

# Hilfestellung für Jugendschutzprogramme

## Software spürt gefährdende Inhalte im Netz auf

Der Jugendmedienschutz-Staatsvertrag befreit Anbieter von der Verpflichtung, jugendbeeinträchtigende Inhalte nur zur Nachtzeit ins Netz zu stellen, wenn diese technisch für ein Jugendschutzprogramm gekennzeichnet sind. Bisher sind allerdings nur sehr wenige Angebote tatsächlich gekennzeichnet, sodass bei einer strengen Einstellung des Programms zu viele und bei einer großzügigen Einstellung zu wenige geblockt werden. Intelligente Software könnte hier helfen, ein

größeres Angebot als bisher in kürzerer Zeit zu analysieren. Was kann eine solche Technik leisten und wie kann sie optimal eingesetzt werden? *tv diskurs* sprach darüber mit Dr. Rolf Bardeli – beim Fraunhofer-Institut in Sankt Augustin zuständig für „künstliche Intelligenz“ –, der mit seinem Team zu diesem Thema im Auftrag der Initiative „sicher online gehen“ eine Machbarkeitsstudie durchgeführt hat.

**Sie beschäftigen sich in Ihrem Arbeitsfeld mit „künstlicher Intelligenz“. Was ist das genau?**

„Künstliche Intelligenz“ ist ja ein gefährlich klingender Begriff, der aufgrund der großen Versprechungen, die gemacht worden sind, für viele Jahre abgelehnt wurde und erst jetzt langsam wieder sexy wird. Ich sage momentan, dass ich mich mit der Mustererkennung auf multimedialen Daten beschäftige.

**Wie kann man sich diesen Job vorstellen?**

Meine Arbeit besteht aus ganz unterschiedlichen Teilen, was auch in unserer Finanzierung begründet liegt. In ungefähr der Hälfte unserer Zeit beschäftigen wir uns mit Forschungsprojekten, die wir beantragt und erhalten haben.

**Kommen die dazugehörigen Forschungsideen von Ihnen selbst oder sind es externe Aufträge?**

Diese Projekte kommen entweder von uns selbst oder von der Community, mit der wir zusammenarbeiten. Wir kooperieren aber auch mit der Wirtschaft, d. h. wir bekommen auch ganz konkrete Anfragen von Wirtschaftspartnern, die gelöst werden müssen.

**Wie sieht so etwas genau aus? Sie bekommen ein Forschungsziel und schauen, ob es mit Maschinen realisierbar ist?**

Dass man sich auf eine sogenannte Machbarkeitsstudie einigt und schaut, ob es funktioniert, ist ein mögliches Forschungsziel. Es gibt sicherlich auch einige Bereiche, bei denen man schon vorher erahnen kann, dass es funktionieren wird. Wir arbeiten dann daran, wie man es umsetzen kann.

**Computer und Roboter können bereits eine ganze Menge: in wahnsinniger Geschwindigkeit fast unendlich viele Variablen abchecken, durchdenken und entscheiden, welche Lösung für das Problem, das ihnen gestellt ist, die beste ist. Die Frage ist nun aber: Geht es nicht eher um die schnelle Analyse von Daten nach bestimmten vorgegebenen Kriterien?**

Zu dieser Frage gibt es unterschiedliche Lager. Ich selbst befinde mich in einem der konservativsten Lager. Ich glaube nicht an „künstliche Intelligenz“, ich habe noch keine gesehen. Für mich hat das viel zu tun mit der Berechnung von Statistiken, mit Zählen und Modellbildung, aber eben typischerweise mit sehr beschränkten Anwendungsszenarien. Die Stärke des Computers ist tatsächlich, dass er

sehr viele Daten einfach durchprozessieren kann, er erstellt darüber Statistiken und kann Modelle lernen. Was für ihn ganz schwierig ist und wo ich bisher auch noch nicht viel Überzeugendes sehe, sind die Themen „Kreativität“ und „Übertragbarkeit von Ideen“.

**Neben seiner Intelligenz gehören zu den wesentlichen Merkmalen des Menschen seine Gefühle. In dem Film Nummer 5 lebt geht es um einen kleinen Roboter, der durch einen Blitzschlag plötzlich eine Art Mensch wird. Wesentlich ist dabei, dass er Gefühle und Ängste hat. Ist so etwas für eine Maschine überhaupt denkbar?**

Es gibt sicherlich Wissenschaftler, die diesen Traum träumen und auch in diese Richtung forschen. Ich persönlich glaube nicht daran. Für mich ist das eine chemische Sache und nichts, was irgendwie in der Welt der Daten lebt. Ein Rechner muss zwar nicht immer systematisch sein, aber es ist doch etwas ganz anderes als das Datenprozessieren.

**Könnte es sein, dass wir eines Tages Rechner haben, die nicht mehr mit Festplatte und elektrischen Dingen funktionieren, sondern in einem Biosystem, das dann Gefühle eher ermöglicht?**

Das ist eine zweiteilige Frage: Dass man einen biologischen Computer hat, kann man sich sehr gut vorstellen. Dazu gibt es bereits erste Ansätze. Die zweite Frage aber, ob daraus Intelligenz, Emotion oder gar Bewusstsein entstehen, ist sehr schwierig. Eine Vorhersage ist kaum möglich.

**Kommen wir zu der praktischen Anwendung: Sie haben ein Forschungsprojekt zu der Frage durchgeführt, ob „künstliche Intelligenz“ uns im Bereich der Jugendschutzprogramme weiterhelfen könnte.**

In diesem Bereich sind wir viel optimistischer. Im Internet haben wir wahnsinnig viele Daten – ein Punkt, in dem der Computer gut und der Mensch schlecht ist. Ein Jugendschützer braucht sehr lange, um mehrere Millionen Webseiten anzuschauen, während ein Computer diese in null Komma nichts analysiert. Der Computer kann dann auch ganz spektakuläre Dinge durchführen, beispielsweise einen Text bewerten, Bilder erkennen und möglicherweise klassifizieren, ob sie in gewisse jugendgefährdende Kategorien gehören oder ob sich bestimmte Logos wiederfinden, bei denen man weiß, dass sie zu jugendgefährdenden Inhalten gehören. Der Computer wird letztlich nicht entscheiden können, ob ein Inhalt jugendgefährdend ist, aber er wird das Material vorfiltern und einschränken können, sodass der Mensch eben nicht mehr Millionen von Webseiten anschauen muss, sondern vielleicht nur noch 100, die besonders wahrscheinlich jugendgefährdend sind.



**Kann ein solches Programm auch Bilder erkennen und lesen?**

Ja, das kann es. In der Studie haben wir es an einem Beispiel durchgeführt, bei dem es um das Thema „Ritzen“, also selbstgefährdendes Verhalten, geht, zu dem im Internet zahlreiche Bilder gepostet werden. Wir haben uns die Frage gestellt, was dem Ritzen optisch ähnlich ist und kamen z. B. auf Tätowierungen. Die Frage war: Kann der Computer dies auseinanderhalten, also Bilder, in denen man das Ergebnis des Ritzens sieht, und Bilder, in denen Tätowierungen gezeigt werden? Es funktioniert! Wenn man genug Trainingsdaten hat, bei denen man weiß, dass das entweder das eine oder das andere ist, und einen Computer darauf trainiert, kann er mit relativ hoher Genauigkeit klassifizieren, ob es Ritz- oder Tätowierungsbilder sind.

**Der Computer weiß also nicht, dass es sich in dem einen Fall um Tätowierungen und im anderen um das Ritzen handelt, sondern er vergleicht einfach Daten und stellt fest, ob von Muster A oder Muster B mehr Anteile enthalten sind. Er sucht somit heraus, was die Jugendschützer sich anschauen sollen.**

Genau, so könnte man das verstehen. Da ist der Computer agnostisch, was die eigentlichen Konzepte angeht. Er versteht sie auch nicht. Das hat den Nachteil, dass er nur sagen kann, ob ein Inhalt wahrscheinlich zu Kategorie A oder B gehört. Gleichzeitig hat es aber den Vorteil, dass man dies für eine ganz breite Klasse von möglichen Kategorien durchführen kann, ohne dem Computer jedes Mal beibringen zu müssen, was diese Kategorie eigentlich bedeutet.

**Aber damit kann der Computer auch keinen Sinn erkennen. Wenn er Ritzbilder erkennt, weiß er nicht, ob die Bilder zur Nachahmung animieren oder ob es sich z. B. um eine Aufklärungskampagne handelt.**

Wenn der Computer nur die Bilder hat und er nur auf den Umgang mit Bildern trainiert wurde, erkennt er den Kontext nicht. Das könnte ein Mensch, wenn er nur Bilder und keinen Kontext hat, aber auch nicht. Geht man einen Schritt weiter und versucht, den Computer auch den umgebenden Text analysieren zu lassen, könnte man hier ähnlich wie bei den Bildern vorgehen: Ist es ein neutraler Text oder ist es ein zum Ritzen anregender Text? Dafür bräuchte man einfach mehr Daten. In unserer Studie haben wir diese Unterscheidung anhand von Texten versucht, die entweder aus einem nationalsozialistisch gesinnten oder aus einem neutralen Umfeld kommen. Wir hatten viele Trainingstexte und konnten entsprechende Modelle bauen, sodass eine Klassifizierung von neuen Webseiten möglich war.

**Im Jugendschutz haben wir es mit sehr komplexen und komplizierten Bereichen zu tun. Hier muss man sich vermutlich im Klaren darüber sein, dass der Computer letztlich nicht entscheiden kann oder darf, sondern lediglich eine Art Vorauswahl trifft, die dann vom Menschen angeschaut werden muss?**

Genauso sehen wir das auch. Der Rechner ist ein Mittel der Arbeitserleichterung und -reduktion. Wir wollen möglichst viel Material vorfiltern, um den Blick des Menschen möglichst gezielt auf die Stellen zu lenken, die interessant sind. Dabei kann man sich fragen, ob bei 500 Kategorien alle behandelt werden müssen oder ob es vielleicht zehn Kategorien gibt, die 90 % des Materials ausmachen. Wenn man so etwas identifizieren und in diesen zehn Kategorien das Material um 90 % reduzieren kann, hat man einen ganz großen Schritt gemacht.

**Aus der praktischen Perspektive betrachtet, wäre es vielleicht hilfreich, wenn es um die unzulässigen Sendungen ginge, etwa um Kinderpornografie.**

Dazu gibt es auch schon einige Arbeiten, selbst im kommerziellen Bereich. Gerade bei „Pornografie“ und „Kinderpornografie“ besteht ein großer Bedarf nach einer Lösung – und entsprechend viel wird hier auch geforscht.

**Wie ist es bei Gewalt? Hier könnte ein solches System in einigen Kontexten auch funktionieren. Der Computer könnte beispielsweise erkennen, ob Körperteile oder Köpfe abgetrennt werden.**

Genau, bis zu einem gewissen Grad könnte man das machen. Da kommen auch noch andere Aspekte ins Spiel, z. B. die immer wieder vorkommenden Aggregator-Webseiten, die solches Material sammeln und verbreiten. Diese blenden häufig ihr Logo ein, das wiedererkannt werden kann. Und somit kann schon recht viel Material einer solchen Webseite zugeordnet werden. Auch die Wiedererkennung von bekanntem, bereits bewertetem Material, das an anderen Stellen neu gepostet wird, funktioniert.

**Vor Kurzem habe ich von einer Software gehört, die in der Lage sein soll, Hate Speech zu erkennen. Halten Sie so etwas für möglich und für hilfreich?**

Ich halte es für ein nützliches und effektives Hilfsmittel, aber nicht für eine alleinige Lösung. Die Computeranalysen sind nicht perfekt. Wenn man annimmt, dass er 95 % dessen, was er beurteilt, richtig beurteilt, dann hat man große Mengen an Material, bei dem man einigermäßen sicher sein kann, ob man sich das anschauen sollte oder nicht. Es bleiben die fehlerhaft einsortierten Inhalte, bei denen man damit leben muss, dass man jemanden zu Unrecht entfernt hat oder dass es einen gewissen Anteil an gefährdenden Angeboten gibt, die durchrutschen. Wenn diese gemeldet werden, muss man sich mit ihnen auch von Hand beschäftigen. Das Problem ist nicht hundertprozentig automatisch gelöst, aber man hat wahrscheinlich deutlich weniger Arbeit, als wenn man alles als Mensch analysieren müsste.

**Kann ein Computersystem gerade im Bereich von Hate Speech unterscheiden, ob es sich um normalen Sprachgebrauch oder böse Degradierung handelt? Könnte es nicht sein, dass sich Unternehmen auf die Technik verlassen und so eine ganze Reihe von Mobbinggeschichten durch das Raster rutscht?**

Einerseits bewegen wir uns von dem bloßen Suchen nach Schlüsselwörtern ganz klar weg. Wenn man genug Daten hat, kann man auch Wortkontexte, sogar ganze Satz- oder Absatzkonstruktionen, implizit bewerten, modellieren und

aus mehr als nur dem Vorkommen der Wörter auf die Art der Texte schließen. Trotzdem kann man sich darauf nicht hundertprozentig verlassen. Möglicherweise gibt es bestimmte Kategorien, die man nicht abdecken kann, gerade wenn man vorher keine Idee hatte, dass es diese Kategorien überhaupt gibt. Vor allem Mobbing ist etwas sehr Subtiles. Das Computersystem befreit einen also nicht davon, mit menschlichem Verstand auf die Dinge zu schauen, aber wenn man es richtig einsetzt, hat man gute Chancen, dass weniger Menschen sich damit beschäftigen müssen.

**Angenommen, man würde diese Technik in Jugendschutzprogramme integrieren wollen, wie ginge man vor?**

Wir im Fraunhofer-Institut bringen aktuelle Forschungsergebnisse in die Anwendung. Wir könnten also Derartiges als Lösung konzipieren und bauen. Software muss natürlich auch gewartet werden, das können wir nicht leisten. Hier käme eine professionelle Softwarefirma zum Einsatz. Letztlich sind wir mit unserer Studie schon einen Schritt in diese Richtung gegangen, indem wir kurz- und mittelfristige Verfahren bereits klassifiziert haben und eben solche, bei denen noch geforscht werden muss. Es gibt einfach Dinge – wie z. B. das genannte Wiedererkennen bereits klassifizierter Inhalte, aber auch die Selektion und Wartung von Schlagwortlisten –, bei denen wir sagen: Das könnten wir sofort machen. Dann gibt es Bereiche, bei denen wir sehen, dass es funktioniert, aber die Software kommt gerade aus der Forschungswelt, deshalb muss man sie erst einmal in den täglichen Einsatz bringen.

**Die bereits verfügbare Software müsste – laienhaft gesprochen – so geschaltet werden, dass sie sich im Netz umschauf und nach bestimmten Zufallssystemen alle möglichen Webseiten in hoher Geschwindigkeit anschaut und dann bei bestimmten Bildern oder Begriffen „Stopp“ sagt.**

Das kommt ein bisschen auf das Material und die zu erkennende Kategorie an, ob man das zentral löst, z. B. durch einen Webcrawler, der sich Webseiten anschaut und diese zentral bewertet und das Ergebnis dann für die Jugendschutzprogramme zur Verfügung stellt. Manche Inhalte lassen sich auch live anschauen, wo das Jugendschutzprogramm also in dem Moment, in dem die Daten heruntergeladen werden, diese selbst bewerten kann.

**Es werden also nicht alle Dinge vorab bewertet, sondern das System erkennt im Moment der Nutzung, ob ein Kind einer entsprechenden Altersklasse einen Inhalt anschaut oder ein Erwachsener.**

Genau, das ist eine Möglichkeit. Ich würde mir jedoch eine Kombination vorstellen. Einmal gibt es Dinge, bei denen das Jugendschutzprogramm in seiner Liste, die möglicherweise von oder mithilfe eines Programms erstellt worden ist, nachschauen kann und weiß, dass diese Seite ein Kind nicht sehen darf, weil sie jugendgefährdend ist. Das kann daran liegen, dass ein Mensch sie gesehen und so eingestuft hat; es kann aber auch sein, dass man dem Computer vertraut oder dass der Computer es herausgesucht hat und ein Mensch noch einmal nachgeschaut und die Einschätzung des Computers bestätigt hat. Und es kann Material geben, auf das noch niemand geschaut hat, aber wir vertrauen z. B. dem Verfahren zum Wiedererkennen gefährdender Bilder. Das ist eine Datenbank, die ganz kompakt speichert, welche Bilder man bereits gesehen hat und bei denen der Klassifikator sagt: Das kenne ich schon.

**Wenn wir alle zu dem, was wir bereits haben, diese von Ihnen beschriebenen Programme einsetzen würden und genug Menschen hätten, die anschließend entscheiden, ob der Computer richtigliegt: Könnte man es damit schaffen, ein kinderfreundliches Internet zu entwickeln?**

Ich denke nicht, dass man es überhaupt schaffen kann, denn die Menge des Materials wächst ja unglaublich. Zudem ist es auch eine Frage, wie viel technischen Einsatz man sich leisten kann.

**Gibt es zwischen Ihnen und den Herstellern der Jugendschutzprogramme oder dem Bundesministerium Überlegungen, das Ganze weiterzuführen oder ist die Sache mit dem Bericht, den Sie vorgelegt haben, erst einmal beendet?**

Es wäre sehr sinnvoll, wenn man es weitermachen würde. Im Moment scheint zwar auf allen Seiten das Interesse da zu sein, aber es muss auch in irgendeiner Weise finanziert werden. Hier entsprechende Möglichkeiten zu finden, ist derzeit einer der großen Knackpunkte.

**Gibt es eine Richtung, wie viel so etwas kosten würde?**

Ich habe dafür keinen konkreten Preis im Kopf. Sicherlich ist es so, dass man mit einer Handvoll 10.000 Euro erste Bausteine auf den Weg bringen könnte. Das Ganze funktioniert modular. Von daher wäre es schön, wenn man ein, zwei solcher Bausteine als Proof of Concept starten könnte, um damit überzeugen zu können.

Das Interview führte Prof. Joachim von Gottberg.