

Künstliche Intelligenz zeigt ihre dunkle Seite

Ein Datenleck zeigt, wie der russische Machtapparat Hightech zur Zensur einsetzt. Anderswo wird die KI Chat-GPT als Weg in die Zukunft gefeiert. Auf diesem Weg lauern autoritäre Versuchungen.

Kommentar von [Jannis Brühl](#) 9. Februar 2023

Es gibt keine böse [Technologie](#). Entscheidend ist, wie der Mensch sie einsetzt. Das zeigt sich in diesen Tagen an den Nachrichten über künstliche Intelligenz (KI): lernfähiger Software, die gigantische Datenmengen verarbeiten kann, gehört die Zukunft. Nur muss noch ausgefochten werden, wie unangenehm diese Zukunft wird.

In dieser Woche eröffneten die Tech-Konzerne [Microsoft und Google](#) mit ihren eloquenten Chatbots [lautstark das Rennen um das große KI-Versprechen](#). Programme [wie Chat-GPT](#) oder Bard sollen Menschen und Computer noch enger zusammenrücken lassen. Suchmaschinen, Text-, Bildproduktion, Terminplanung - all das soll KI übernehmen. In der Tech-Branche herrscht Euphorie wie lange nicht mehr.

Zugleich lenkt ein Datenleck den Blick auf die dunkle Seite künstlicher Intelligenz: Der russische Staat arbeitet an selbständig lernenden Systemen, modelliert am menschlichen Gehirn. Sie sollen erfassen, wer in Videos oder Bildern den Angriff auf die Ukraine kritisiert, wer Präsident [Wladimir Putin](#) lächerlich macht, wer gleichgeschlechtliche Beziehungen lobt - was 2023 in Russland eben verboten ist. Das geht [aus den "Russian Censor Files" hervor](#), einem großen Datenleck aus der Telekommunikations- und Zensurbehörde Roskomnadsor, das die *Süddeutsche Zeitung* mit dem russischen Medium *iStories* [ausgewertet hat](#).

Der Computer soll lernen, was der Despot nicht mag. Ein digitaler Handlanger des Systems, Befehlsverweigerung ausgeschlossen. Wer den Krieg kritisiert, soll erfasst, der Beitrag von Online-Plattformen gelöscht werden. Zur Analyse von Textbeiträgen im Netz ist KI bei Roskomnadsor schon im Einsatz.

Künstliche Intelligenz zeigt ihre dunkle Seite

Dazu spannt Russland Technologie-Unternehmen aus dem Land ein und macht Druck auf internationale Plattformen, den Zensurgelüsten nachzukommen. In Moskau [bedrängten Agenten des Kremls vor dem Krieg sogar Managerinnen von Google und Apple persönlich](#). Sie mussten eine App, über die die Opposition Unterstützer mobilisierte, aus den App-Stores ihrer Konzerne werfen. Die Kontrolle der Social-Media-Sphäre wird auf absehbare Zeit ein Machtfaktor sein. Der Tech-Sektor wird zur politischen Beute.

Die moderne Diktatur ist Hightech

KI bietet neue Möglichkeiten, staatliche Späh- und Propagandaaktionen zu verschleiern. Roskomnadsor arbeitet an einer Art Bot-Fabrik, die Software-Programme - Bots - erzeugt. Sie sollen zum Beispiel in geschlossenen Chatgruppen menschliche Nutzer imitieren. So sollen Chats von Regimekritikern unterwandert werden. Was für ein Kontrast zu den Bots von Google und Microsoft! Selbsthilfe-Gurus und Produktivitäts-Extremisten verklären sie als Mittel, den Büromenschen von den Mühen der Arbeit zu befreien.

Ein Kult um Technologie [galt Umberto Eco](#) schon vor 30 Jahren als eines der Kennzeichen des Faschismus. Auch Diktaturen des 21. Jahrhunderts werden auf futuristische Hightech nicht verzichten, selbst wenn sie ihre Anhänger mit Hokusfokus wie der Beschwörung alter Imperien und Religionen bei Laune halten. Während der Angriff auf die Ukraine größtenteils mit traditionellen Waffen ausgetragen wird [und das Internet als Schlachtfeld eine unerwartet kleine Rolle spielt](#), funktioniert die Kontrolle der eigenen Bevölkerung heute nur noch über den Zugriff auf die Technologiekonzerne. Bestes Beispiel ist China, das mit seiner Abschottung des eigenen Netzes deutlich weiter ist als Russland.

Zu glauben, dass Demokratien frei von solchen autoritären Träumen sind, wäre arrogant. [Von Polen über Frankreich bis Italien](#) fantasieren Rechte und halbrechte Politiker von einem Algorithmus, der Arbeitslose automatisch in Jobs drängt, und der totalen Durchleuchtung von Migranten.

Künstliche Intelligenz zeigt ihre dunkle Seite

Die Interessen der Betroffenen wären der Maschine egal - und damit auch der Bürokratie.

Zudem ist unklar, wie gut und fair die Systeme sind, die westliche Konzerne wie Facebook auf Druck der Politik einsetzen, um sogenannte Hassrede herauszufiltern. Hier bieten der Rechtsstaat und neue Gesetze wie der "Digital Services Act" der EU wenigstens ein Mittel der Selbstverteidigung gegen übergriffige Plattformen, die unter dem Druck von Politik und Strafverfolgern stehen. Nutzer können Löschungen mit guten Argumenten rückgängig machen. Ein Luxus, den die Bürger Russlands nicht haben, wenn die KI auf sie anspringt.

© SZ/

Im Maschinenraum der russischen Zensur

Die Aufsichtsbehörde Roskomnadsor soll alle Informationskanäle in und nach Russland kontrollieren. Das klappt nicht so gut, wie sich der Kreml das wünscht. Interne Dokumente zeigen, wie künstliche Intelligenz bei der Zensur helfen soll.

Am 16. Oktober 2022 kommentiert die Youtuberin Tetiana aus Rheinland-Pfalz ein Video der Deutschen Welle. Am gleichen Tag postet Rudolf K. in Augsburg einen Beitrag auf dem russischen sozialen Netzwerk Odnoklassniki. Sie schreiben auf Russisch, beklagen Opfer und Gewalt des Angriffskriegs gegen die Ukraine. Die Tabelle ist eines von mehr als zwei Millionen internen Dokumenten, die von der Hackergruppe [Belarusian Cyber Partisans](#) an die *Süddeutsche Zeitung* geleakt wurden. Sie bieten einen einmaligen Einblick in einen modernen, digitalen Zensurapparat. Mehr als 1,5 Terabyte Mails und Protokolle legen die Methoden und Technologien offen, mit denen der russische Staat überwacht und löscht, was seine Bürgerinnen und Bürger im Internet verbreiten. Und nicht nur die. Auch Internetnutzer aus Deutschland sind betroffen. Interne Projektberichte zeigen, dass die RKN-Verantwortlichen ihre Kontrolle noch weiter ausbauen wollen und intensiv an künstlicher Intelligenz (KI) für die Überwachung und Zensur des Internets arbeiten.

Künstliche Intelligenz zeigt ihre dunkle Seite

Die Zensurbehörde RKN soll das russische Internet sauber halten, frei von Verstößen gegen die Gesetze und Regeln des Landes. Dazu gehören pornografische Inhalte mit Minderjährigen und Gewaltdarstellungen. Wegen solcher Inhalte ermitteln auch deutsche Staatsanwaltschaften und Polizeien, doch damit enden die Gemeinsamkeiten. Denn im russischen Internet ist es auch verboten, Bilder von nicht-heterosexuellen Beziehungen zu veröffentlichen oder den Angriffskrieg gegen die Ukraine als solchen zu bezeichnen. Artikel, Videos und andere Beiträge, die vom offiziellen Narrativ des Kremls abweichen, soll RKN schnellstmöglich zensieren.

Das Internet ist eine Herausforderung für die Medienaufsicht, denn Informationen verbreiten sich über die vielen verschiedenen sozialen Netzwerke rasend schnell. Zwar verfolgt der Kreml spätestens seit 2019 den Plan eines souveränen Internets, einer Parallelstruktur zum weltweiten Web, wie sie China mit seiner „Great Firewall“ gebaut hat. [Doch von so einem hermetisch abgeriegelten Modell ist Russland weit entfernt.](#) Die Zensoren haben nur direkten Zugriff auf russische Internetanbieter und Plattformen. Um auch die Inhalte ausländischer Medien und Unternehmen zensieren zu können, ist RKN auf deren Kooperation angewiesen. Anzeigen, Geldstrafen und die Drosselung des Datenverkehrs dienen als Druckmittel, um zu versuchen, Google und Facebook dazu zu bewegen, den Löschanfragen von RKN nachzukommen. Hilft auch das nicht, kann die Zensurbehörde Websites blockieren. So ist das Karrierenetzwerk LinkedIn seit 2016 nicht mehr aus Russland erreichbar, da es gegen russische Gesetze verstoßen haben soll. Das alles macht das Land zunehmend unattraktiv für ausländische Unternehmen. Einige russische Unternehmen beteiligen sich hingegen an der Zensur.

Brand Analytics durchkämmt für RKN das Internet

Die geleakten Dokumente zeigen, dass RKN seit 2022 eng mit dem Dienstleister Brand Analytics zusammenarbeitet. Das Unternehmen mit Sitz in Sankt Petersburg bietet Technik für die Analyse von Social-Media-Plattformen an.

Künstliche Intelligenz zeigt ihre dunkle Seite

Auf seiner Website wirbt es damit, eine Brücke zwischen Unternehmen und Menschen zu schlagen, von der beide Seiten profitieren. Der Medienaufsichtsbehörde hilft Brand Analytics allerdings, kritische Stimmen zu zensieren.

Täglich geht im Mail-Postfach von RKN ein Bericht von Brand Analytics ein. Darin zusammengefasst sind Social-Media-Inhalte in allen Sprachen der GUS-Länder zu Themen wie: „Aufrufe zum Protest“, „Krise der russischen Wirtschaft“ oder „Töten von Zivilisten“.

Ein Programm durchsucht Youtube, Facebook und weitere Plattformen gezielt nach Wörtern von einer Liste, die RKN-Mitarbeiter vorgeben.

Enthält ein Post beispielsweise die Schlagwörter „Russland“ und „Terrorist“, wird er unter dem Themenkomplex „Tötung von Zivilisten“ aufgeführt.

Mit Computerlinguistik-Software wird bewertet, ob verdächtige Beiträge in einem negativen, neutralen oder positiven Ton verfasst sind. Das soll bei der Eingrenzung auf Kritik helfen. Die automatisierte Zuordnung von „neutral“ oder „positiv“ ist aber nicht fehlerfrei.

Zum Schluss erstellt das Programm eine Liste der verdächtigen Inhalte und schickt diese automatisiert per Mail an RKN.

Am 16. Oktober 2022 enthält die Mail auch den Kommentar der Youtuberin Tetiana. Die Ukrainerin lebt in Rheinland-Pfalz, ist mit einem Deutschen verheiratet. Die Monitoring-Software findet ihren Kommentar unter einem Video des russischsprachigen Kanals der Deutschen Welle. Thema ist die Flucht von Russinnen und Russen in die EU. Tetiana kommentiert auf Russisch, Deutschland solle keine Menschen aufnehmen, die schweigend der Ermordung und Vergewaltigung ukrainischer Zivilisten zugeschaut hätten. Im Bericht von Brand Analytics ist Deutschland als Absenderland erkenntlich, der Kommentar ist zudem mit dem Hinweis „enthält Aggression“ versehen.

Künstliche Intelligenz zeigt ihre dunkle Seite

Grenzen der Zensur

RKN-Mitarbeiter gehen die Listen verdächtiger Posts durch, um die Stimmung in der Bevölkerung zu beobachten, frühzeitig auf regionale Protestaufrufe aufmerksam zu werden und Unruhestifter zu identifizieren. Verstößt ein Post oder Kommentar gegen russische Gesetze, stellt die Behörde Löschanfragen an Youtube, Facebook oder Odnoklassniki. Dank der Automatisierung von Brand Analytics spare man sich dabei vier bis acht Stunden Arbeit pro Tag, heißt es zufrieden in einer internen Notiz. Um auch geschlossene Gruppen und private Kanäle überwachen zu können, arbeitet RKN bisher mit händisch angelegten Fake-Konten. Weil die von den Plattformen schnell erkannt und gelöscht würden, arbeite man gemeinsam mit russischen Mobilfunkanbietern bereits an einer eigenen Bot-Farm, einem System, das automatisiert Fake-Accounts anlegt, heißt es in einem Projektbericht aus dem Leak. Die so generierten Social-Media-Bots sind kleine Softwareprogramme, die selbständig normales Benutzerverhalten simulieren sollen, indem sie regelmäßig Fotos und Texte posten oder auf andere Inhalte reagieren. So können sie länger unentdeckt spionieren.

Dass sie in einem Datenleck der russischen Zensurbehörde auftaucht, schockiert Tetiana aus Rheinland-Pfalz nicht: „Ich kenne Roskomnadsor nicht. Aber klar, ich weiß, dass es in Russland so läuft.“ Einschüchtern lassen wolle sie sich nicht. „Alle Kommentare, die ich irgendwann geschrieben habe, sind meine Meinung. Und mit allen meinen Kommentaren unterstütze ich die Ukraine.“ Auf Youtube bleibt ihre Meinung sichtbar, denn auf diese Plattform hat RKN nur begrenzten Einfluss. Interne Berichte aus dem Leak belegen, dass Google einer Löschanfrage für den russischen Kanal der Deutschen Welle bisher nicht nachkam. Anders sieht es bei dem Beitrag von Rudolf K. aus Augsburg auf der russischen Plattform Odnoklassniki aus. Ob aus freiem Willen gelöscht oder von RKN zensiert: Weder sein Post noch sein Account sind noch aufrufbar, der Link aus der Tabelle führt ins Leere.

Künstliche Intelligenz zeigt ihre dunkle Seite

Seit vergangenem Jahr ist die Entfernung von Informationen über den Angriffskrieg eine neue Hauptaufgabe von RKN. Einer geleakten Präsentation zufolge sollen keinerlei Aufrufe zu Protesten gegen den Krieg oder andere unerwünschte Informationen über die militärischen Aktionen auffindbar sein. Laut dem Bericht identifiziert RKN zwischen 24. Februar, dem ersten Kriegstag, und 10. November mehr als 200 000 solcher Beiträge, von denen etwa 60 Prozent erfolgreich entfernt wurden. Neben der Messenger-App Telegram bereite vor allem Youtube Probleme. Die Google-Plattform sei besonders unkooperativ bei der Löschung von Kanälen, unerlaubte Inhalte erreichten so weiterhin 60 Millionen Nutzer und Nutzerinnen. Zwar hat Russland zuletzt häufiger mit der Blockade von Youtube gedroht. Doch noch sind die ausländischen Plattformen zu beliebt unter Russinnen und Russen, können hauseigene Alternativen wie die Videoplattform Rutube nicht mithalten. Die RKN-Mitarbeiter schlagen deshalb vor, zunächst nur den Datenverkehr von Youtube zu drosseln. Einem Sprecher der russischen Non-Profit-Organisation Roskomsvoboda (übersetzt: Roskom-Freiheit) zufolge blockiert, verlangsamt und manipuliert RKN den Datenverkehr immer häufiger, sodass Websites nur mit Störungen und langer Ladezeit genutzt werden können. Außerdem werden diese neuerdings ohne Ankündigung blockiert. Roskomsvoboda setzt sich für digitale Rechte ein und dokumentiert die Sperrung von Onlinediensten. Die Netzaktivisten vermuten, dass RKN derzeit testet, wie ausländische Dienste noch effektiver verboten und Informationszugänge gestört werden können.

Der Traum von der umfassenden KI-Lösung

Trotz der Unterstützung durch Tech-Unternehmen wie Brand Analytics kann die Zensurbehörde nur einzelne Feuer der Meinungsfreiheit löschen. Mit der Entwicklung hauseigener KI verspricht sie sich, künftig auch die ersten Funken zu ersticken, die zu einer Flamme werden könnten. Einzelne KI-Werkzeuge sind den geleakten Unterlagen nach schon im Einsatz.

Künstliche Intelligenz zeigt ihre dunkle Seite

So durchsuche seit 2020 ein lernendes System mit dem Namen „Sauberes Internet“ die russische Suchmaschine Yandex nach verbotenen Inhalten. Eine ähnliche Software werte die Rezensionen und Beschreibungen von Audio- und Videomaterial nach Hinweisen auf Verstöße aus.

Mit den neuen Projekten, die Chiffren wie „Okulus“ und „Argus“ als Namen tragen, will RKN zukünftig auch Verstöße in Videos und Bildern erkennen. Eine aus Algorithmen bestehende Nachahmung des menschlichen Gehirns – ein sogenanntes neuronales Netz – ist laut internen Berichten bereits im Aufbau. Mitarbeiter haben 2022 damit begonnen, einem solchen lernfähigen Software-System beizubringen, illegale Inhalte zu erkennen.

Dafür sammeln die Mitarbeiter Videos und Bilder und klassifizieren sie als „enthält illegale Inhalte“ oder „enthält keine illegalen Inhalte“. Anschließend werden diese Daten dem neuronalen Netz als Trainingsdatensatz zugefügt.

Das neuronale Netz analysiert die Inhalte und versucht, Unterschiede zwischen dem Bildmaterial mit und dem ohne illegalen Inhalt festzustellen. Mit jedem neuen Bild lernt es mehr darüber, was die verbotenen Inhalte gemeinsam haben und was sie von erlaubtem Material unterscheidet.

Auf Grundlage dieser Analyse stellt das Programm Regeln auf und ordnet neue Bilder in die Kategorien „enthält illegale Inhalte“ oder „enthält keine illegalen Inhalte“ ein.

Das vom Programm selbst klassifizierte Bildmaterial prüfen RKN-Mitarbeiter dann händisch nach. Sie untersuchen, wie gut die künstliche Intelligenz schon funktioniert, und in welchen Fällen sie noch Fehlentscheidungen trifft.

Mit dem Informationssystem MIR-1 plant RKN zudem einen Alleskönner. Er soll mit KI das Risiko beziffern, dass ein verbotener Inhalt sich im Netz weiterverbreitet. Sogar Memes, also betextete Bilder, die nur mit Kontext zu verstehen sind, soll die KI interpretieren können, so steht es zumindest in dem Konzept. Das solle auch das Aufspüren von Beleidigungen des Präsidenten

Künstliche Intelligenz zeigt ihre dunkle Seite

Wladimir Putin erleichtern. Für alle geplanten Systeme ist ein straffer Zeitplan vorgesehen, die meisten sollen noch 2023 fertig werden.

Eine Genauigkeitsanalyse vom Januar 2022 attestiert den Systemen für automatisierte Videoerkennung Trefferquoten zwischen 13 und 72 Prozent. So richtig zuverlässig laufen sie also noch nicht. Die Wahrscheinlichkeit, dass Menschen zufällig ins Fadenkreuz geraten, ist bei diesen Quoten hoch.

Weder RKN noch Brand Analytics reagierten auf Anfragen der *Süddeutschen Zeitung*. Dass der Kreml aktiv versucht, unliebsame Inhalte aus dem Internet zu entfernen, ist seit dem Angriffskrieg gegen die Ukraine offensichtlicher denn je. Die Dokumente aus dem Datenleck der russischen Medienaufsichtsbehörde zeigen, dass Russland den Zugang zu unabhängigen Informationen künftig noch viel gezielter und schneller einschränken könnte. Doch solange ausländische Plattformen wie Youtube für Russinnen und Russen zugänglich bleiben und die Unternehmen dem Druck des Kremls nicht nachgeben, können weiter unabhängige Informationen ins Land kommen.