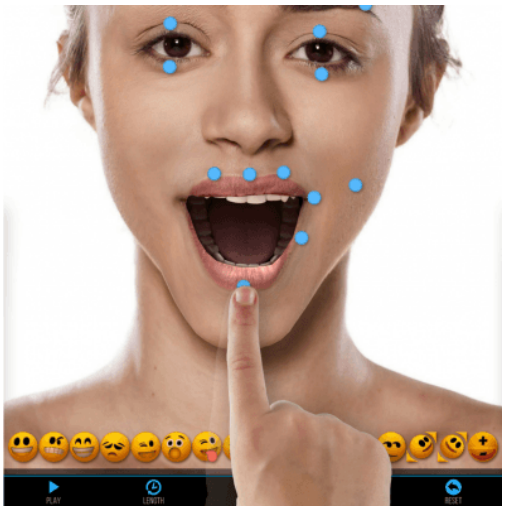
Bild 2: Deepfake erstellen per App (Quelle: MugLife)

Den Begriff Deepfake setzt sich aus den Wörtern "Deep Learning" und "Fake" (engl. für Unwahrheit) zusammen und bezeichnet Medieninhalte, die mithilfe von künstlichen neuronalen Netzen automatisiert erstellt wurden.
Die vielen Beispielen, die in jüngster Zeit im Netz kursierten [5], unter anderem eine gefälschte Obama-Rede, beweisen, wie weit fortgeschritten diese Technik schon ist.
Mit den entsprechenden Programmen wie bspw. der "Fake App" oder „Mug Life", die mit Googles Open-Source-Programmbibliothek für künstliche Intelligenz „Tensor Flow" operiert, kann schon heute jeder kinderleicht Deepfakes erzeugen (Siehe Bild 2).

Innerhalb wenigen Sekunden lässt sich damit aus einem einzelnen Foto eine gefälschte Animation basteln.
Momentan ist es noch nicht so einfach die Stimme einer Person zu manipulieren. Die Softwarefirma Adobe jedoch arbeitet bereits an einem "Photoshop für Audionachrichten" mit dem Namen Adobe Voco, mit dem es möglich sein wird, eine Sprachaufnahme in Textform zu wandeln, diese zu verändern und mit derselben Stimme wieder vorlesen zu lassen [6].

Big Nudging

Der Begriff Big Nudging setzt sich aus „Big Data“ und „Nudging“ zusammen.
Nudging ist ein Begriff aus der Verhaltensökonomie und geht davon aus, dass der Mensch nicht in der Lage ist selbstständig Entscheidungen optimal zu treffen. Die Theorie bezieht sich auf den Paternalismus, der eine Handlung beschreibt, die gegen den Willen, aber auf das vermeintliche Wohl eines anderen ausgerichtet ist [7].
Um die Entscheidung eines Menschen zu unterstützen, wählt Nudging jedoch nicht den Weg des Informierens und Überzeugens. Vielmehr werden psychologische Unzulänglichkeiten ausgenutzt, um jemanden zu bestimmten Verhaltensweisen zu bringen. Anders ausgedrückt: Der Mensch wird ausgetrickst.
Mit den oft ohne unser Einverständnis gesammelten persönlichen Daten offenbart sich, was wir denken, wie wir fühlen und wie wir manipuliert werden können. Diese Insiderinformationen werden ausgenutzt, um uns zu Entscheidungen zu bringen, die wir sonst vermutlich nicht treffen würden.

Big Data

Big Data steht für die Sammlung von einer riesigen Menge an Daten über eine große Anzahl von Menschen. Diese Daten liegen zu einem Großteil in unstrukturierter Form vor. Alles, was wir treiben, ob im Netz oder außerhalb, hinterlässt digitale Spuren. Jeder Einkauf mit der Karte, jede Google-Anfrage, jede Bewegung mit dem Handy in der Tasche und jeder Like wird gespeichert.
Inzwischen steht der Begriff Big Data für eine vollkommen neue Ära digitaler Kommunikation, in der es dank entsprechender Analyse Tools und Machine Learning recht einfach ist diese Menge an Informationen zu filtern und auszuwerten.

Mit einem auf das Individuum zugeschnittenen Newsfeed unter dem Suchfeld, möchte Google z.B. in nächster Zeit Antworten auf Fragen geben, die noch nicht einmal gestellt wurden [8]. Bekanntgebungen wie diese beweisen, wie viele Daten die Internetfirmen schon über uns gesammelt haben und dass es möglich ist, diese Informationen dafür einzusetzen unsere Persönlichkeit einzuschätzen und unsere Interessen vorauszusagen.

Lange war nicht ganz klar, wozu diese Daten gut sein sollen, außer dass in unserem Facebook-Feed Schuhe beworben werden, die wir uns am Tag vorher auf Amazon angeschaut haben.
Ebenso war unklar, ob Big Data eine Gefahr oder ein Gewinn für die Menschheit darstellt. Aber spätestens nach 2016, nach dem erfolgreichen Online-Wahlkampf von Donald Trump kennen wir die Antwort. Denn sowohl im US-Wahlkampf, als auch bei der Brexit-Kampagne wurden persönliche Informationen aus dem Netz gezielt eingesetzt [9].

Psychometrie

Die dort angewandte Vorgehensweise basiert auf der sogenannten Psychometrie. Psychometrie ist der wissenschaftliche Versuch, die Persönlichkeit eines Menschen zu

vermessen. In der modernen Psychologie ist dafür die sogenannte Ocean-Methode zum Standard geworden. Die meisten Persönlichkeitstests funktionieren nach diesem Modell [10].
Es sagt im Grunde aus, dass sich jeder Charakterzug eines Menschen anhand von fünf Persönlichkeits-Dimensionen messen lässt: Offenheit (d.h. wie aufgeschlossen jemand gegenüber Neuem ist), Gewissenhaftigkeit (Wie perfektionistisch ist die Person), Extraversion (oder Geselligkeit), Verträglichkeit (also wie rücksichtsvoll und kooperativ man ist) und Neurotizismus (Ist die Person leicht verletzlich oder ängstlich?).
Anhand dieser Dimensionen kann man relativ genau sagen, mit was für einem Menschen man es zu tun hat, welche Bedürfnisse und Ängste er hat und wie er sich in bestimmten Situationen tendenziell verhalten wird.
Das Problem war lange Zeit die Datenbeschaffung, denn zur Bestimmung musste man einen langen und sehr persönlichen Fragebogen ausfüllen. Mit den sozialen Medien und Big Data ergaben sich neue Wege.

Der erste, der Psychometrie und Big Data zusammenbrachte war Michael Kosinki. Zusammen mit seinem Team von der University of Cambridge entwickelten sie eine künstliche Intelligenz, die mit den Ocean-Werten von Testpersonen und deren Facebook- und Twitter-Profilen trainiert wurde.
Anschließend konnte anhand der Facebook Likes einer Person mit dem entwickelten Programm ein extrem zuverlässiges Persönlichkeitsprofil des Nutzers erstellt werden [11].
Ein Mann, der zum Beispiel, eine Kosmetikmarke liked, stuft die KI mit hoher Wahrscheinlichkeit als homosexuell ein. Einer der besten Indikatoren für Heterosexualität ist dagegen das Liken von der Hip-Hop Gruppe “Wu-Tang Clan”.

Im Jahr 2012 war Kosinkis Künstliche Intelligenz so weit entwickelt, dass sie aus durchschnittlich 68 Facebook-Likes eines Users seine Hautfarbe (95- prozentige Treffsicherheit), seine sexuelle Vorliebe (88-prozentige Wahrscheinlichkeit), seine politische Ausrichtung (85 Prozent), seine Intelligenz, seine Religion und seinen Alkohol-, Zigaretten- und Drogenkonsum voraussagen.
Bekannt geworden ist das folgende Zitat von Kosinski: „70 Likes reichen, um die Menschenkenntnis eines Freundes zu überbieten“ [12] und mit mehr als 300 Likes lässt sich angeblich sogar übertreffen, was Menschen von sich selber zu wissen glauben.

Anstatt psychologische Profile aus Personen zu erstellen ließ sich der entwickelte Ansatz nun auch umkehren. So konnte nach Personen gesucht werden, die einer bestimmte Persönlichkeiten entsprachen, z.B. nach allen unentschlossenen Demokraten. Für diese Menschen-Suchmaschine interessierten sich nach der Veröffentlichung vor allem die Firma SCL. Konsinski lehnte damals deren Angebot von 10 Millionen Dollar jedoch ab.

Cambridge Analytica

Cambridge Analytica (CA) war ein amerikanisches Datenanalyse-
Unternehmen mit dem Ziel durch Microtargeting Wähler zu rekrutieren.
In der Präsidentschaftswahl der USA in 2014 war Cambridge Analytica (CA) an 44 US-Wahlkampf-Kandidaturen beteiligt [13].
In der Öffentlichkeit wurde CA 2015 erstmals sichtbar, als es Ted Cruz bei seiner Bewerbung als Präsidentschaftskandidat der republikanischen Partei als ersten bedeutenden Kunden gewinnen konnte. Nach seiner Niederlage im Wahlkampf konzentrierte CA sich auf den Kandidaten Donald Trump. Die Grundlage für die Kontaktierung möglicher

Wähler waren die 50 Millionen Datensätze von 2014, die sie mit 1 Million Dollar erstellt hatten.
Im weiteren Verlauf der Wahlvorbereitung wurden die Datensätze auf 220 Millionen Bürger erweitert. Dies kostete Donald Trump Berichten nach umgerechnet 5,9 Millionen Dollar.
Seit 2. Mai 2018 ist das Unternehmen insolvent.
Es stellt sich die Frage, wie die Daten von einer so großen Menge von Nutzern bzw. möglichen Wählern gesammelt werden konnte.

Grober Ablauf

Cambridge Analytica kontaktierte eine große Anzahl potenzieller Wähler mit einem Persönlichkeitstest. Durch die Beantwortung des Tests hatten die Teilnehmer die Möglichkeit bis zu fünf Dollar zu verdienen. Die Testfragen beinhalteten den Bereich Politik und Fragen zur Einschätzung der eigenen Persönlichkeit. Um die Beantwortung einfach zu machen, bot Cambridge Analytica den Usern an, sich mit dem Facebook Account in der bereitgestellten App anzumelden. Ohne ihr Wissen wurden sämtliche Facebook Daten gesammelt, sowie die aller Facebook-Freunde der angemeldeten Person. Mögliche persönliche Daten waren dabei die verschiedenen Likes, Wohnort, Alter, Familienstand sowie Name einer betroffenen Person. Algorithmen kombinieren nun die Testantworten mit den Facebook Informationen und erstellten riesige Datensätze über Millionen von Menschen.
Die dadurch entstandenen psychologischen Persönlichkeitsprofile jeder Person führten zu personalisierten politischen Inhalten und Werbung, mit dem Ziel die Überzeugungen an die des Auftraggebers anzupassen.
Die beiden verwendeten Instrumente, um gezielte Personen mit Inhalten zu adressieren, sind sogenannte *dark posts*, sowie Microtargeting. *Dark Posts* sind gekaufte Inserate auf Facebook, die nur bei einer ausgewählten Zielgruppe erscheinen. Microtargeting beschreibt die an das Individuum und deren Persönlichkeiten angepasste Adressierung mit Werbung oder anderen Inhalten [14].

Social Bots

Soziale Netzwerke sind heute Podium für Politiker, Celebrities und Revolutionäre mit denen Millionen von Web-Usern auf einmal erreicht werden können. In den falschen Händen kann dies einfach zur Verbreitung von Unwahrheiten und Propaganda genutzt werden.

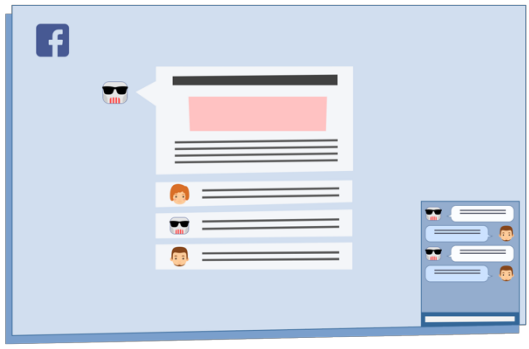

Bild 3: Social Bots auf Facebook (selbst entworfen)

Das übliche Mittel sind in solchen Fällen Social Bots. Das sind im Grunde kleine künstlich intelligente Software Programme, die sich hinter falschen Profilen auf sozialen Plattformen wie Twitter, Facebook oder Instagram für reale Personen ausgeben und sich somit unbemerkt durch das Netz bewegen. Sie versuchen im Prinzip das Verhalten echter Nutzer zu imitieren, um möglichst glaubhaft zu erscheinen [15].
Die Bots sind in der Lage eigenständige Entscheidungen zu treffen, wie z.B. welchen Nutzern sie folgen und welche Beiträge sie liken oder teilen. Außerdem besitzen sie die Fähigkeit die Texte von anderen einigermaßen gut zu verstehen und eigene Texte zu

verfassen. Dabei greifen sie auf eine Menge an vorgefertigten Texten zurück.
Social Bots analysieren also massenweise Posts und Tweets im Netz und werden dann automatisch aktiv, wenn sie bestimmte Hashtags oder Keywords im Netz entdecken, auf die sie vom Entwickler angesetzt wurden.
Dann liken, teilen oder verfassen sogar eigene Beiträge in der Absicht, ein Thema extrem zu pushen (siehe Bild 3).
Die Konsequenz ist, dass die Meinung von Facebook- oder Twitter-Nutzer durch die massenhaften Postings und die überproportionale Präsenz eines bestimmten Meinungsbildes leicht manipuliert und für eigene Zwecke missbraucht werden [16].
So sind viele Experten beispielsweise der Meinung, dass Social Bots mitverantwortlich für den Ausgang des Brexits in Großbritannien und der US-amerikanischen Präsidentschaftswahl 2016 zugunsten Donald Trumps sind.

Zahlen und Fakten zum Brexit

Die in der Fachzeitschrift "Social Science Computer Review" veröffentlichten Studie "The Brexit Botnet and User-Generated Hyperpartisan News" [17] veröffentlichte interessante Zahlen zur Aktivität von Social Bots vor den Abstimmungen zum Brexit. Es wurde offenbart, dass sich über 40000 Twitter-Konten kurz nach der Wahl selbst löschten, den Benutzernamen änderten oder durch Twitter blockiert wurden. Diese Anzeichen sprechen für eine Nutzung von Fake Accounts, die durch Bots gesteuert werden. Dabei war sichtbar, dass nur ein kleiner Teil für die eigentlichen Posts sorgten, während sich der Rest des Bot-Netzwerks für das retweeten und verbreiten der Posts verwendet wurde. Diese Annahme wird unterstützt durch das Ergebnis, dass 54 Prozent der Bots keinen Original-Tweet verfassten und es sich bei den meisten Beiträgen um Retweets handelte. Zudem wurden die über gefälschte Konten verschickten Tweets im Schnitt bis zu sechshundert mal retweetet. Ein weiterer Beweis für die Nutzung von Bots ist, dass zehn Prozent der Bot-Tweets über nur fünf Konten verbreitet wurden.
Ein oft genutztes Instrument der Bots sind zudem Posts mit falschen Nachrichten und temporär erstellten Inhalten. Dies wird durch die Statistik unterstrichen, dass 63 Prozent der URLs in Bot-Tweets nach der Wahl nicht mehr existieren oder funktionieren.

Erkennen von Bots im Alltag

Einige Maßnahmen im Kampf gegen die Verbreitung von Fake News wurden schon EU weit eingeleitet [18], dennoch hängt der Erfolg maßgeblich damit zusammen, inwieweit die Webuser in der Erkennung von Bots sensibilisiert sind.
Um im Alltag in sozialen Netzwerken zu erkennen, ob ein Account von einem Bot oder einer echten Person geführt wird, gibt es sechs Fragen, die es sich zu stellen gilt [19]:

1. *Wie seriös ist der Account?*
Kennt man die Person, die dort angeblich twittert? Kenne ich Follower des Accounts? Ist der Account verifiziert?

2. *Was verrät die Profilbeschreibung?*
Gibt es viele Details zur Person? Hat der Account Freunde oder Familie?

3. *Wie oft postet der Account?*
Bot-Accounts verbreiten oftmals eine sehr große Anzahl von Beiträgen in geringer Zeit und konstanten Abständen.

4. *Wie schnell reagiert der Account?*
Ein Bot reagiert schneller als jeder Mensch. Durch Algorithmen werden Kommentare oftmals automatisch innerhalb sehr kurzer Zeit veröffentlicht.

5. Wie viel gefällt dem Account?

Auch bei Likes übersteigen die Zahlen von Bot-Accounts oftmals eine realistische Anzahl.

6. Wie reagiert der Account mit Kontextfragen?

Kaum ein Chatbot kann mit Fragen, die räumliches Denken erfordern, zurechtkommen.

REFERENCES

[1] Arnaudo, D. (2017). *Computational Propaganda in Brazil: Social Bots during Elections*Washington

[2] Duden. (n.d.). Fake News

[3] Lange, S. (2015). Böhmermann katapultiert sich in den Medien-Olymp, *Welt*, from https://www.welt.de/vermischtes/prominente/article138565662/Boehmermann-katapultiert-sich-in-den-Medien-Olymp.html, accessed 16-3-2019

[4] Maier, S. (2018). Deepfakes – manipulierte Videos als Zeitbombe, *Main-Spitze*, from https://www.main-spitze.de/panorama/leben-und-wissen/deepfakes-manipulierte-videos-als-zeitbombe_18932335, accessed 16-3-2019

[5] Faked Obama Speech. (2018). *Youtube*, from https://www.youtube.com/watch?v=cQ54GDm1eL0, accessed 16-3-2019

[6] Adobe. (2018). Adobe Voco The Voice Manipulation Software, *Youtube*, from https://www.youtube.com/watch?v=E5EoMhv1XE0, accessed 16-3-2019

[7] Steinmann, M. (2015). Nudging – Paternalismus als Wolf im Schafspelz, *Ludwig von Mises Institut*, from https://www.misesde.org/?p=10199, accessed 16-3-2019

[8] Hurtz, S. (2018). Google will Antworten geben, bevor jemand Fragen stellt, *Sueddeutsche*, from https://www.sueddeutsche.de/digital/google-discover-1.4181596, accessed 16-3-2019

[9] Scott, M. (2018). Cambridge Analytica helped 'cheat' Brexit vote and US election, claims whistleblower, *Politico*, from https://www.politico.eu/article/cambridge-analytica-chris-wylie-brexit-trump-britain-data-protection-privacy-facebook/, accessed 16-3-2019

[10] smart-digits. (2017). Cambridge Analytica und das Ocean-Modell, from https://www.smart-digits.com/2017/02/cambridge-analytica-und-das-ocean-modell/, accessed 16-3-2019

[11] Drösser, C. (2017). Welche Datenspuren verraten unsere Persönlichkeit?, *Deutschlandrundfunk Kultur*, from https://www.deutschlandfunkkultur.de/psychometrie-welche-datenspuren-verraten-unsere.976.de.html?dram:article_id=376252, accessed 16-3-2019

[12] Grassegger, H.; Krogerus, M. (2016). Ich habe nur gezeigt, dass es die Bombe gibt, *Das Magazin*, from https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/?reduced=true, accessed 16-3-2019

[13] Sellers, F. S. (2015). Politics Cruz campaign paid $750,000 to 'psychographic profiling' company, *Washington Post*, from https://www.washingtonpost.com/politics/cruz-campaign-paid-750000-to-psychographic-profiling-company/2015/10/19/6c83e508-743f-11e5-9cbb-790369643cf9_story.html?noredirect=on&utm_term=.c8dc684b0cae, accessed 16-3-2019

[14] Kind, S.; Weide, S. (2017). *Microtargeting: psychometrische Analyse mittels Big Data*

[15] Social Bots – die Technik hinter Fake-News. (2018). *Ionos*, from https://www.ionos.de/digitalguide/online-marketing/social-media/social-bots-was-koennen-die-meinungsroboter-wirklich/, accessed 16-3-2019

[16] Mirau, F. (2016). Wie uns Social Bots beeinflussen und warum sie so gefährlich sind, *Basic Thinking*, from https://www.basicthinking.de/blog/2016/10/17/social-bots/, accessed 16-3-2019

[17] Bastos, M. T.; Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News, *Social Science Computer Review*, Vol. 37, No. 1

[18] Otto, T.; Hübert, H. (2018). EU will gegen Desinformation vorgehen, *Deutschlandfunk*, from https://www.deutschlandfunk.de/expertengruppe-zu-falschnachrichten-eu-will-gegen.2907.de.html?dram:article_id=412776, accessed 16-3-2019

[19] Sickert, T. (2017). So erkennen Sie Meinungsroboter, *Spiegel Online*, from http://www.spiegel.de/netzwelt/web/social-bots-entlarven-so-erkennen-sie-meinungsroboter-a-1129539.html, accessed 16-3-2019