Leonhardt Appel ist Tester im Qualitätsmanagement der International Age Rating Coalition (IARC), einem Zusammenschluss von Jugendschutzinstitutionen verschiedener Länder zur Altersbewertung von Onlinespielen und Apps. Er hat mit ATLAS (Artificial Intelligence Testing, Learning And Scraping) ein KI-Tool entwickelt, welches für die Recherche nach falsch eingestuften Apps Methoden des Machine Learning einsetzt, um solche Angebote schneller und in größerer Zahl zu entdecken. Am Beispiel von ATLAS und IARC wird deutlich, wie sich Algorithmisierung und der Einsatz von KI im Jugendmedienschutz sinnvoll mit dem gesellschaftlichen Diskurs und der Arbeit in Gremien verschränken lassen.

TEXT: CLAUDIA MIKAT UND CHRISTINA HEINEN

Es geht darum, die schwarzen Schafe zu finden!

Seit 2013 ist die Unterhaltungssoftware Selbstkontrolle (USK) Teil der IARC. Vor dem Hochladen einer App in einen Store füllen Programmentwickler*innen einen IARC-Fragebogen aus. Auf Grundlage ihrer Antworten wird automatisiert ein Alterskennzeichen vergeben – basierend auf der in jedem Land unterschiedlichen, kulturspezifischen Auswertungsmatrix. Der Fragebogen wird nur einmal ausgefüllt, die nach den gegebenen Antworten errechneten Einstufungen können jedoch von Land zu Land unterschiedlich sein, d.h. ein und derselbe Inhalt kann in Deutschland zu einer anderen Einstufung führen als beispielsweise in den USA oder in Brasilien. In Deutschland ist die Spruchpraxis der Gremien der USK Grundlage für die Auswertungsmatrix des Fragebogens. Dieser wird regelmäßig auf seine Aktualität geprüft und kann bei neueren Entwicklungen auch angepasst werden. So finden sich beispielsweise die im novellierten Jugendschutzgesetz (JuSchG) erwähnten In-App-Käufe oder Nutzerinteraktionen schon seit jeher im Fragebogen und werden als Deskriptoren ausgespielt. Gleichzeitig wurde das System im vergangenen Jahr aber auch darauf vorbereitet, auf eine neue Spruchpraxis reagieren zu können. Beim Erstellen der Ratings kommt es in seltenen Fällen zu Fehleinschätzungen aufgrund von – meist versehentlich gegebenen – falschen Antworten der App-Entwickler*innen, z.B. aufgrund von kulturellen Missverständnissen. In den Bereichen „Sex“ (Indien: sehr streng, USA: streng, Deutschland: liberal), „Gewalt“ (Deutschland: streng, USA: liberal) und „Drogen“ (Australien: sehr streng) sind die Unterschiede in der Bewertung besonders groß. In Deutschland hat das von der USK verantwortete Qualitätsmanagement der IARC die Aufgabe, falsche Ratings zu entdecken und zu korrigieren, indem Tester*innen falsch eingestufte Apps identifizieren und in Folge den Fragebogen erneut, diesmal korrekt, ausfüllen. Sämtliche IARC-Institutionen in den beteiligten Ländern erhalten daraufhin eine Nachricht über das berichtigte Rating und können dieses – entsprechend den kulturellen Anpassungen mithilfe ihrer länderspezifischen Auswertungsmatrix – übernehmen.

Machine Learning

Die Idee, ATLAS zu entwickeln, entstand aus der konkreten Arbeit im Qualitätsmanagement: Neben der Prüfung der Topdownloads, der meistgespielten Onlinespiele und beliebtesten Apps, zufälligen Stichproben und dem Filtern von Spielen und Apps nach verdächtigen Stichwörtern beinhaltet die durchaus zeitaufwendige Suche nach falschen Ratings, dass Tester*innen sich in Stores umsehen. Dort fahnden sie nach offensichtlichen Ungereimtheiten, einem Action-Shooter-Spiel beispielsweise, bebildert mit Waffen und Ego-Shooter-Perspektiven, mit einer Freigabe ab 0 Jahren. Da bestimmte Auffälligkeiten sich zwar sehr klar benennen lassen und wiederkehren, die Suche nach ihnen aber wie gesagt aufwendig und zeitraubend ist, entstand die Überlegung, eine künstliche Intelligenz nach diesen Mustern suchen zu lassen. Gelernt hat ATLAS das sogenannte Targeted Testing, die zielgerichtete Suche nach falschen Ratings, mit 40.000 korrekt klassifizierten Apps der Altersstufen 0–18 und den dazugehörigen Bildern und Texten aus den Stores. Das Besondere des Machine Learning ist dabei, dass die KI nicht nur wie einfache Algorithmen nach bekannten Mustern sucht, sondern in großen Datenmengen eigenständig neue, teilweise auch nur für Computer sinnhafte Muster erkennt, die dann zum verbesserten Aufspüren falscher Ratings genutzt werden.

Grobe Verstöße entdecken

Einer der Grundsätze des Qualitätsmanagements ist es, sich vorrangig auf grobe Jugendschutzverstöße zu konzentrieren. Um die Entdeckung unzulässiger Inhalte zu gewährleisten, wurde ein System von Alerts bei bestimmten Antworten und Antwortkombinationen im Fragebogen entwickelt. Dazu die Geschäftsführerin der USK Elisabeth Secker: „Wenn ein Entwickler angibt, unzulässige Inhalte in seinem Spiel zu haben, bekommen wir einen Alarm und schauen uns das an." In 98 % der Fälle sei der Alarm unbegründet, der Entwickler habe die Frage missverstanden und den Bogen falsch ausgefüllt, aber man wolle in diesem Bereich der unzulässigen Inhalte eine möglichst hohe Entdeckungswahrscheinlichkeit gewährleisten. Leonhardt Appel unterstreicht: „Es geht darum, die schwarzen Schafe zu finden. Wenn irgendwo auf der Welt ein Entwickler angibt, seine App enthalte sexuelle Gewalt in Verbindung mit Minderjährigen, wird ein Alarm ausgelöst – und wir reagieren."

Gleichzeitig, so Elisabeth Secker, würden aber auch die Stores wie z.B. der Google Play Store sehr sorgfältig darauf achten, dass rechtswidrige Inhalte von ihrem Angebot ausgeschlossen bleiben. Weitere Alerts würden bei sich widersprechenden Antwortkombinationen ausgelöst. Diese Automatisierungen im Qualitätsmanagement ließen sich über einfache Algorithmen bewerkstelligen.

Der Fokus im Qualitätsmanagement liegt auf groben Verstößen, aber auch Over Rating ist ein Thema. Dazu Elisabeth Secker: „Wichtig ist, dass das System nachvollziehbar bleibt. Deshalb korrigieren wir auch Einstufungen nach unten."

Technologischer Sprung

ATLAS wird zur Vorauswahl beim zielgerichteten Testen eingesetzt. Zu diesem Zweck erfolgt ein Abgleich des Ratings mit Bildern und Texten, die das Spiel im Store bewerben. ATLAS liest dabei Bilder (Image Recognition) und Texte (Natural Language Processing) aus, analysiert diese und trifft eine Vorhersage, wie hoch die Wahrscheinlichkeit ist, dass die Bilder und Texte zum Rating passen bzw. dass es sich um ein falsch eingestuftes Spiel handelt. Dabei liegt die KI in 85–90 % der Fälle mit ihrer Prognose richtig.

Grenzen sind erreicht, wenn es um Feinabstufungen etwa zwischen einer Freigabe ab 16 gegenüber einer Freigabe ab 18 Jahren geht. Diese stark kontextabhängige, feine Differenzierung im Bereich höherer Altersfreigaben findet sich in den Bildern nicht wieder. ATLAS nimmt eine automatisierte Vorauswahl möglicherweise kritischer Titel vor, die im Anschluss durch Tester*innen bei der USK händisch überprüft werden. Inzwischen wird Leonhardt Appels KI auch international von allen IARC-Institutionen eingesetzt. Bis Ende 2021 wurden durch ATLAS 500.000 Apps und 2,3 Mio. Bilder gescannt, über 5.000 Apps werden Tag für Tag von ATLAS geprüft.

Elisabeth Secker betont, dass die IARC eine Spruchpraxis mit dahinter liegenden Kriterien abbilde, die sich über die Jahre entwickelt habe. Beirat und IARC-Ausschuss profitieren auch von der Gremienarbeit der USK, durch die sie auf neue Phänomene und Themenschwerpunkte aufmerksam werden, wie z.B. die Casino-Apps als Trend der letzten Jahre. Dies ist wichtig mit Blick auf eventuelle Anpassungen des Fragebogens. Nach dem neuen Jugendschutzgesetz kann IARC nun auch der zuständigen Obersten Landesjugendbehörde zur Anerkennung vorgelegt werden und Anbietern von Spieleplattformen die rechtskonforme Erfüllung der Kennzeichnungspflicht ermöglichen.

Claudia Mikat ist Geschäftsführerin der Freiwilligen Selbstkontrolle Fernsehen (FSF).
Christina Heinen ist Hauptamtliche Vorsitzende in den Prüfausschüssen der FSF.