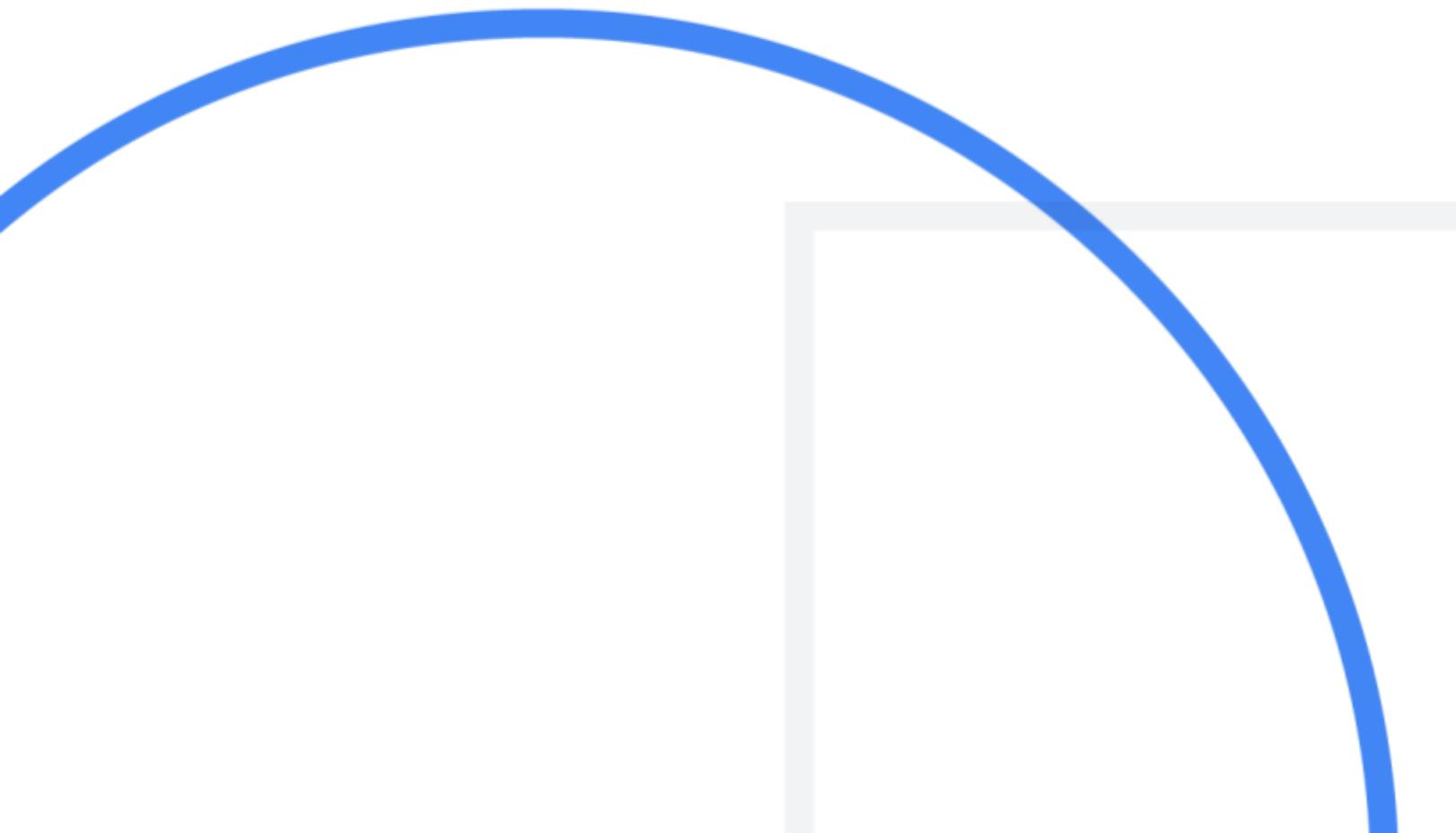


Informationsqualität &
verantwortungsvoller Umgang
mit Inhalten



Einleitung

Trotz seiner noch recht kurzen Geschichte präsentiert sich das Internet heute als herausragendes Kommunikationsmedium. Einst als Vehikel für den Datentransfer zwischen Computern konzipiert – unabhängig von Inhalt oder Zweck dieser Daten –, entwickelte es sich in wenigen Jahren zum Grundpfeiler der modernen Welt. Heute verbindet es Menschen und Gemeinschaften rund um den Globus. Es kann das Beste in unseren Gesellschaften hervorbringen, indem es den Zugang zu Wissen demokratisiert, unternehmerisches Handeln fördert und Kunst und Kreativität neue Möglichkeiten eröffnet.

Auch Google arbeitet unermüdlich daran, sein Angebot noch hilfreicher zu machen, für jede Nutzerin und jeden Nutzer überall auf der Welt. Konkret bedeutet das Hunderte von Milliarden Suchanfragen, pro Minute mehr als 500 Stunden Videomaterial hochgeladen auf YouTube, Milliarden von monatlichen Direktverbindungen zwischen Unternehmen und Kunden sowie viele hundert Milliarden US-Dollar an Wertschöpfung.¹ Hinter dieser Auflistung stehen zahllose Geschichten von Erkenntnis und Vernetzung, von Gemeinschaften, die Unterstützung für ihre Anliegen erhalten, und Menschen, die Antworten auf große und kleine Fragen finden.

Wir sehen uns in der Verantwortung gegenüber unseren Nutzerinnen und Nutzern, denn diese verlassen sich darauf, dass wir ihnen vertrauenswürdige, hilfreiche und für sie relevante Informationen bereitstellen. Wie jedes andere Kommunikationsmedium kann aber auch das Internet missbraucht werden. Die niedrigen Einstiegshürden, die seinen weltweiten Erfolg erst möglich gemacht haben, erweisen sich auch als Einfallstor für böswillige Akteurinnen und Akteure, die anderen Menschen Schaden zufügen wollen, sei es aus Profitgier oder aus ideologischen Gründen. Die Frage, wie man ein dezentral aufgebautes Internet so handhaben kann, dass positive Effekte begünstigt und negative Effekte verhindert werden, wird seit seinen Anfängen kontrovers diskutiert.

¹Siehe auch economicimpact.google.com;
www.blog.google/around-the-globe/google-europe/helping-europeans-succeed-googles-impact-in-europe

Ein Gegenstand dieser Debatten sind Inhalte wie Texte, Bilder, Videos und Webseiten, die Menschen miteinander teilen; Inhalte, die immer häufiger erzeugt und weltweit immer schneller verfügbar gemacht werden.

Unsere Mission bei Google ist es, diese Informationen zu organisieren und allgemein zugänglich und nutzbar zu machen. Im Mittelpunkt dieser Mission stehen die Relevanz und Qualität der Informationen, die wir den Nutzerinnen und Nutzern zugänglich machen. Unsere vielfältigen Plattformen sind dazu da, ihnen auf verschiedene Arten „hochwertige“ Informationen verfügbar zu machen. Darunter verstehen wir die jeweils nützlichsten, vertrauenswürdigsten und relevantesten Inhalte zum passenden Zeitpunkt. Zugleich arbeiten wir aktiv daran, einzelne Menschen und ganze Gesellschaften vor Schaden zu schützen, indem wir die Reichweite von „minderwertigen“ Informationen beschränken, d. h. von Inhalten, die weit unterhalb dieses Qualitätsanspruchs liegen.

Die Unterscheidung zwischen Informationen von hoher und niedriger Qualität ist eine anspruchsvolle, dynamische Aufgabe, für die es keine perfekte Lösung gibt. Die schiere Menge an Online-Informationen macht es unmöglich, jedem Inhalt dasselbe Maß an Aufmerksamkeit, menschlicher Urteilskraft und Sorgfalt zukommen zu lassen. Und auch wenn dies möglich wäre, könnten selbst verständige und sachkundige Personen ganz unterschiedliche Vorstellungen davon haben, was ein passendes Ergebnis ist und was nicht. Aus dem gleichen Grund kann es weder ein perfektes Ranking-System noch absolute Einigkeit über die Kriterien geben, nach denen es priorisiert.

Wir kennen dieses Dilemma und setzen uns damit auseinander, Tag für Tag, seit der Gründung von Google. Mit dem vorliegenden Whitepaper wollen wir unsere Grundsätze zu Informationsqualität und dem verantwortungsvollen Umgang mit Inhalten darlegen. Es formuliert die Prinzipien, die unseren Produkt-, Richtlinien- und Durchsetzungsentscheidungen zugrunde liegen, und benennt die vier Säulen, nach denen wir diese Prinzipien auf unsere Dienste anwenden. Es gibt zudem Einblicke in die Maßnahmen, die wir gemeinsam mit anderen Technologieunternehmen und mit unseren Partnerinnen und Partnern aus Wissenschaft, Zivilgesellschaft und Behörden ergreifen, um eine hohe Informationsqualität im Internet zu gewährleisten.

Wir freuen uns auf das Feedback zu diesem Whitepaper und werden auch künftig transparent über die hier behandelten Themen berichten.

Unser Verständnis von Informationsqualität

Die Mission von Google ist es, die Informationen dieser Welt zu organisieren und allgemein zugänglich und nutzbar zu machen. Diese Aufgabe bringt zwangsläufig Konflikte mit sich. Es gilt, die schwierige Balance zwischen dem ungehinderten Informationsfluss und Sicherheit, Effizienz und Genauigkeit, aber auch gesellschaftlichen Interessen zu wahren.

Die Produkt-, Richtlinien- und Durchsetzungsentscheidungen, die wir in diesem komplexen Umfeld fällen, folgen einer Reihe von Grundsätzen, die für alle unsere Produkte und Dienste gleichermaßen gelten:

- **Offenheit und Zugänglichkeit:** Wir wollen den Zugang zu einem offenen und diversen Ökosystem von Informationen bereitstellen. Dies bedeutet jedoch nicht, dass mit unseren Diensten beliebig umgegangen werden kann. Wie wir weiter unten noch ausführlich beschreiben werden, ist das Löschen von Inhalten ein wichtiges Instrument zur Sicherstellung der Informationsqualität. Es ist allerdings nur eines von mehreren Instrumenten, die uns zur Verfügung stehen, und wir verwenden es mit größter Sorgfalt, insbesondere in Zusammenhang mit der Google Suche. Ein vernünftiger und verantwortungsbewusster Ansatz zur Sicherung der Informationsqualität sollte unserer Ansicht nach darauf abzielen, Inhalte und Informationen frei zugänglich zu halten.
- **Interessen unserer Nutzerinnen und Nutzer respektieren:** Wer sich für konkrete Inhalte interessiert, die weder rechtswidrig sind noch unseren Richtlinien widersprechen, sollte diese auch finden können. Das gilt auch dann, wenn alles darauf hindeutet, dass es sich um Inhalte von vergleichsweise niedriger Qualität handelt. Eine deutlich höhere Messlatte legen wir an, wenn jemand eine weniger konkrete Suchanfrage stellt.

- **Alle Menschen im Blick behalten:** Unsere Dienste werden rund um den Globus von Nutzerinnen und Nutzern mit diversen kulturellen, sprachlichen und anderweitigen Hintergründen und in ganz unterschiedlichen Lebensphasen in Anspruch genommen. Während einige in eine Welt der Smartphones hineingeboren wurden, haben andere den Großteil ihres Lebens ohne Internet verbracht. Bei der Entwicklung unserer Produkte und Richtlinien sowie bei unseren Entscheidungen zur Durchsetzung dieser Richtlinien beachten wir die Vielfalt unserer Nutzerinnen und Nutzer und streben danach, ihren Bedürfnissen möglichst gut gerecht zu werden.

Diese Prioritäten sind seit unserer Gründung fest in unserer Arbeit verankert. Sie spiegeln sich in unserer Herangehensweise an Informationsqualität wider und berücksichtigen den Wandel der Nutzungserwartungen und Normen, die zunehmende Raffinesse von Akteuren mit böswilligen Absichten und die Weiterentwicklung des Internets insgesamt.

Jedes Produkt und jeder Dienst von Google hat seinen ganz eigenen Zweck. Genauso hat jede Nutzerin und jeder Nutzer ihre und seine ganz eigenen Erwartungen bezüglich der Inhalte, die dort präsentiert werden. Bei der Festlegung, welche Inhalte mit den einzelnen Produkten und Diensten bereitgestellt werden, gehen wir daher äußerst sorgfältig vor.

Unsere Produkte und Dienste bewegen sich in einem breiten Spektrum, für das fein abgestufte Beschränkungen gelten. Google Chrome ermöglicht den Zugang zur gesamten Bandbreite der Informationen im Internet. Warnhinweise gibt es nur bei Seiten, die Malware enthalten könnten. Die Google Suche dient als Index für alle im offenen Web abrufbaren Seiten. Nutzerinnen und Nutzer können hier jede legale Seite finden, die zu ihrer Suchanfrage passt. Innerhalb des erwähnten Spektrums befindet sich die Google-Suchmaschine an dem Ende, für das die wenigsten Beschränkungen gelten. Am anderen Ende befinden sich unsere Produkte rund um Werbung. Für sie gelten die strengsten Beschränkungen, da wir nicht von Anbietern profitieren wollen, die schädliche Inhalte oder belastende Nutzungserfahrungen erzeugen. Die übrigen Produkte und Dienste liegen zwischen diesen beiden Extremen. So gibt es bei Gmail z. B. nur minimale Beschränkungen für Inhalte, während YouTube als Plattform zum Hochladen und Teilen von Inhalten innerhalb einer Community deutlich mehr Schutzmechanismen erfordert als die Google Suche.

Auch die einzelnen Funktionen eines Produkts können zu verschiedenen Bereichen des Spektrums gehören. Einige Funktionen der Google-Suchmaschine wie beispielsweise die Autovervollständigung („Autocomplete“) liefern Informationen, die Nutzerinnen und Nutzer möglichst schnell zu den gewünschten Ergebnissen führen sollen.² Zugleich sollen jedoch Nutzerinnen und Nutzer, die nicht ausdrücklich danach gesucht haben, keinesfalls mit potenziell verstörenden Inhalten konfrontiert werden. Für diese Funktionen haben wir daher Richtlinien entwickelt, die das Anzeigen z.B. von pornografischen Darstellungen, verbaler oder körperlicher Gewalt ausschließen. Diese Vorsichtsmaßnahmen bedeuten jedoch nicht, dass Nutzerinnen und Nutzer nicht nach beliebigen Inhalten suchen können.

Um die Informationsqualität sicherzustellen und Inhalte für die Vielzahl unserer Produkte und Dienste zu moderieren, orientieren wir uns an vier einander ergänzenden Prinzipien:

-  **Entfernen:** Wir definieren verantwortungsbewusste Richtlinien für alle unsere Produkte und Dienste und leiten Maßnahmen gegen Inhalte und Verhaltensweisen ein, die gegen diese Richtlinien verstoßen. Wir kommen außerdem den gesetzlichen Verpflichtungen zur Löschung von Inhalten nach.
-  **Sichtbar machen:** Wir stufen hochwertige Inhalte und zuverlässige Quellen hoch, wo und wenn es darauf ankommt.
-  **Reduzieren:** Wir schränken die Verbreitung von potenziell schädlichen Informationen dort ein, wo wir Inhalte vorstellen oder empfehlen.
-  **Belohnen:** Wir setzen hohe Qualitäts- und Zuverlässigkeitsstandards für Inhalteersteller, die ihre Inhalte monetarisieren oder bewerben möchten.

Mit diesen vier Prinzipien können wir einheitliche Kriterien auf unser gesamtes Produktportfolio anwenden. Zugleich können wir sie gezielt einsetzen, um den individuellen Anwendungen und deren Erfordernissen gerecht zu werden. In den folgenden Kapiteln beschreiben wir ihre Funktionsweise in der Praxis.

² support.google.com/websearch/answer/106230

Auch bei der Bewältigung der COVID-19-Pandemie im Jahr 2020 haben diese vier Säulen sich als robuster Rahmen erwiesen – eine Zusammenfassung unserer Maßnahmen gegen Fehlinformationen zum Coronavirus beschreiben wir am Ende dieses Whitepapers.

Entfernen: Entwicklung und Durchsetzung unserer Richtlinien

Eine Säule, mit der wir die Informationsqualität unserer Produkte und Dienste sicherstellen, besteht in der vollständigen Entfernung bestimmter Inhalte von einer Plattform. Inhalte können aus zwei Gründen entfernt werden: wenn sie gesetzwidrig sind oder wenn sie gegen unsere Richtlinien für ein Produkt oder einen Dienst verstoßen.

Wir halten uns an die Gesetze jedes Landes, in dem wir aktiv sind, und sperren Inhalte für Länder, wenn diese in dem betreffenden Land illegal sind. Jedes Land hat nicht nur seine eigene Kultur und Geschichte, sondern auch eigene Gesetze, die regeln, welche Meinungsäußerungen zulässig sind und welche nicht. So steht in Deutschland, Frankreich und Österreich z. B. die Leugnung des Holocausts unter Strafe. Während manche Länder umfassende Gesetze zum Schutz gegen Verleumdung erlassen haben, sind andere hier weniger restriktiv. Ein weiteres Beispiel sind die in der Europäischen Union und Russland geltenden Regelungen zum Datenschutz, die Individuen ein „Recht auf Vergessenwerden“ einräumen. Suchmaschinenbetreiber (wie Google) sind demnach verpflichtet, auf eine Person bezogene Suchergebnisse unter bestimmten Voraussetzungen zu entfernen.

Ergänzend hierzu definieren wir Richtlinien, die genau festlegen, welche Inhalte und Verhaltensweisen für ein Produkt oder einen Dienst zulässig sind und welche nicht. Diese auch als Content- oder Community-Richtlinien bezeichneten Regeln werden von uns fortlaufend weiterentwickelt. Sie sind Bestandteil der YouTube Nutzungsbedingungen, als solche klar und eindeutig formuliert und für alle Nutzerinnen und Nutzer sowie Creator leicht zugänglich, unabhängig davon, ob es sich um Video-Creator, Webmaster, App-Entwickler oder Werbetreibende handelt. Die Richtlinien benennen Aufgabe und Zweck eines bestimmten Produkts oder Dienstes und bilden

somit einen wesentlichen Bestandteil der jeweiligen Besonderheit des Produkts. Sie erläutern auch, welche Inhalte und Verhaltensweisen nicht zulässig sind und wie der Prozess aussieht, mit dem ein Inhalt entfernt oder die Person, die diesen erstellt hat, von einem Dienst ausgeschlossen wird. Die folgenden Abschnitte machen deutlich, wie wir unsere Regeln und Richtlinien entwickeln und durchsetzen.

Umfassende Entfernung von Inhalten

Wir setzen unsere Richtlinien für Inhalte in großem Maßstab durch. Hierfür ergreifen wir täglich Millionen von Maßnahmen gegen Inhalte, die die Richtlinien für eines oder mehrere unserer Produkte verletzen.

- 2020 wurden **über 34 Millionen Videos** von YouTube entfernt, da sie gegen unsere Community-Richtlinien verstoßen haben.³
- 2020 blockierten und entfernten wir außerdem mehr als **3,1 Milliarden unzulässige Werbeanzeigen** („Bad Ads“) aus unseren Systemen und sanktionierten über **1,7 Millionen** Konten der betreffenden Werbetreibenden. Werbeanzeigen von mehr als **1,3 Milliarden Webseiten** unseres Publisher-Netzwerks wurden wegen Verletzung der Richtlinien entfernt.⁴
- Die Google Play-Richtlinien für Entwicklerinnen und Entwickler⁵ untersagen zahlreiche Arten von irreführenden Verhaltensweisen und Inhalten. Dies gilt insbesondere für die Verbreitung von Anwendungen aus den Bereichen Medizin und Gesundheit. Wenn Entwicklerinnen und Entwickler diese Richtlinien verletzen, werden ihre Apps aus dem Google Play Store entfernt. 2019 verhinderte Google Play zudem die Veröffentlichung von über **790.000** nicht regelkonformen Apps in seinem Play Store.

³ transparencyreport.google.com/youtube-policy/removals

⁴ blog.google/products/ads-commerce/ads-safety-report-2020

⁵ play.google.com/about/developer-content-policy

- 2019 identifizierte und entfernte Google Maps mehr als **75 Millionen** unzulässige Rezensionen und **4 Millionen** gefälschte Unternehmensprofile. Google Maps entfernte zudem über **580.000** Rezensionen und **258.000** Firmeneinträge, für die uns Regelverstöße gemeldet wurden. Darüber hinaus überprüften und entfernten wir mehr als **10 Millionen** Fotos und **3 Millionen** Videos in Google Maps wegen Verletzung der Content-Richtlinien und deaktivierten mehr als **475.000** Konten wegen missbräuchlicher Nutzung.

Entwicklung von Richtlinien und Schutzvorkehrungen

Mit den Richtlinien für unsere Produkte und Dienste wollen wir einerseits Schaden von unseren Nutzerinnen und Nutzern abwenden, andererseits aber auch den zweckmäßigen Gebrauch unserer Produkte ermöglichen. Die Richtlinien für unsere Produkte und Dienste sind so formuliert, dass sie eine angemessene Balance zwischen dem freien Zugang zu einer Vielfalt von Äußerungen und der Beschränkung von schädlichen Inhalten und Verhaltensweisen wahren.

Diese Balance kann schon deswegen von Produkt zu Produkt variieren, weil sich problematische Inhalte in verschiedenen Diensten und Kontexten unterschiedlich auswirken. Auch etwas, das übergreifend als "schädlich" angesehen wird und entsprechend in allen Produkten und Diensten verboten ist, kann sich in jedem Produkt oder Dienst in anderer Form äußern. Aus diesem Grund müssen wir für jedes Produkt individuell bestimmen, welche Schutzvorkehrungen erforderlich sind, und unsere Richtlinien entsprechend gestalten. Dies gilt hinsichtlich der Folgen für Einzelpersonen, aber auch mit Blick auf mögliche Auswirkungen auf die Gesellschaft als Ganzes, z. B. durch die versuchte Beeinflussung von Wahlen oder gesellschaftlichen Prozessen.

Unsere Richtlinien für Produkte und Dienste beachten dabei folgende Aspekte:

- **Ermutigung zu schädlichem oder gefährlichem Verhalten:**
Hierunter fallen Inhalte, die besonders schädliche oder gefährliche

Verhaltensweisen zeigen oder Nutzerinnen und Nutzer zu einem solchen Verhalten ermuntern.

- **Hassrede:** Hierbei handelt es sich um Inhalte, die Gewalt gegen Einzelne oder Gruppen unterstützen oder befürworten und sich dabei auf Merkmale wie z. B. ethnische Herkunft, Geschlechtsidentität, sexuelle Orientierung oder religiöse Überzeugung beziehen.
- **Bedrohung, Belästigung und Cybermobbing:** Dies sind Inhalte, die eine direkte Bedrohung anderer Personen, Erpressung oder die Veröffentlichung privater Daten beinhalten bzw. darauf abzielen, andere Personen zu belästigen oder zum Schweigen zu bringen.
- **Gewalttätige oder grausame Inhalte:** Der primäre Zweck dieser Inhalte besteht darin, zu schockieren, Sensationslust zu befriedigen und unnötig drastische, unangemessene oder beleidigende Darstellungen zu zeigen; auch Inhalte, die von einer Terrororganisation erstellt werden bzw. diese unterstützen, können in diese Kategorie fallen.
- **Nacktheit und pornografische Inhalte:** Hierbei handelt es sich um Beschreibungen oder bildliche Darstellungen von Nacktheit und expliziten sexuellen Handlungen.
- **Die Bedeutung von Kontext:** Kontext ist für alle Inhalte wichtig, insbesondere für extreme Inhalte. Diese können zulässig sein, wenn sie in einen pädagogischen oder wissenschaftlichen Kontext oder in Nachrichten oder Dokumentationen eingebunden sind.
- **Spam, Betrug und irreführende Praktiken:** In diese Kategorie fallen Aktivitäten, die darauf abzielen, unsere Produkte missbräuchlich zu nutzen, Vorkehrungen zum Schutz der Benutzerdaten zu umgehen, Bewertungssysteme zu manipulieren oder im weiteren Sinne ungültigen Traffic zu erzeugen, der nicht aus einem echten Nutzerinteresse herrührt.
- **Identitätsdiebstahl, Falschdarstellung und Betrug:** Hierzu zählen Aktivitäten, die die Identität einer Person, eines Unternehmensstandortes, des Landes, in dem ein Unternehmen tätig ist, oder den Verkauf von Waren und Diensten vortäuschen.

Entwicklung von Richtlinien

Um neuartige Gefahren und Schwachstellen in unseren Richtlinien zu identifizieren, greifen wir auf Expertenmeinungen, Nutzerfeedback und gesetzliche Vorgaben zurück. Dabei stützen wir uns auf die Forschungsarbeit von Analysten, die sich mit neuen Strategien böswilliger Akteure befassen, auf Trends bei anderen Plattformen sowie auf kulturelle Entwicklungen, die wir im Auge behalten müssen. Darüber hinaus stehen wir in ständigem Dialog mit Gesetzgebern in aller Welt. Ihre Sichtweisen und Bedenken finden unmittelbaren Niederschlag bei der Weiterentwicklung unserer Richtlinien.



In einem weiteren Schritt sammeln wir möglichst viele Beispiele dafür, wie sich schädliche Inhalte in unseren Diensten darstellen hat oder wie sie sich künftig zeigen könnten, und suchen nach Gemeinsamkeiten. Dabei betrachten wir auch Gegenbeispiele in Gestalt von Inhalten, die zwar ähnlich aussehen wie der schädliche Inhalt, bei genauerer Betrachtung jedoch unproblematisch oder von erheblichem öffentlichem Interesse sind. Dank dieser Methode können wir sowohl gemeinsame Merkmale bestimmen, die den Inhalt oder das Verhalten schädlich machen, als auch Risiken erkennen, die sich aus einer zu weit gefassten Richtlinie ergeben könnten.

Auf diese Weise entwickeln wir vorläufige Standards und Leitlinien für die Durchsetzung der Richtlinien, die wir dann anhand der Gegenbeispiele überprüfen, um falsche Einstufungen zu vermeiden. Parallel dazu beraten wir uns mit Expertinnen und Experten aus den verschiedenen Google-Fachgebieten sowie von außerhalb des Unternehmens, die ihre jeweils eigenen Sichtweisen einbringen. Im nächsten Schritt versuchen wir, etwaige Widersprüche zwischen den verschiedenen Einschätzungen aufzulösen, und formulieren aus den vorläufigen Standards und Leitlinien eine kohärente Richtlinie. Erst nach ausführlichen Tests, mit denen wir ein

hohes Maß an einheitlicher Anwendung sicherstellen, folgt die nächste Phase der tatsächlichen Richtlinienanwendung.

Diesen Prozess der Beurteilung und Verfeinerung wiederholen wir so lange, bis eine verständliche, vorhersagbare und wiederholbare Vorgehensweise erreicht ist. Alle Nutzerinnen und Nutzer oder Creator sollen bei einer Änderung unserer Richtlinien verstehen können, worum es sich handelt (Verständlichkeit). Sie sollen zudem feststellen können, ob ihre Inhalte oder ihre Verhaltensweisen von der geänderten Regelung betroffen sind (Vorhersagbarkeit). Die neue Regelung sollte außerdem so allgemein formuliert sein, dass sie einheitlich auf viele voneinander unabhängige Fälle anwendbar ist (Wiederholbarkeit).

Vor der endgültigen Einführung und Durchsetzung der neuen Richtlinie ermitteln wir, ob sie die betreffenden Inhalte gezielt bekämpft und wie sich die Änderung auf Nutzerinnen und Nutzer auswirkt. Wir überlegen außerdem, wie wir über die Änderung informieren können, und richten die notwendigen Mechanismen für ihre Durchsetzung ein.

All dies ist ein sehr zeitaufwendiger Prozess. Es kann Monate dauern, bis eine Richtlinie so klar und robust ist, dass wir sie anwenden und durchsetzen. Für dieses Verfahren, in das eine Fülle an Fachwissen von innerhalb und außerhalb des Unternehmens einfließt, sind in der Regel unsere Trust & Safety Teams zuständig. Ihre Mitglieder bekämpfen Online-Missbrauch und -Betrug, indem sie die Richtlinien entwickeln und durchsetzen, die unsere Produkte sicher und zuverlässig machen. In ihren Reihen finden sich Produktexperten, Ingenieure, Rechtsanwälte, Datenwissenschaftler und viele andere, die weltweit zusammenarbeiten und dabei ein Netzwerk von internen und externen Spezialisten für Sicherheits- und anderweitige Fachfragen nutzen.

Wir sind uns darüber im Klaren: Es wird immer Personen geben, die nicht mit den Entscheidungen einverstanden sind, die wir treffen, um die richtige Balance zwischen der Vermeidung von Schaden und dem Recht auf freien Informationszugang zu finden. Eine einfache, verbindliche Antwort auf diese Fragen zu finden, ist oft kaum möglich, und über den angemessenen Umgang mit Inhalten kann man verschiedener Meinung sein. Dies gilt besonders angesichts des ständigen Wandels, der für das Internet und damit für die Nutzungserwartungen typisch ist. Wir sind überzeugt: Inklusive Prozesse, transparente Vorgehensweisen und die Bereitschaft, unsere

Richtlinien immer wieder auf den Prüfstand zu stellen, werden den Interessen unserer Nutzerinnen und Nutzer sowie der Gesellschaften, in denen wir aktiv sind, am besten gerecht.

Entwicklung einer Richtlinie zum Verbot von spekulativen und experimentellen medizinischen Behandlungsformen in Google Ads

Bei der Entwicklung unserer Richtlinien berücksichtigen wir zahlreiche Gesichtspunkte und Erwägungen. Ein Beispiel hierfür ist die Überarbeitung der Google Ads-Richtlinie im September 2019, mit der wir die Bewerbung von spekulativen und experimentellen medizinischen Behandlungen wie der Stammzellen- oder Gentherapie untersagt haben.⁶

Auch wenn wir wissen, dass bahnbrechende medizinische Erkenntnisse oft als unbewiesene Ideen ihren Anfang nehmen, sind wir dem Schutz und der Sicherheit unserer Nutzerinnen und Nutzer verpflichtet. Wie wichtig dies ist, zeigt das Beispiel von Therapien, die in einigen Ländern bereits zugelassen worden waren, während wir zeitgleich vermehrt Angebote für ungeprüfte und irreführende Behandlungsformen feststellen mussten. Hier bestand die Gefahr, dass Menschen erhebliche Summen für wissenschaftlich nicht fundierte Therapien ausgeben, die keinen medizinischen Nutzen haben oder gar ernsthafte Gesundheitsschäden verursachen können. Nach ausgiebiger Beratung mit externen Expertinnen und Experten untersagen wir die Bewerbung solcher Behandlungen. Damit verhindern wir den Missbrauch unserer Werbepattformen für Zwecke, die unseren Nutzerinnen und Nutzern ernsthafte finanzielle und gesundheitliche Schäden hätten zufügen können.

⁶ support.google.com/google-ads/answer/9475042

Proaktive Erkennung & Durchsetzung

Zahlen und Fakten

Zwischen Oktober und Dezember 2020:

- wurden **über 9,3 Millionen Videos** von YouTube entfernt, die gegen unsere Community-Richtlinien verstoßen haben.
- wurden **94 %** dieser Videos zuerst von unseren automatischen Systemen gemeldet. Von den maschinell erkannten Videos wurden **36 %** überhaupt nicht und ca. **72%** weniger als zehn Mal aufgerufen.
- wurden mehr als **906 Millionen** Kommentare auf YouTube entfernt, wobei es sich mehrheitlich um Spam handelte. **99,6 %** der gelöschten Kommentare wurden automatisch erkannt.

Fallstudie: Erkennung von extremistischen Inhalten auf YouTube durch maschinelles Lernen (ML)

- Im Juni 2017 setzten wir erstmals ML-Technologien ein, um extremistische Inhalte auf YouTube zu identifizieren. Dabei sichteten unsere Teams **mehr als 2 Millionen Inhalte**, mit denen unsere ML-Klassifikatoren trainiert wurden.
- Im **ersten Quartal 2017** verzeichneten **8 %** der Videos, die aufgrund einer Verletzung unserer Richtlinie zu gewalttätigem Extremismus gelöscht wurden, weniger als zehn Aufrufe.
- Im **ersten Quartal 2018** lag dieser Wert bei über **50 %**.
- Im **vierten Quartal 2019** wurden ca. **90 %** der hochgeladenen Videos mit extremistischen Inhalten entfernt, bevor sie zehn Mal aufgerufen wurden.

Erhöhte Erkennung durch maschinelles Lernen



Prozentsatz der auf YouTube entfernten Videos im Zusammenhang mit gewalttätigem Extremismus bevor sie 10 Aufrufe erzielten

Zur Durchsetzung unserer Richtlinien im großen Maßstab nutzen wir sowohl maschinelle als auch menschliche Ressourcen, um problematische Inhalte zu identifizieren. Neben der manuellen Meldung durch Nutzerinnen und Nutzer spielen automatisierte Technologien eine wichtige Rolle bei der großflächigen Erkennung dieser Inhalte. Unsere automatisierten Systeme werden sorgfältig trainiert, um Spam und unzulässige Inhalte schnell identifizieren und entsprechende Maßnahmen einleiten zu können. Das schließt die Meldung von potenziell problematischen Inhalten ein, die dann durch menschliche Prüferinnen und Prüfer untersucht werden. Das Urteilsvermögen dieser Personen wird für die vielen Entscheidungen benötigt, die eine nuancierte Betrachtung erfordern. Bei der Beurteilung von Qualität oder Zweck eines Inhalts kommt es sehr auf den Kontext an, in dem er erzeugt oder geteilt wird. Besonders relevant ist dies für Inhalte, die in Zusammenhang mit Bildung und Wissenschaft sowie zu künstlerischen, dokumentarischen oder journalistischen Zwecken verwendet werden, die aber in anderem Kontext gegen unsere Richtlinien verstoßen würden.

Zusätzlich befassen sich unsere weltweiten Expertenteams mit den Strategien böswilliger Akteurinnen und Akteure, die alles daran setzen, die automatisierten Abwehrmechanismen unserer Produkte zu überwinden. Zur Bekämpfung der fortlaufend neu entstehenden Formen von missbräuchlicher Nutzung und Bedrohung braucht es den menschlichen Geist, um eine erste Einschätzung vorzunehmen und Maßnahmen zu entwerfen. Erst dann kommt ein automatisiertes System zum Zuge, das Angriffe großflächig abwehren kann. Zudem beobachten und analysieren

unsere Teams wie beispielsweise die Google Threat Analysis Group⁷ Bedrohungen und gezielte Angriffe jeglicher Art und liefern die notwendigen Erkenntnisse, damit wir bei der Richtlinienentwicklung- und Durchsetzung immer einen Schritt voraus sein können.

In den vergangenen zwei Jahrzehnten haben wir erhebliche Summen in die umfassende Gefahrenerkennung und das Durchsetzen der Richtlinien investiert und unseren Ansatz immer weiter verfeinert. Zugleich bedeutet die Offenheit und Reichweite unserer Produkte und des Internets, kombiniert mit dem Erfindungsreichtum und der Hartnäckigkeit böswilliger Akteurinnen und Akteure, dass es schlicht unmöglich ist, alle problematischen Inhalte und Aktivitäten zu erkennen. Um mehr solcher Inhalte schneller zu identifizieren, könnten wir noch stärker auf automatisierte Systeme setzen. Ein solcher Ansatz hat allerdings auch Nachteile.

Durch die vermehrte Nutzung von automatisierten Systemen steigt die Gefahr von fehlerhaften Entscheidungen. In der Folge könnten Inhalte entfernt werden, die de facto nicht gegen unsere Richtlinien verstoßen. Dies könnte z. B. die Diversität der Nutzerstimmen und -meinungen beeinträchtigen oder Inhalte treffen, die von hohem Interesse für die Öffentlichkeit sind. Die Löschung könnte darüber hinaus bei Creatoren und anderen Partnerinnen und Partnern für Verwirrung sorgen, denn diese sind auf klare und eindeutige Richtlinien angewiesen.

Wenn wir andererseits unsere Trust and Safety Teams mit der Überprüfung im Grunde harmloser Inhalte überfrachten, besteht die Gefahr, dass raffinierte Akteurinnen und Akteure unbemerkt bleiben und mehr Raum für schädliche Handlungen haben. Umgekehrt hat auch die Beschränkung auf ein engeres Gefahrenspektrum ihre Tücken. Wir könnten das große Ganze aus dem Blick verlieren und so riskieren, dass blinde Flecken entstehen, die von Dritten ausgenutzt werden, um Schaden zu verursachen.

Auch wenn es keine einfache Lösung für dieses Dilemma gibt, wissen wir um die Verantwortung, die eine Tätigkeit dieser Größenordnung mit sich bringt. Eine einzige falsche Entscheidung kann erhebliche Auswirkungen für unsere Nutzerinnen und Nutzer, Entwicklerinnen und Entwickler, Creator und Werbetreibenden haben. All dies spornt uns an, unsere Tools und Ressourcen

⁷ blog.google/threat-analysis-group

für die Gefahrenerkennung und das Durchsetzen der Richtlinien kontinuierlich weiterzuentwickeln.

Als Ergänzung zu unseren eigenen Anstrengungen ermöglichen wir Nutzerinnen und Nutzern und vertrauenswürdigen Organisationen die Meldung von potenziell problematischen Inhalten. Solche manuell gemeldeten Inhalte werden von unserem Trust & Safety Team überprüft, um sicherzustellen, dass tatsächlich Handlungsbedarf besteht.

Neben einzelnen Inhalten und Verhaltensweisen sanktionieren wir auch die betreffenden Akteurinnen und Akteure durch zielgerichtete Maßnahmen. Bei besonders schwerwiegenden und wiederholten Richtlinienverletzungen werden wir auf Nutzerkonto-Ebene aktiv. Damit greifen wir das Problem an der Wurzel an und sorgen für einen effektiven Schutz unserer Nutzerinnen und Nutzer. In unseren “Bad Ads” Reports erläutern wir beispielsweise, wie wir mit Machine Learning-Technologien die Konten von Werbetreibenden identifizieren und entfernen, die sich nicht an unsere Richtlinien halten.⁸

Transparente Durchsetzung von Richtlinien

Unsere Richtlinien funktionieren am besten, wenn die Nutzenden sie kennen und wissen, wie wir sie durchsetzen. Deshalb arbeiten wir daran, diese Informationen möglichst gut verständlich und für alle verfügbar zu machen.

Unsere umfangreichen Hilfsangebote, Webseiten mit Community-Richtlinien und Blogs informieren über die konkreten Bestimmungen unserer Richtlinien. Ergänzend dazu veröffentlichen wir regelmäßig Berichte, in denen wir über die Durchsetzung der Richtlinien und die Überprüfung von Inhalten, die lokale Gesetze und Vorschriften verletzen, Auskunft geben.

- Der **Transparenzbericht “YouTube-Community-Richtlinien und ihre Anwendung”** enthält vierteljährliche Updates zur Anzahl der entfernten Videos, Kanäle und Kommentare. Er nennt außerdem die einzelnen Richtlinien, auf deren Basis die Inhalte entfernt wurden. Der Bericht beschreibt im Detail, wie wir unzulässige Videos identifizieren (z. B. mit automatisierten Systemen oder über Nutzermeldungen) und wie viele dieser Videos entfernt werden konnten, ohne dass sie zuvor Aufrufe erzielt haben.⁹

⁸ Report 2021: blog.google/products/ads-commerce/ads-safety-report-2020

⁹ transparencyreport.google.com/youtube-policy/removals

- Unser jährlicher **“Bad Ads” Report** gibt Einblicke in die Maßnahmen, mit denen wir unsere Werberichtlinien durchsetzen. Dazu zählt die Anzahl der entfernten Anzeigen, die Anzahl der Seiten, auf denen Werbung blockiert wurde, sowie die Anzahl der Nutzerkonten von Werbetreibenden und Publishern, die im Berichtszeitraum gelöscht wurden. Der Bericht gibt außerdem an, wie oft wir unsere Richtlinien im Jahresverlauf überarbeitet haben.¹⁰
- Der **vierteljährlich erscheinende Bericht (Quarterly Coordinated Influence Operations Bulletin)**¹¹ der Threat Analysis Group (TAG) schlüsselt auf, wie wir gegen Nutzerkonten vorgehen, die wir koordinierter Einflussnahme durch inländische oder ausländische Kampagnen zuordnen können.
- Im **Google Transparenzbericht** veröffentlichen wir Informationen zu behördlichen Ersuchen um Entfernung von Inhalten aus unseren Diensten. Wir legen außerdem dar, wie sich die Maßnahmen von Behörden und Unternehmen auf Datenschutz, Sicherheit und den Online-Zugang zu Informationen auswirken.¹²
- Als weitere Ressource stellen wir einen öffentlich zugänglichen, durchsuchbaren und herunterladbaren **Transparenzbericht zu Inhalt und Finanzierung von Wahlwerbung** auf unseren Plattformen bereit.¹³ Mit diesem Bericht wollen wir das Vertrauen der Wählerinnen und Wähler in die politische Werbung auf unseren Werbepattformen stärken — auch und gerade angesichts aktuellerer Debatten rund um politische Werbung und der Bedeutung von Vertrauen in demokratische Prozesse.

Wir werden unsere Bemühungen um maximale Transparenz fortsetzen. Sie sind ein wichtiger Beitrag zu einem informierten öffentlichen Dialog über die Rolle, die unsere Dienste in der Gesellschaft spielen.

Beschwerde gegen Durchsetzungsmaßnahmen

Bei der Durchsetzung unserer Richtlinien können uns Fehler unterlaufen, in deren Folge Inhalte irrtümlich aus unseren Diensten entfernt werden. Um hier vorzubeugen, informieren wir Creator wo immer möglich über die von uns

¹⁰ blog.google/products/ads-commerce/ads-safety-report-2020

¹¹ Siehe Veröffentlichungen im Blog der Threat Analysis Group: blog.google/threat-analysis-group

¹² transparencyreport.google.com

¹³ transparencyreport.google.com/political-ads/home

ergriffenen Maßnahmen und geben ihnen Gelegenheit, Beschwerde einzulegen und zusätzliche Informationen einzureichen. Die Entscheidung wird dann von einem anderen Mitglied unseres Trust and Safety Teams geprüft und entweder bestätigt oder zurückgenommen.

Beschwerde im Rahmen unserer Richtlinie zu Falschdarstellung bei Werbung

Wir prüfen Werbeschaltungen mit einer Kombination aus automatischen und manuellen Prozessen. Um die Einhaltung der Richtlinien sicherzustellen, führen wir von Zeit zu Zeit Nachprüfungen durch. Mitunter wird das Geschäftsmodell eines Werbetreibenden bei der Erstprüfung als Regelverstoß eingestuft. Dies kann der Fall sein, wenn wir die Einzelheiten der angebotenen Dienste oder Produkte nicht abschließend klären können. Wenn der Werbetreibende Beschwerde gegen unsere Entscheidung einlegt und uns weitere Informationen zu seinen Geschäftspraktiken bereitstellt, kann sich diese Einstufung nachträglich ändern. So verhindert unsere Richtlinie zu Falschdarstellung¹⁴ beispielsweise, dass Nutzerinnen und Nutzer durch irreführende Werbeanzeigen getäuscht werden. Wenn sich die behauptete Beziehung zu einem anderen Unternehmen nicht direkt aus der Landingpage des Werbetreibenden ableiten lässt, können die betreffenden Anzeigen zunächst blockiert werden. Der Wahrheitsgehalt der Behauptung kann dann durch zusätzliche Angaben nachgewiesen werden. In der Praxis heißt das, dass uns Werbetreibende z. B. zusätzliche Informationen zu ihrem Geschäftsmodell und ihrer Dienstleistungskette zur Verfügung stellen. Dies kann z. B. ein offizieller Nachweis ihrer Beziehung zu einem anderen Unternehmen oder ihre offizielle Teilnahme an einem Partnernetzwerk sein.

Gutwillige Akteurinnen und Akteure sollen unsere Regeln möglichst leicht verstehen und einhalten können. Den böswilligen wollen wir es dagegen möglichst schwer machen, unsere Regeln ungestraft zu übertreten. Aus diesem Grund sehen wir bei der Richtliniendurchsetzung einen gewissen Spielraum vor, wenn es sich um irrtümlich entstandene Verstöße handelt.

¹⁴ <https://support.google.com/adspolicy/answer/6020955>

Jede und jeder von uns kann einen Scherz zu weit treiben oder nicht sofort erkennen, dass etwas Problematisches getan oder weitergegeben wurde. Wenn z. B. eine einzelne App gegen unsere Richtlinien verstößt, sanktionieren wir normalerweise diese App und nicht das Konto des Entwicklers.

Wenn es sich hingegen um schwerwiegende, mehrfache oder betrügerische Verstöße handelt, können wir Maßnahmen gegen die Website, den Kanal oder die App als Ganzes ergreifen. In besonders schwerwiegenden Fällen sperren wir Nutzerkonten.

Auf **YouTube** können bestimmte Verletzungen unserer Community-Richtlinien zu einer Verwarnung führen. Ein Creator kann dann während einer Woche keine Inhalte auf der Plattform veröffentlichen oder erzeugen. Wenn es innerhalb von 90 Tagen zu einer weiteren Verwarnung kommt, wird die Sperre auf zwei Wochen verlängert. Bei einer dritten Verwarnung wird der YouTube-Kanal gelöscht. Creator können Beschwerde gegen unsere Entscheidungen einlegen.

Zwischen Oktober und Dezember 2020 wurden **über 2 Millionen Kanäle** von YouTube entfernt, da sie gegen unsere Community-Richtlinien verstoßen haben.

Seelisches und körperliches Wohlbefinden der Prüfteams unterstützen

Bei der Sicherstellung einer angemessenen Informationsqualität und dem Umgang mit Inhalten kommen automatische Systeme und Technologien zum Einsatz, aber die Prüfung durch Menschen spielt eine entscheidende Rolle. Unsere Prüfteams führen jährlich Milliarden von Sichtungen durch. Damit schaffen sie die Grundlage für fundierte Durchsetzungsentscheidungen und helfen dabei, Trainingsdaten für unsere Machine Learning-Modelle zu gewinnen.

Obwohl die moderierten Inhalte überwiegend weder gewalttätig noch schockierend sind, können sie verstörend und verletzend wirken. Manche Prüferinnen und Prüfer arbeiten bewusst in besonders schwierigen Bereichen, da es ihnen ein Anliegen ist, die betreffenden Inhalte zu finden und aus dem Web zu entfernen.

Um ihnen die Arbeit zu erleichtern, erhalten die Prüfteams vor allem bei extrem belastenden Aufgaben technische Hilfsmittel. Automatisierte Meldesysteme ermöglichen uns die schnelle und präzise Erkennung und Entfernung von Inhalten. Damit entlasten wir die Menschen in den Prüfteams und verkürzen die Zeit bis zur Entfernung problematischer Inhalte. So wurden **über 94 %** der Videos, die wir im vierten Quartal 2020 wegen einer Verletzung der Community-Richtlinien von YouTube entfernt haben, zuerst durch automatisierte Systeme erkannt und gemeldet. Wir arbeiten ständig daran, diese Systeme noch besser und genauer zu machen.

Die Menschen, die problematische Inhalte überprüfen, leisten einen entscheidenden Beitrag dazu, dass digitale Plattformen für jede und jeden von uns sicherer werden. Dabei kann ihre Arbeit ausgesprochen kompliziert und emotional schwierig sein. Google fördert das Wohlbefinden dieser Prüferinnen und Prüfer durch ein umfassendes Unterstützungsprogramm, die konsequente Durchsetzung seiner Standards bei allen Partnerunternehmen sowie durch wissenschaftliche Forschung und technologische Innovation. Damit stellen wir sicher, dass unsere Prüfteams über die Ressourcen verfügen, die sie für ihre seelische und körperliche Gesundheit brauchen.

Maßnahmen für seelisches und körperliches Wohlbefinden

Unsere Prüfteams bewerten den Kontext und die Nuancen unbekannter Inhalte, sie nehmen Unterscheidungen vor und fällen Urteile. Es ist unsere Verpflichtung, diesen Personen größtmögliche Unterstützung zu bieten, und haben hierfür erhebliche Summen investiert.

Konkret setzen wir auf folgende Maßnahmen:

- Prüferinnen und Prüfer mit entsprechendem Bedarf erhalten Zugang zu einzel- oder gruppentherapeutischen Angeboten. Sie können außerdem spezielle Ruhebereiche nutzen und am Arbeitsplatz oder telefonisch eine zu jeder Tages- und Nachtzeit verfügbare individuelle psychologische Betreuung in Anspruch nehmen.
- Auf Wunsch können Prüferinnen und Prüfer vom Umgang mit besonders schrecklichen Inhalten freigestellt werden.
- Wir bieten sportliche Aktivitäten an (mit festem Plan oder spontan wählbar).
- Alle Standorte müssen über jederzeit zugängliche Ruheräume und Gemeinschaftsbereiche verfügen.

Einhaltung des Verhaltenskodex sicherstellen

Wir arbeiten mit externen Anbietern und Auftragnehmern zusammen, um unsere Richtliniendurchsetzung in vollem Umfang, rund um die Uhr und mit der gebotenen Expertise in zahlreichen Landessprachen zu gewährleisten; so, wie es von einer globalen Plattform erwartet wird. Bei der Zusammenarbeit mit diesen Firmen führen wir regelmäßige Standortinspektionen und Audits durch, um sicherzustellen, dass unsere Richtlinien und unser Verhaltenskodex für Lieferanten eingehalten werden. Durch Einzelgespräche und Fokusgruppen mit Prüferinnen und Prüfern erhalten wir dabei ein unmittelbares und vertrauliches Feedback. Alle externen Partner stellen ihren Mitarbeitenden sowohl Beschwerde- und Schlichtungsmechanismen als auch Zugang zu einer Vertrauensperson bereit. Die Mitarbeiterinnen und Mitarbeiter unserer externen Anbieter können sich mit ihren Anliegen zudem an dieselben Beratungsstellen wie unsere eigene Belegschaft wenden und dort auch anonym Meldung erstatten.

Wissenschaftliche Forschung & technologische Innovation

Neben dem direkten Feedback von Mitarbeitenden und der Nutzung von professionellem Rat und Input setzen wir bei der Durchsetzung der Richtlinien auch auf aktuelle Forschung und technologische Neuerungen. Ein Beispiel hierfür ist ein 2019 veröffentlichter Forschungsbericht. Er befasst sich damit, wie

durch die Darstellung von Bildern in Graustufen ("Grayscale Transformation") deren emotionale Wucht verringert werden kann. Prüferinnen und Prüfer, die im Rahmen ihrer Tätigkeit mit gewalttätigen und extremistischen Inhalten konfrontiert werden, fühlen sich demnach deutlich weniger belastet, wenn die Bilder nicht in Farbe, sondern in Graustufen angezeigt werden. Aufgrund dieser Erkenntnis haben wir eine Graustufen-Funktion in unsere Review-Programme integriert. Da jeder Mensch anders reagiert, lässt sich diese Funktion nach Wunsch aktivieren oder deaktivieren und ermöglicht so mehr Flexibilität bei der Überprüfung von Inhalten. Mittlerweile arbeiten 70 % der Prüferinnen und Prüfer, die gewalttätige und extremistische Inhalte auf Google Drive, Google Photos usw. zu sehen bekommen, dauerhaft mit aktivierter Graustufen-Anzeige. Aufgrund dieses Erfolgs werden wir die Funktion nun auch an anderen Stellen einführen.

Allerdings ist auch die Graustufen-Anzeige kein Allheilmittel und funktioniert nicht für alle Prüferinnen und Prüfer und alle Arten von belastenden Inhalten. Aus diesem Grund suchen wir nach alternativen Wegen zur Lösung des Problems. Ein möglicher Ansatz bestand z. B. darin, bestimmte Inhalte unscharf darzustellen. In der Praxis fühlten sich viele Prüferinnen und Prüfer jedoch von der Unschärfe irritiert oder klagten gar über Schwindel, so dass sie die Option deaktivierten. Gegenwärtig experimentieren wir mit einem Schieberegler, mit dem sich der Unschärfegrad flexibel einstellen lässt. Ergänzt wird dies durch eine Mouseover-Funktion, die den Bildausschnitt scharf stellt, über den der Mauszeiger geführt wird. Die ersten Resultate sind vielversprechend und könnten bei einigen Personen gut funktionieren.

Um die langfristigen emotionalen Auswirkungen dieser Tätigkeit zu verstehen, sind jedoch weitere Maßnahmen erforderlich. Wir setzen unsere Forschungsarbeit fort und werden unsere Erkenntnisse weiterhin veröffentlichen und eng mit anderen Unternehmen der Branche kooperieren.

Diese Überprüfung von Inhalten ist ein vergleichsweise junges Tätigkeitsfeld mit einer stark steigenden Zahl von Beschäftigten. Dazu gehören auch unsere Anti-Missbrauchs-Teams. Der Ausbau dieses Bereichs ist eine wichtige Maßnahme in unserem Kampf gegen böswillige Akteurinnen und Akteure und für den Schutz unserer Nutzerinnen und Nutzer vor schädlichen Online-Inhalten. Unsere Verantwortung erstreckt sich jedoch auch auf unser wachsendes Anti-Missbrauchs-Team, das im Interesse unserer Nutzerinnen und Nutzer gegen Missbrauch vorgeht. Wir werden diese Maßnahmen mit vollem Einsatz weiterführen.

Sichtbar machen: Zuverlässige Inhalte für unsere Nutzerinnen und Nutzer

Ob Google Suche, YouTube, Google Play, Google Maps oder andere Dienste: Unsere Produkte erfüllen die Anforderungen der Nutzenden, indem sie gigantische Mengen an Informationen durchsuchen. Diese Informationen stammen von etablierten Publishern, neu hinzukommenden Creatoren und Einzelpersonen, die Online-Inhalte in Form von Kommentaren, Rezensionen oder Beiträgen in Foren und sozialen Medien beisteuern. Wir verwenden Algorithmen, um diese Inhalte so zu organisieren, wie dies nach unserem Verständnis der Absicht und den Wünschen unserer Nutzerinnen und Nutzer entspricht.

- Der Google-Suchindex umfasst **mehr als 100 Millionen Gigabyte** an Daten und erfasst Hunderte Milliarden von Webseiten. In Buchform ergäbe dies einen Stapel, der zwölf Mal der Entfernung Erde – Mond und zurück entspricht.
- Weltweit gibt es täglich Milliarden von Suchanfragen, und **15% aller Suchanfragen** sind uns neu und in dieser Form noch nie dagewesen.
- In jeder Minute werden über **500 Stunden Videomaterial** auf YouTube hochgeladen.

Um die Nützlichkeit eines Inhalts beurteilen zu können, müssen wir zunächst die Absicht der Nutzerin bzw. des Nutzers ermitteln. Sie lässt sich z. B. daran erkennen, dass in das Eingabefeld der Google Suche ein bestimmter Begriff

getippt wurde, oder daran, dass jemand ein Fußballspiel auf YouTube anschaut und dadurch indirekt Interesse für Videos über Fußball oder andere Sportarten signalisiert.

In der Regel kommen mehrere Inhalte in Frage, um der Absicht eines Nutzers oder einer Nutzerin zu entsprechen. Aus diesem Grund betrachten wir eine Reihe weiterer Faktoren, um diese Inhalte in eine Rangordnung zu bringen. Unsere Bewertungsalgorithmen suchen nach Hinweisen auf Sachkenntnis, Zuverlässigkeit und Vertrauenswürdigkeit (Expertise – Autoritativeness – Trustworthiness) der einzelnen Inhalte, um Ergebnisse, die für die Suchenden zu diesem Zeitpunkt am relevantesten sind, ganz oben in der Trefferliste anzuzeigen. Ein frühes und besonders bekanntes Beispiel für einen solchen Algorithmus ist PageRank. Die Wichtigkeit einer Webseite wird hier anhand der Links beurteilt, die von anderen Seiten auf sie verweisen.

Wir arbeiten kontinuierlich an der Verbesserung unserer Ranking-Systeme. 2019 führten wir bei der Google Suche mehr als 383.605 Prüfungen durch, um die Qualität der Suchergebnisse zu messen, und nahmen mehr als 3.600 Änderungen an den zugrundeliegenden Algorithmen vor.¹⁵

Damit unsere Suchalgorithmen höchsten Relevanz- und Qualitätsstandards entsprechen, wenden wir strenge Verfahren an, die sowohl Live-Tests als auch Test durch Tausende von speziell geschulten externen Evaluatoren, sogenannten Search Quality Ratern, rund um den Globus beinhalten.¹⁶ Unsere erstmals 2013 veröffentlichten Leitlinien für Evaluatoren (“Search Quality Rater Guidelines”) definieren die Ziele, die wir mit unseren Ranking-Systemen verfolgen. Sie benennen auch die Kriterien, anhand derer unsere Evaluatoren die Sachkompetenz, Zuverlässigkeit und Vertrauenswürdigkeit von Webseiten beurteilen.¹⁷ Anhand der Bewertungen dieser Prüferinnen und Prüfer können wir die Qualität unserer Suchergebnisse vergleichen und sicherstellen, dass für alle Nutzerinnen und Nutzer der Google-Suchmaschine weltweit gleichermaßen hohe Standards gelten.

Die Zuverlässigkeit oder wissenschaftliche Genauigkeit einer Seite ist nicht für alle Nutzungserfahrungen und jeden Kontext gleichermaßen wichtig.

¹⁵ Weitere Informationen zu diesen Tests, ihrer Durchführung und der Bewertung der Google Suchergebnisse finden Sie unter www.google.com/search/howsearchworks

¹⁶ www.google.com/search/howsearchworks/mission/users

¹⁷ Weitere Informationen zu unseren Search Quality Raters Guidelines finden Sie unter www.google.com/search/howsearchworks

Wenn eine Nutzerin oder ein Nutzer z. B. nach Unterhaltungsangeboten auf YouTube sucht oder sich damit beschäftigt, spielt die Zuverlässigkeit des Inhalts eine untergeordnete Rolle. Wenn es sich jedoch um Inhalte zu Themen wie persönlicher Lebensunterhalt, gesellschaftliche Teilhabe oder um Nachrichten geht, hat die Vertrauenswürdigkeit einen sehr viel höheren Stellenwert. Unzuverlässige Informationen könnten hier die Gesundheit, die finanzielle Lage, die künftigen Lebensumstände oder auch die Sicherheit eines Menschen beeinträchtigen. Für diese Themenbereiche haben wir daher eine eigene Kategorie definiert: „Your Money or Your Life“ (YMYL).¹⁸

Bei den YMYL-Themen gehen wir davon aus, dass die Nutzenden ein Höchstmaß an Vertrauenswürdigkeit und Sicherheit von uns erwarten. Wenn unsere Algorithmen erkennen, dass eine Suchanfrage ein YMYL-Thema betrifft, räumen die Ranking-Systeme bei der Trefferanzeige den von uns definierten Faktoren Fachlichkeit, Zuverlässigkeit und Vertrauenswürdigkeit automatisch ein höheres Gewicht ein. Wenn eine Nutzerin oder ein Nutzer z. B. nach einer bestimmten medizinischen Information oder Beratung sucht, versuchen wir, Inhalte von zuverlässigen Quellen wie Ärzten und medizinischen Organisationen anzuzeigen.

Sichtbar machen von zuverlässigen Informationen

Mit speziellen Funktionen oder Änderungen beim Ranking von Inhalten sorgen wir dafür, dass zuverlässige Informationen bevorzugt angezeigt werden. Hierzu zählen:

- **DIE BEREICHE TOP-NEWS UND EILMELDUNGEN AUF YOUTUBE:** Auf **YouTube** haben wir eine **Top-News-** und **Eilmeldungen-**Funktion eingeführt, die zuverlässige Nachrichten zu wichtigen Ereignissen an prominenter Stelle anzeigt. Die Top-News-Anzeige wird durch Suchanfragen nach Nachrichten ausgelöst und stellt Inhalte von Nachrichtenquellen bereit, die den Inhaltsrichtlinien für Google News¹⁹ entsprechen. Bereich Eilmeldungen erscheint automatisch auf der YouTube-Homepage, sobald in einem bestimmten Land

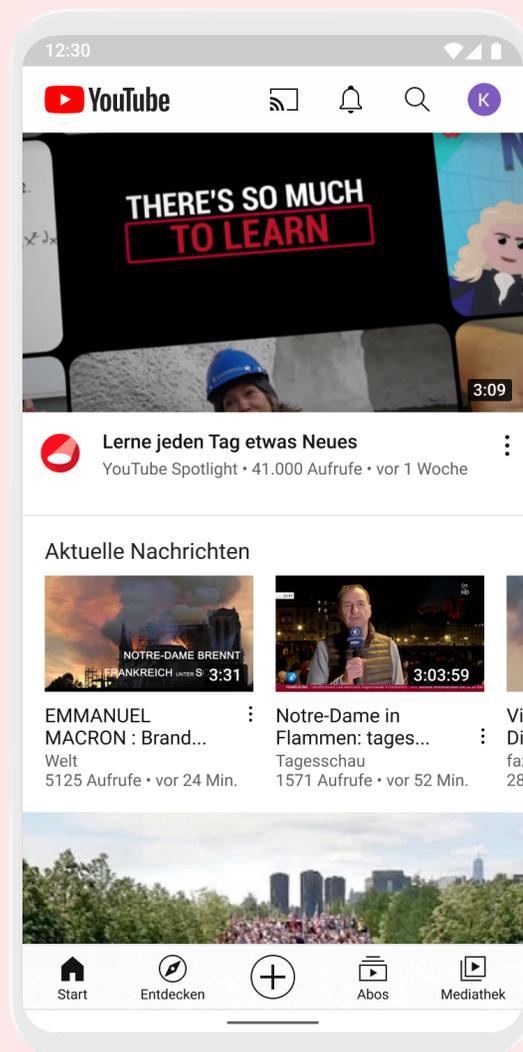
¹⁸ Siehe hierzu unsere Search Quality Rater Guidelines: static.googleusercontent.com/media/googleusercontent.com/en/searchqualityevaluatorguidelines.pdf

¹⁹ support.google.com/news/producer/answer/6204050

ein wichtiges Nachrichtereignis stattfindet. Auch diese Anzeige nutzt zuverlässige und überprüfte Nachrichtenquellen.

- Die Bereiche Top-News und Eilmeldungen ist gegenwärtig in mehr als **40 Ländern** verfügbar.
- 2019 verzeichneten die Nachrichtenkanäle von zuverlässigen YouTube-Partnern ein Nachfrageplus von **über 60%**.

Bereich Eilmeldungen



- **ZUSÄTZLICHE SCHUTZMASSNAHMEN BEI AKTUELLEN NACHRICHTENEREIGNISSEN:** Sowohl für die **Google Suche** als auch für **YouTube** gilt: Aktuelle

Nachrichtenergebnisse und das damit einhergehende erhöhte Nutzerinteresse üben oft eine geradezu magnetische Anziehungskraft auf böswillige Akteurinnen und Akteure aus. Häufig verbreiten sich innerhalb kürzester Zeit Spekulationen, noch während seriöse Nachrichtenquellen die Sachlage recherchieren. Parallel dazu veröffentlichen böswillige Akteurinnen und Akteure Inhalte in Foren und sozialen Medien, mit denen sie die nach Informationen suchenden Menschen in die Irre führen und Aufmerksamkeit erregen wollen. Um diese Art von Inhalten weniger sichtbar zu machen, haben wir unsere Systeme so ausgelegt, dass sie während einer aktuellen Krisenlage dem Faktor Zuverlässigkeit mehr Gewicht beimessen.

Ein weiteres Mittel für die Bereitstellung zuverlässiger Inhalte sind Kontextinformationen. Anhand dieser Informationen können Nutzerinnen und Nutzer selbst feststellen, inwieweit die angezeigten Inhalte vertrauenswürdig sind. Zwar sind Kontextinformationen nicht immer verfügbar, wo sie uns aber vorliegen, erlauben sie tiefere Einblicke in eine Story oder einen Inhalt.

Hier einige Beispiele:

- In der Google Suche und auf YouTube können in den Suchergebnissen Infoboxen, sogenannte Knowledge Panels und Infobereiche, angezeigt werden. Sie enthalten Kontext- und Basisinformationen zu Personen, Orten oder anderen Dingen.
- Auf YouTube-Kanälen von Fernsehsendern, die von den Regierungen ihres Landes finanziert oder betrieben werden, wird unter jedem Video ein Infobereich angezeigt. Es macht deutlich, dass der Kanal staatliche oder öffentliche Gelder bezieht.

- Seit Juni 2018 wurden auf YouTube Milliarden solcher Infobereiche eingeblendet.
- In Google News stellen wir unter den Links zu einzelnen Artikeln eine Option „Mehr zum Thema“ zur Verfügung. Mit dieser Option erhalten Nutzerinnen und Nutzer, die sich weitergehend informieren möchten, schnell und einfach Zugang zu nicht personalisierten, ausführlichen Artikeln zu dem betreffenden Thema. Dazu gehören häufig eine Zeitleiste, Tweets oder auch Faktenchecks, die weiteren Kontext bereitstellen.
- In der Google Suche heben wir seit fast drei Jahren Faktenchecks gut sichtbar hervor, um den Nutzenden ein fundiertes Urteil über Online-Inhalte zu ermöglichen. Die Faktenchecks werden jährlich viele Milliarden Mal angezeigt und wurden durch ähnliche Features auf YouTube und bei der Google Bildersuche ergänzt.

Vermeiden von Voreingenommenheit in unseren Ranking-Systemen und Richtlinien

Unsere Produkte richten sich an alle. Die weltweit mehr als 100.000 Beschäftigten bei Google haben naturgemäß eine Fülle an verschiedenen Ansichten. Daher treffen wir Vorkehrungen um sicherzustellen, dass bei der Entwicklung und Durchsetzung unserer Richtlinien kein Raum für Voreingenommenheit ist.

Wie oben erwähnt, sollen unsere Suchalgorithmen höchsten Relevanz- und Qualitätsstandards entsprechen. Aus diesem Grund verwenden wir strenge Verfahren, die sowohl Live-Tests als auch das Feedback von Tausenden speziell geschulter externer Evaluatoren rund um den Globus umfasst.²⁰

Unsere Leitlinien für Evaluatoren, die Search Quality Rater Guidelines²¹, definieren die Ziele, die wir mit unseren Bewertungssystemen verfolgen. Die Leitlinien benennen auch die Kriterien, anhand derer unsere Evaluatoren die Sachkenntnis,

²⁰ www.google.com/search/howsearchworks/mission/users/

²¹ static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

Zuverlässigkeit und Vertrauenswürdigkeit von Webseiten beurteilen.²² Die Kriterien enthalten keine politischen Überzeugungen und weisen die Prüfer ausdrücklich darauf hin, dass *„Bewertungen auf Basis der in diesen Richtlinien enthaltenen Anweisungen und Beispielen erfolgen sollen. Bewertungen dürfen nicht auf Ihren persönlichen Meinungen, Vorlieben, religiösen Überzeugungen oder politischen Ansichten beruhen.“* Darüber hinaus unterscheiden unsere Suchalgorithmen nicht danach, ob Werbung von einem Unternehmen, einer Einzelperson oder einer Organisation geschaltet wird. Bei der Analyse von Webseiten durch unsere Suchalgorithmen und bei der Durchsetzung unserer Richtlinien dulden wir keinerlei bevorzugte Behandlung einzelner Werbetreibender. Auch eine etwaige Bezahlung, mit der eine spezielle Behandlung erreicht werden soll, weisen wir kategorisch zurück.

Des Weiteren führen wir Live-Traffic-Tests durch, um festzustellen, wie reale Nutzerinnen und Nutzer eine neue Funktion einsetzen, bevor wir sie in größerem Umfang freischalten. Die Ergebnisse dieser Tests werden von erfahrenen Google-Technikerinnen und -Technikern sowie von Suchanalytistinnen und Suchanalysten überprüft. Gemeinsam ermitteln sie, ob die neue Funktion praxisreif ist. 2019 führten wir über **460.000 Tests** mit speziell geschulten externen Evaluatoren sowie Live-Tests durch, aus denen sich mehr als **3.600 Verbesserungen** der Google-Suchmaschine ergaben.

Unser Engagement geht weit über das bloße Ranking hinaus. So lassen wir uns während der Richtlinienentwicklung von einem Gremium aus verschiedenen externen und internen Personen beraten. Unser Prozess beinhaltet zahlreiche Google-Teams und bindet Führungskräfte in die abschließende Formulierung einer neuen oder aktualisierten Richtlinie ein. Wir informieren in unseren Online-Hilfen und anderen Foren über unsere Produktrichtlinien und sonstigen Leitfäden, so dass die

²² Weitere Informationen zu unseren Search Quality Raters Guidelines finden Sie unter www.google.com/search/howsearchworks

Nutzerinnen und Nutzer jederzeit wissen, welche Regeln für unsere Produkte gelten.

Darüber hinaus setzen wir unsere Richtlinien konsequent um, unabhängig vom Gegenstand der Richtlinie und von den beteiligten Parteien. Wenn ein Fall im „Graubereich“ angesiedelt ist, wird er von mehreren Personen überprüft, um eine angemessene Entscheidung sicherzustellen. Grundsätzlich wenden wir bei allen Produkten einen rigorosen Qualitätssicherungsprozess an. Mit gleicher Sorgfalt entwickeln und verwenden wir unsere Sicherheitslisten. Mit ihnen sorgen wir z. B. dafür, dass eine Website, die sich wegen besonders schwerwiegender Verstöße gegen die Werberichtlinien nicht mehr über unsere Dienste finanzieren darf, auch nicht über einen anderen Google-Dienst erneut Geschäfte machen kann.

Wir nehmen selbstverständlich nicht für uns in Anspruch, dass unsere Produkte „neutral“ sind. Jedes Ranking beinhaltet automatisch eine Klassifizierung anhand bestimmter Zielvorgaben oder Faktoren. Deshalb veröffentlichen wir die Zielvorgaben und Faktoren, für die unsere Produkte optimiert werden, und sind jederzeit offen für Feedback. So legen unsere Leitlinien für Evaluatoren, die Search Quality Rater Guidelines, beispielsweise dar, wie wir die Kriterien Sachkompetenz, Zuverlässigkeit und Vertrauenswürdigkeit in der Google Suche anwenden.

In manchen Fällen gibt es keine sachkompetenten, zuverlässigen Quellen, die wir als Reaktion auf eine Suchanfrage oder eine vermutete Nutzerabsicht anzeigen könnten. Diese Fälle, in denen die verfügbaren Antworten auf eine Suchanfrage beschränkt, nicht existent oder äußerst problematisch sind, werden von Fachleuten als „Data Void“ oder „Datenleere“ bezeichnet.²³ Bei wichtigen, nachrichtenrelevanten Ereignissen gelingt es uns mittlerweile besser als zuvor, mit solchen Data Voids umzugehen. Die Ranking-Systeme der Google Suche und von YouTube sind darauf trainiert, derartige Ereignisse zu erkennen und während einer Krisenlage zuverlässige Suchergebnisse

²³ Data Voids: Where Missing Data Can Easily Be Exploited – Danah Boyd, Michael Golbiewski, 2019 – datasociety.net/wp-content/uploads/2018/05/Data_Society_Data_Voids_Final_3.pdf

stärker zu gewichten. Dessen ungeachtet suchen wir weiterhin nach Möglichkeiten, wie wir mit Data-Void-Problemen in unseren Produkten und Services umgehen können.

Natürlich kommt es auch vor, dass Inhalte gesucht und aufgerufen werden, die unsere Systeme als qualitativ minderwertig einstufen würden. Solange sie jedoch nicht gegen geltendes Recht oder die Richtlinien verstoßen, sind sie trotzdem auf unseren Plattformen zu finden. Wie weiter oben betont, ist es unserer Ansicht nach von grundsätzlicher Bedeutung, die Entscheidung der Nutzerinnen und Nutzer zu respektieren.

Reduzieren: Weniger Reichweite für grenzwertige Inhalte

Bei der Verbreitung von Inhalten auf unseren Plattformen legen wir eine hohe Messlatte an. Grundsätzlich sollte jeder Inhalt, der über unsere Dienste verfügbar ist, für aktiv danach suchende Nutzerinnen und Nutzer auch auffindbar sein. Andererseits ist nicht jeder Inhalt empfehlenswert - und wir sind auch nicht dazu verpflichtet, jeglichen Inhalt zu empfehlen. Unsere Systeme und Richtlinien sind daher so ausgelegt, dass sie die Nutzenden nicht von sich aus mit Inhalten konfrontieren, die schädlich sein könnten.

Dieser Sicherungsmechanismus gilt für Funktionen, die Inhalte empfehlen (z. B. die Empfehlungsfunktion von YouTube), die Inhalte prominent präsentieren (z. B. die hervorgehobenen Snippets bei Google Suchergebnissen) oder die Inhalte von Partnern oder kuratierte Inhalte anzeigen (z. B. die Knowledge Panels in der Google Suche bzw. die Infobereiche auf YouTube). Er gilt auch für Funktionen, mit denen wir Nutzerinnen und Nutzern auf Grundlage realer Google Suchanfragen dabei helfen, eine beabsichtigte Suchanfrage zu formulieren (z. B. die „Autocomplete“-Funktion).

Weniger Empfehlungen von grenzwertigen Inhalten und schädlichen Fehlinformationen auf YouTube

Seit unserer Ankündigung im Januar 2019²⁴ gehen wir auf YouTube aktiv gegen Empfehlungen von grenzwertigen Inhalten und Videos vor, die Nutzerinnen und Nutzer falsch informieren und sie damit schädigen könnten. Als „grenzwertig“ werden

²⁴ youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html

Inhalte und Videos bezeichnet, die sich an der Grenze zum Richtlinienverstoß bewegen, ohne die Richtlinien tatsächlich zu verletzen. Diese Selbstverpflichtung erweitern wir nun auf Länder außerhalb der Vereinigten Staaten, darunter Großbritannien, Irland, Südafrika und andere englischsprachige Märkte. Entsprechende Initiativen gibt es auch für nicht englischsprachige Länder wie Brasilien, Frankreich, Deutschland, Mexiko und Spanien. Um die Empfehlung von grenzwertigen Inhalten und schädlichen Fehlinformationen zu reduzieren, nahmen wir mehr als **30 Änderungen** an unseren Empfehlungssystemen vor. In der Folge ging die Wiedergabezeit („Watch Time“) solcher aus nicht abonnierten Empfehlungen stammenden Inhalte in den USA im Jahr 2019 um **mehr als 70 %** zurück.

Die Einstufung von Informationen und Inhalten als „grenzwertig“ oder schädlich kann schwierig sein, insbesondere angesichts der enormen Bandbreite der Videos, die auf YouTube hochgeladen werden. Für diese anspruchsvolle Aufgabe setzen wir externe Evaluatoren aus der ganzen Welt ein, die uns Informationen zur Qualität bestimmter Videos liefern. Für diese Prüferinnen und Prüfer gelten dieselben Richtlinien wie für die Evaluatoren bei der Google Suche.²⁵ Jedes evaluierte Video erhält bis zu neun verschiedene Bewertungen, wobei manche Inhalte von Expertinnen und Experten mit ausgewiesener fachlicher Kompetenz beurteilt werden müssen. So wird die Richtigkeit von Videos zu bestimmten Therapieformen beispielsweise von Ärztinnen und Ärzten untersucht, um die Verbreitung von medizinischen Fehlinformationen einzudämmen. Basierend auf dem Konsens dieser Prüferinnen und Prüfer, erstellen wir anhand bewährter Machine Learning-Systeme Modelle, die uns bei der Analyse von Hunderttausenden Videostunden pro Tag unterstützen und dabei helfen, grenzwertige Inhalte zu erkennen und ihre Verbreitung zu reduzieren. Die Genauigkeit dieser selbstlernenden Systeme nimmt im Laufe der Zeit immer weiter zu.

²⁵ support.google.com/youtube/answer/9230586

Mehr Online-Sicherheit für Kinder und Familien

Unsere Produkte für Kinder und Familien sind von Grund auf so gestaltet, dass sie Eltern und pädagogische Fachkräfte bei der Einrichtung von möglichst sicheren Erfahrungsräumen für Kinder im Internet helfen.

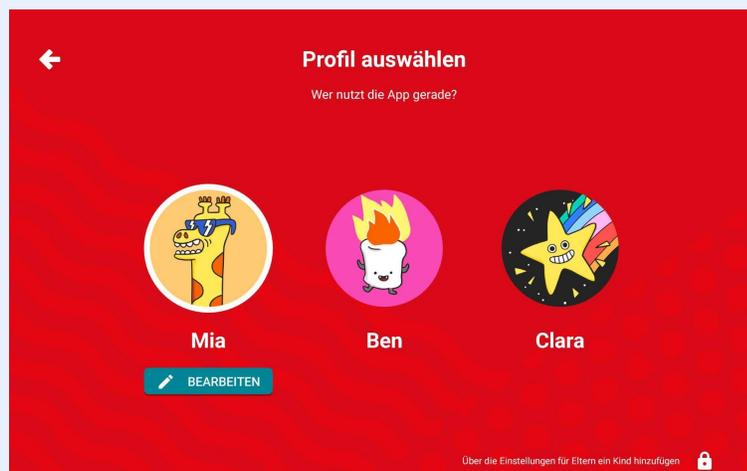
- Die **Family Link**²⁶-App ist standardmäßig in der jeweils neuesten Android-Version enthalten. Sie erleichtert es Eltern, ein wachsames Auge auf ihren Nachwuchs zu haben, wenn dieser mit einem kompatiblen Endgerät im Internet unterwegs ist. Mit Family Link können Eltern Grundregeln für die Nutzung digitaler Medien aufstellen. Dabei können sie z. B. bestimmte Apps freigeben, die Online-Zeit anzeigen lassen, die tägliche Nutzungsdauer begrenzen oder Schlafenszeiten festlegen.
- **Google Assistant für Familien**²⁷: Kinder, die sich über Family Link in ihrem eigenen Account anmelden, erhalten mit Google Assistant kindgerechte Antworten und Funktionen, die speziell für ihre Altersgruppe entwickelt wurden.
- Für **Google Play** gelten spezielle Richtlinien, die sicherstellen, dass Apps für Kinder nur geeignete Inhalte anbieten, dass kindgerechte Werbung geschaltet wird und personenbezogene Informationen geschützt werden. Die Richtlinien verringern auch die Gefahr, dass Kinder versehentlich mit Apps in Kontakt kommen, die nicht für sie gedacht sind. Inzwischen fordern wir Entwicklerinnen und Entwickler auf anzugeben, ob Kinder zu ihrer Zielgruppe gehören, und wir verlangen, dass In-App Werbung für Kinder von einem Werbenetzwerk mit zertifizierten Standards stammt.

²⁶ families.google.com/familylink

²⁷ support.google.com/assistant/answer/9071584

YouTube Kids

- **YouTube Kids**²⁸ bietet ein eigenes YouTube-Erlebnis, das speziell für Kinder entwickelt wurde und das die Eltern kontrollieren können. Die App verwendet eine Mischung aus Filtern, Nutzer-Feedback und manueller Überprüfung, um die Videos in YouTube Kids familienfreundlich zu machen und dem Nachwuchs einen Katalog von Inhalten in einer sichereren Umgebung bereitzustellen. Mit speziellen Einstellungen können Eltern selbst festlegen, welche Inhalte ihre Kinder sehen dürfen. Alternativ können sie Inhalte aus externen Sammlungen auswählen, die von Expertinnen und Experten und etablierten Anbietern von Inhalten für Kinder zusammengestellt wurden.



- Mit dem **YouTube-Konto mit Elternaufsicht**²⁹ haben Eltern die Möglichkeit, verschiedene Einstellungen für ältere Kinder und Jugendliche vorzunehmen und so festzulegen, dass sie nur bestimmte Inhalte sehen und Funktionen nutzen können.

Weitere Informationen zu unserem Engagement für Familien finden Sie unter safety.google/families

²⁸ youtube.com/kids

²⁹ youtube.com/myfamily

Belohnen: Höhere Standards für die Monetarisierung

Wir stellen besonders hohe Anforderungen an die Informationsqualität von Diensten, in denen Werbeanzeigen geschaltet und Inhalte vermarktet werden, etwa mit Hilfe von Google Ads und AdSense. Wir streben weder zu unserem eigenen Nutzen noch zum Nutzen anderer danach, mit schädlichen Inhalten oder Verhaltensweisen Einnahmen zu erzielen. Da gerade böswillige Akteurinnen und Akteure durch schädliche Inhalte oft finanzielle Vorteile erlangen wollen, können wir durch eine hohe Schwelle für den Zugriff auf Monetarisierungsmöglichkeiten den Missbrauch unserer Dienste für diesen Personenkreis weniger attraktiv machen.

Hassrede und irreführende Verhaltensweisen sind in unseren Werbeprodukten ausdrücklich untersagt. Dazu gehört auch der Ausschluss von Publishern, die unsere Dienste nutzen wollen, um belästigende und auf Cybermobbing abzielende Werbung oder Werbung mit anderweitig gefährlichen oder herabsetzenden Inhalten zu veröffentlichen. Dieses Verbot gilt auch für Publisher, die den eigentlichen Zweck ihrer Website verschleiern wollen. So identifizierten wir 2017 z. B. eine Gruppe von Publishern in Mazedonien, deren Websites vorgaben, amerikanische Nachrichtenkanäle zu sein. Nachdem wir dies festgestellt hatten, wurden die Publisher von der Monetarisierung ausgeschlossen.

Um finanziellen Schaden von unseren Nutzerinnen und Nutzern abzuwenden, schränken wir auch die Nutzung unserer Werbeprodukte durch bestimmte Arten von Unternehmen stark ein. Studien zeigen, dass gewinnorientierte Kautionsanbieter in den USA ihr Geld größtenteils in Gegenden mit einer überwiegend einkommensschwächeren Bevölkerung verdienen. Dabei nutzen sie die Notlage der Menschen mit undurchsichtigen Finanzierungsangeboten aus, die zu einer langfristigen Verschuldung führen können. Nach intensiven Beratungen mit Fachleuten entschieden wir, den

Nutzerschutz mit einer neuen Richtlinie zu stärken und die Anzeigen für Kautionsdienstleistungen auf unseren Plattformen einzuschränken.

Auf YouTube können Creator unser Partnerprogramm³⁰ nutzen, um mit ihren Inhalten Einnahmen zu erzielen und mit zusätzlichen Tools ihre Kanäle zu optimieren. Für die Teilnahme an dem Programm müssen sie eine bestimmte Mindestzahl an Abonnentinnen und Abonnenten sowie Wiedergabezeit ihrer öffentlichen Videos erreichen und die Monetarisierungsrichtlinien für YouTube einhalten. Verwarnungen aufgrund vorheriger Verstöße gegen die Richtlinien, Spam- oder anderweitige Missbrauchsmeldungen werden von uns sorgfältig beobachtet, um Hinweise auf problematisches Verhalten zu erhalten. Verstöße gegen die Richtlinien können dazu führen, dass Werbung in bestimmten Videos blockiert, das AdSense-Konto eines Kanals deaktiviert, der Creator aus dem YouTube-Partnerprogramm ausgeschlossen oder der Kanal vollständig gelöscht wird.

Maßnahmen gegen die Monetarisierung von Inhalten minderer Qualität

Die Bewerbung oder Monetarisierung von qualitativ minderwertigen Inhalten entspricht weder dem Selbstverständnis von Google noch liegt es in unserem Geschäftsinteresse. Auch die meisten Werbetreibenden wollen von derartigen Inhalten nicht profitieren oder damit Einnahmen erzielen. Wir selbst haben ein hohes Interesse daran, von diesen Unternehmen als vertrauenswürdiger Partner gesehen zu werden, der die Integrität ihrer Marken zuverlässig schützt. Aus diesem Grund investieren wir erhebliche Summen in menschliche und technische Unterstützung, um die Monetarisierung solcher Inhalte zu verhindern. Weder wir selbst noch die Werbetreibenden, die unsere Plattformen nutzen, möchten mit qualitativ minderwertigen Inhalten in Verbindung gebracht werden.

³⁰ support.google.com/youtube/answer/72851

2018 führten wir zum Beispiel folgende Maßnahmen durch:

- Fast **734.000 Publisher und App-Entwickler** wurden aus unserem Werbenetzwerk ausgeschlossen.
- Aus annähernd **1,5 Millionen Apps** wurden Werbeschaltungen komplett entfernt.
- Die Mindestvoraussetzungen für die Teilnahme an unserem YouTube-Partnerprogramm wurden auf **4.000 Stunden Wiedergabezeit** innerhalb der letzten zwölf Monate und **1.000 Abonnenten** erhöht.

2020 wurden **Werbeanzeigen aus über 1,3 Milliarden Webseiten entfernt**, die Teil unseres Publisher-Netzwerks waren und gegen unsere Richtlinien verstoßen hatten.

Mit der kontinuierlichen Weiterentwicklung unserer Richtlinien wollen wir sicherstellen, dass mit diesen und anderen Diensten keine schädlichen Informationen oder Erfahrungen bereitgestellt oder verbreitet werden. Niemand – und am wenigsten Google selbst – soll damit Einnahmen erzielen.

Zusammenarbeit mit Partnern

Das Sicherstellen von Informationsqualität und der verantwortungsvolle Umgang mit Inhalten in Google-Produkten und -Diensten erfordert erhebliche Ressourcen. Erschwert wird diese Aufgabe durch die Geschwindigkeit, mit der Inhalte erzeugt und geteilt werden, und durch die ausgeklügelten Strategien böswilliger Akteurinnen und Akteure. Glücklicherweise müssen wir sie nicht alleine schultern.

Wir arbeiten mit einer Fülle an kompetenten Partnern aus der Technologiebranche, Behörden und Zivilgesellschaft zusammen, um bestmögliche Resultate zu erzielen. Gemeinsam setzen wir alles daran, dass unsere Richtlinien effektiv sind, dass wir einschlägige Best Practices definieren, sie teilen und daraus lernen und dass es uns gelingt, neue Herausforderungen erfolgreich zu bewältigen. Die folgenden Abschnitte geben einen Überblick über diese Arbeit.

Gemeinsam gegen unzulässige Inhalte

Zusätzlich zu unseren eigenen Anstrengungen setzen wir auf ein Netzwerk von Partnern, deren Fachwissen und Erfahrung uns dabei helfen, nicht richtlinienkonforme Inhalte zu erkennen.

Das **YouTube Trusted Flagger-Programm**³¹ wurde ins Leben gerufen, um einzelne Personen, Regierungsbehörden und Nichtregierungsorganisationen (NGOs) zu unterstützen, die besonders häufig YouTube-Inhalte melden, die unsere Community-Richtlinien verletzen. Das Programm stellt den "Trusted Flaggen" eine spezielle Oberfläche bereit, mit der sie mehrere Videos gleichzeitig melden können. Zudem ermöglicht es den regelmäßigen Austausch und das Feedback zu den

³¹ support.google.com/youtube/answer/7554338

Maßnahmen, die YouTube für verschiedene Inhaltsbereiche anwendet.

Das Trusted Flagger-Programm ist ein Element eines Netzwerks aus **mehr als 180 Forschungseinrichtungen, Behörden und NGOs weltweit**, die uns mit ihrem wertvollen Expertenwissen bei der Durchsetzung unserer Richtlinien unterstützen.

Die Teilnehmer am Trusted Flagger-Programm werden darin geschult, wie die Community-Richtlinien von YouTube umgesetzt werden können. Da ihre Meldungen eine höhere Genauigkeit aufweisen als die anderer Nutzerinnen und Nutzer, priorisieren wir Trusted Flagger-Meldungen. Von Trusted Flaggern gemeldete Inhalte unterliegen denselben Richtlinien wie von anderen Nutzerinnen und Nutzern gemeldete Inhalte und werden von unseren geschulten Teams geprüft. Diese entscheiden, ob die betreffenden Inhalte gegen unsere Community-Richtlinien verstoßen und entfernt werden müssen.

Bei der Verfolgung und Dokumentation der Aktivitäten von böswilligen Akteurinnen und Akteuren, die unsere weltweiten Produkte und Dienste ins Visier nehmen, wenden wir uns außerdem an Organisationen, die auf solche Aufgaben spezialisiert sind. Zum Schutz der betreffenden Unternehmen und ihrer Beschäftigten veröffentlichen wir hierzu meist nur wenige Informationen. Ein bekanntes Beispiel ist unsere Zusammenarbeit mit dem Cyber Sicherheitsunternehmen FireEye, in deren Zuge eine Reihe von Sicherheitsvorfällen und versuchter Einflussnahmen aufgedeckt wurden.

Weiterentwicklung und Verbesserung unserer Richtlinien

Angesichts neuer Verhaltensweisen und Erwartungen unserer Nutzerinnen und Nutzer und als Reaktion auf die immer ausgefeilteren Taktiken der Böswilligen müssen wir unsere Richtlinien ständig verfeinern und weiterentwickeln. Dabei greifen wir häufig auf die Unterstützung durch Fachleute in verschiedenen Bereichen und die Erkenntnisse von Wissenschaft, zivilgesellschaftlichen Organisationen und anderen Unternehmen der Branche zurück.

Richtlinien gegen Hassrede auf YouTube

Nach eingehender Beratung mit zahlreichen Expertinnen und Experten zu Themen wie gewalttätigem Extremismus, Rassismus, Bürgerrechten und dem Recht auf freie Meinungsäußerung haben wir unsere Richtlinien für YouTube überarbeitet. Verboten sind demnach Videos, die die Überlegenheit einer bestimmten Personengruppe nahelegen, um andere Personen auf Basis von Merkmalen wie Alter, Geschlechtsidentität, sexueller Orientierung, ethnischer Herkunft, Kastenzugehörigkeit, religiöser Überzeugung oder Veteranenstatus zu diskriminieren, abzusondern oder auszuschließen.

Gemeinsam mit Fachleuten haben wir zudem strengere Richtlinien zu Belästigung und Cybermobbing formuliert. Allein im vierten Quartal 2020 führte dies dazu, dass **über 174.000 Videos und über 182 Millionen Kommentare** wegen Hassrede und Belästigung gelöscht wurden. Trotz dieser Erfolge wissen wir, dass noch viel zu tun bleibt. Wir werden unsere Richtlinien und

Produkte daher fortlaufend auf den Prüfstand stellen, damit sie allen Nutzerinnen und Nutzern gerecht werden.

Best Practices für ein besseres Internet

Um Herausforderungen zu bewältigen, die zahlreiche Produkte und verschiedene Ökosysteme betreffen, kooperieren wir mit anderen Technologieunternehmen und Branchenvertretern. Dabei prüfen wir sorgfältig, wo eine Kooperation hilfreich ist und wo Google mit seinen Ressourcen andere unterstützen kann. Diese Art der Zusammenarbeit ist oft am effektivsten, um auf breiter Front gegen böswillige Akteurinnen und Akteure vorzugehen.

Bekämpfung von sexuellem Missbrauch von Kindern

Um das entsetzliche Phänomen der Darstellung von sexuellem Missbrauch an Kindern im Internet zu bekämpfen, ist Google seit 2006 gemeinsam mit anderen Branchenvertretern in der Technology Coalition³² organisiert. Wir stellen Unternehmen und Nichtregierungsorganisationen kostenlos neueste Technologien bereit, um ihnen dabei zu helfen, solche Darstellungen schneller und in größerem Umfang zu identifizieren, zu löschen und zu melden.³³ Die Mitglieder der Technology Coalition haben in den vergangenen Jahren signifikante Fortschritte dabei erzielt, innovative Technologien für den Kampf gegen Darstellungen des sexuellen Missbrauchs an Kindern zu entwickeln und zu implementieren. 2020 startete die Coalition mit „Project Protect“ einen ambitionierten Investitions- und Strategieplan, der die gemeinsamen Anstrengungen dauerhaft verstärken und vertiefen soll.³⁴

Mit technischen Werkzeugen wie CSAI Match und der Content Safety API, die von Google- und YouTube-Entwicklerinnen und -Entwicklern erstellt wurden, lassen sich potenziell illegale Inhalte gezielter überprüfen. Zugleich können bekannte und neue Darstellungen von sexuellem Missbrauch an Kindern effektiver identifiziert werden. Diese Werkzeuge kommen nicht nur auf unseren eigenen Plattformen, sondern auch bei Unternehmen wie Adobe, Tumblr und Reddit zum Einsatz. Sie ermöglichen die schnellere

³² technologycoalition.org

³³ protectingchildren.google

³⁴ technologycoalition.org/2020/05/28/a-plan-to-combat-online-child-sexual-abuse

Identifizierung potenzieller Opfer sexuellen Missbrauchs und verringern die seelische Belastung der Menschen, die diese Inhalte überprüfen.

Bekämpfung von terroristischen Inhalten

Wir wollen Terroristen wirksam daran hindern, zu terroristischen Handlungen aufzurufen, gewalttätige extremistische Propaganda zu verbreiten und reale Gewalttaten auf unseren Plattformen zu zeigen und zu glorifizieren. Aus diesem Grund haben wir gemeinsam mit anderen Branchenvertretern das Global Internet Forum to Counter Terrorism (GIFCT) ins Leben gerufen.

Gemeinsam gegen
Terrorismus



Neben anderen wichtigen Initiativen ermöglicht es den beteiligten Unternehmen und Organisationen, digitale Fingerabdrücke, sogenannte Hashwerte, von bekannten terroristischen und extremistischen Inhalten in eine Datenbank hochzuladen. Anhand dieser digitalen Fingerabdrücke können die entsprechenden Inhalte dann schnell von allen teilnehmenden Plattformen gelöscht werden.

Durch die gemeinsame Anwendung von Best Practices und die Zusammenarbeit bei plattformübergreifenden Tools konnten wir:

- die GIFCT-Datenbank auf 200.000 Hashwerte erweitern,
- ein weltweites Forschungsnetzwerk knüpfen, um Einblicke in die Art und Weise zu erhalten, wie Terroristen Technologien nutzen, und
- neue Mitglieder für das GIFCT gewinnen sowie mehr als 100 kleinere Technologieunternehmen durch weltweit stattfindende Workshops an dem Projekt beteiligen.

Die Verbreitung terroristischer Inhalte über das Internet ist ein dynamisches Problem, für dessen Bekämpfung wir unsere Tools und Ansätze für Gegenmaßnahmen fortlaufend nachjustieren müssen. GIFCT ist ein zentrales Element in diesem Kampf. Die tragischen Ereignisse von Christchurch im Jahr 2019 zeigten beispielhaft, wie dringlich eine Verbesserung des Informationsaustauschs zwischen Plattformen ist, um Live-Uploads und koordinierte Re-Uploads von terroristischen Inhalten zu verhindern.

Wir sind stolz darauf, zu den Unterzeichnern des „Christchurch-Aufrufs zur Eliminierung von terroristischen und gewalttätigen extremistischen Online-Inhalten“ zu zählen, und arbeiten mit Nachdruck an der Umsetzung der damit verbundenen Verpflichtungen. So entwickelten und implementierten die GIFCT-Teilnehmer z. B. ein Krisenprotokoll, das die Reaktion auf reale Ereignisse wie die Ermordung wehrloser Unschuldiger und Zivilisten regelt. Zum Tragen kommt dieses Protokoll z.B. bei einer terroristischen Gewalttat, die live gefilmt und online übertragen wird. Ende 2019 organisierte YouTube gemeinsam mit der neuseeländischen Regierung einen international besetzten Krisenpräventions-Workshop in der Hauptstadt Wellington, bei dem das Krisenprotokoll weiter verfeinert wurde.

2020 entwickelte sich das GIFCT zu einer unabhängigen Organisation unter Leitung eines Executive Directors und unterstützt durch versierte Teams für Technologie, Terrorabwehr und operative Aufgaben. Auf Basis der 2019 gewonnenen Erkenntnisse setzt sich das neue und unabhängige GIFCT-Konsortium weiterhin für Wissensaustausch, technische Innovation und gemeinsame Forschung mit Expertinnen und Experten, Kräften der Zivilgesellschaft und Regierungsstellen ein. Ein weiterer Aufgabenschwerpunkt ist die Umsetzung der Verpflichtungen aus dem Christchurch-Aufruf. Hierfür unterstützt das GIFCT wissenschaftliche Forschungen, fördert Counter-Speech-Projekte im Netz und hilft einem breiten Spektrum von Technologieunternehmen dabei, den Missbrauch ihrer Plattformen zu verhindern und zu bekämpfen.

Wir wollen diese Kooperationen fortentwickeln, von ihnen lernen und dabei weitere Möglichkeiten finden, wie wir gemeinsam mit unseren Partnern aus Wirtschaft und Politik Best Practices definieren können.

Mehr Informationsqualität durch bessere Regulierung

Von einer durchdachten Regulierung profitiert sowohl die Gesellschaft als auch das Internet selbst. Dabei ist die richtige Balance nirgends so wichtig wie bei der Debatte über Online-Inhalte.

Für Online-Inhalte gelten bereits heute zahlreiche Gesetze, die Verbraucherinnen und Verbraucher schützen, üble Nachrede unter Strafe stellen und den Umgang mit sensiblen Daten regeln. Ein ausgewogen konzipierter Rechtsrahmen für Online-Plattformen spielt auch eine wichtige Rolle bei der Antwort auf die Frage, wie ein angemessener Umgang mit illegalen Inhalten aussehen kann. Effektive Safe-Harbor-Bestimmungen geben z. B. vor, wie Online-Plattformen ihren rechtlichen Verpflichtungen nachkommen können, wenn sie auf illegale Inhalte hingewiesen werden. Sie legen auch fest, dass eine Online-Plattform, die mit alternativen freiwilligen Maßnahmen gegen illegale oder anderweitig schädliche Inhalte vorgeht, nicht belangt wird. All diese Gesetze und Bestimmungen fördern den freien Informationsfluss, Innovation und wirtschaftliches Wachstum, geben den Plattformen aber zugleich die nötige Rechtssicherheit im Kampf gegen problematische Inhalte.

Eine systematische Beobachtung der Erstellung und Durchsetzung von Richtlinien kann ebenfalls eine wichtige ergänzende Funktion haben. Seit es das Internet gibt, haben Unternehmen, Politik und Zivilgesellschaft Regeln für gute Praxis erarbeitet, die den Online-Diensten als Richtschnur für ein angemessenes Verhalten dienen sollen ("Code of Conduct"). In dem Maße, in dem Content-Sharing-Dienste wie soziale Medien und Video-Plattformen relevante Orte für die öffentliche Debatte werden, entwickeln sich auch die

Beobachtungs- und Aufsichtsinstrumente weiter, indem sie z.B. das Verhalten der Plattformen anhand der Best Practices bewerten.

Neue Aufsichtsmethoden können unserer Ansicht nach gut funktionieren, wenn sie sich auf ein spezifisches, klar definiertes Problem und drei Aspekte konzentrieren:

- **Definition von Standards für mehr Transparenz und Best Practices:** Transparenz ist die Voraussetzung für effektive Praktiken und die Grundlage für eine fundierte Diskussion. Da immer neue Technologien und Kommunikationsformen entstehen, sollten Aufsichtsbehörden einen flexiblen, kollaborativen Ansatz verfolgen, der Best Practices unterstützt und Forschung wie Innovation fördert.
- **Bekämpfung von systemischen Versagen anstelle einzelner Vorkommnisse:** Die Größe und Komplexität moderner Online-Plattformen erfordert einen Ansatz, der nicht einzelne Begebenheiten, sondern das große Ganze in den Blick nimmt. Da sich problematische Inhalte nie vollständig aus dem Internet verbannen lassen werden, sollten wir die Fortschritte bei der Bekämpfung solcher Inhalte anerkennen. Und wir sollten datengestützte Methoden anwenden, um zu verstehen, ob bestimmte Vorkommnisse Einzelphänomene oder Ausdruck eines systemischen Problems sind.
- **Förderung der internationalen Zusammenarbeit:** Unsere Zeit ist durch multinationale Internet-Plattformen und Nutzerinnen und Nutzer geprägt, die weltumspannend kommunizieren und auf Informationen zugreifen. Deshalb sollten auch die Staaten Best Practices austauschen und widerstreitende Vorgehensweisen vermeiden, die nur die Einhaltung der Regeln unnötig erschweren. Vielmehr sollten in einer internationalen Kraftanstrengung gemeinsame Prinzipien und Praktiken definiert werden. Zugleich gilt: Verschiedene Staaten werden auf Basis der allgemeinen Menschenrechte das Recht auf freie Meinungsäußerung unterschiedlich auslegen, abhängig von ihren Rechtsgepflogenheiten, ihrer Geschichte und ihren gesellschaftlichen Werten. Inhalte, die in einem Land illegal sind, können anderswo zulässig sind. Kein Land sollte befugt sein, seine eigenen Regeln auf die Bürger eines anderen Staates anzuwenden.

Bekämpfung von Fehlinformationen zu COVID-19

Wir wenden die in diesem Whitepaper beschriebenen Grundsätze und Instrumente an, um die zahlreichen neuen Entwicklungen und Herausforderungen rund um Informationsqualität und den verantwortungsvollen Umgang mit Inhalten in unseren Diensten zu bewältigen. Die Corona-Pandemie ist eine dieser Entwicklungen, mit einer nie dagewesenen und völlig unerwarteten Reichweite und Auswirkung.

Um angemessen darauf reagieren und aufkommende Probleme rechtzeitig erkennen zu können, haben wir unsere Richtlinien und Praktiken einer detaillierten Überprüfung unterzogen. Nachdem wir feststellen mussten, dass sich COVID-19 zum regelrechten Magneten für Online-Betrüger entwickelte, haben wir beispielsweise eine eigene Webseite eingerichtet, auf der sich Nutzerinnen und Nutzer über Betrugsversuche rund um COVID-19 und geeignete Schutzmaßnahmen informieren können.³⁵

In einem im Zusammenhang mit diesem Whitepaper besonders wichtigen Bereich bauen wir auch in Zeiten von Covid-19 auf unseren bestehenden Anstrengungen gegen Fehlinformationen auf³⁶: das Sichtbarmachen zuverlässiger Gesundheitsinformationen und die Bekämpfung medizinischer Fehlinformationen in unseren Diensten.

Sichtbar machen von vertrauenswürdigen COVID-19-Informationen

Wir haben zahlreiche Anstrengungen unternommen, um vertrauenswürdige Informationen in den Vordergrund zu rücken. Darüber hinaus kooperieren wir mit Gesundheitsorganisationen und Regierungen, um unseren Nutzerinnen und Nutzern jederzeit

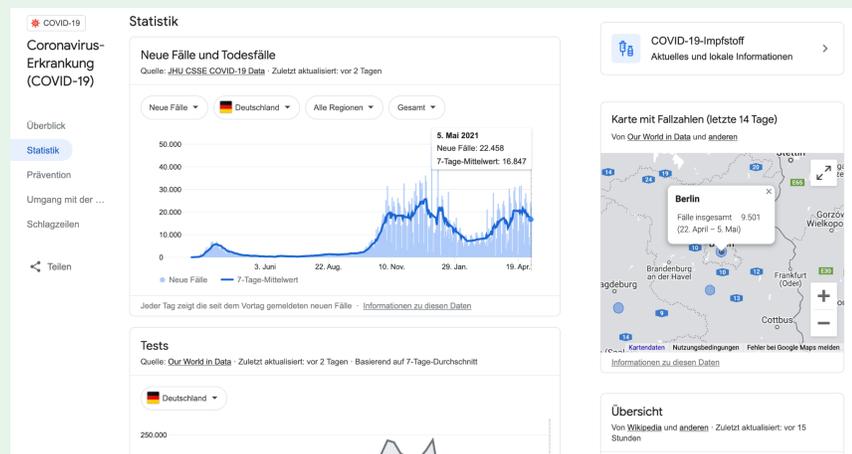
³⁵ safety.google/securitytips-covid19

³⁶ Siehe z. B. unser Whitepaper vom Februar 2019 „Wie Google Desinformation bekämpft“: services.google.com/fh/files/blogs/wie_google_desinformation_bekaempft.pdf

Covid-19 Suchformat,
international verfügbar

zuverlässige Informationen bereitzustellen:

- In der **Google Suche** haben wir ein umfassendes Suchformat zu COVID-19³⁷ eingeführt, das offizielle Gesundheitsinformationen mit neuesten Zahlen und Visualisierungen verbindet. Indem die Suchergebnisse in verschiedenen Kategorien übersichtlich angeordnet werden, können Nutzerinnen und Nutzer leichter durch die verfügbaren Ressourcen navigieren. Zugleich können neue Informationen jederzeit hinzugefügt werden. Das Format ergänzt die vorhandenen Maßnahmen, mit denen die Google Suche³⁸ und Google News³⁹ sensible Ereignisse und Kontexte identifizieren. Unsere Systeme sind generell so ausgelegt, dass sie zuverlässige Quellen für diese Art von Suchanfragen höher gewichten.



- Auf der **Google Startseite** veröffentlichen wir in Absprache mit der Weltgesundheitsorganisation (WHO) und lokalen Gesundheitsbehörden Ratschläge und Informationen, mit denen die Verbreitung von COVID-19 eingedämmt werden soll. Entsprechende Maßnahmen, zu denen auch „Stay Home“-Doodles und Nachrichten auf unserer Startseite gehören, wurden in über 100 Ländern ergriffen.

³⁷ blog.google/products/search/connecting-people-covid-19-information-and-resources

³⁸ g.co/howsearchworks

³⁹ newsinitiative.withgoogle.com/hownewsworks/mission

- Auf **YouTube** räumen wir zuverlässigen Quellen wie der WHO und lokalen Gesundheitsbehörden Vorrang ein, damit Nutzende jederzeit aktuelle COVID-19-Informationen erhalten. Weitere Neuerungen sind ein COVID-19-Bereich auf unserer Homepage mit Beiträgen von vertrauenswürdigen Nachrichtenanbietern und lokalen Gesundheitsbehörden, Infobereichen zu Gesundheitsfragen⁴⁰, die in Suchergebnissen angezeigt werden und Hinweise zu COVID-19-Symptomen, -Prävention und -Behandlung enthalten, sowie Links zu lokalen Gesundheitsbehörden⁴¹ in der Einzelansicht von Videos zu COVID-19. Insgesamt wurden diese Infobereiche **über 400 Milliarden** Mal angezeigt. Wenn unsere Systeme eine gesundheitsbezogene Suchanfrage erkennen, fördert YouTube Inhalte von Kanälen mit nachgewiesener Zuverlässigkeit wie z. B. Nachrichtenorganisationen oder Gesundheitsbehörden.
- In **Google News** gibt es einen neuen COVID-19-Bereich⁴² mit Links zu aktuellen, relevanten Beiträgen, die aus zuverlässigen, internationalen wie lokalen Quellen stammen. Der Bereich ist in 40 besonders stark betroffenen Ländern verfügbar und räumt COVID-19-Beiträgen Vorrang ein, die von lokalen Nachrichtenverlagen in der Region der Nutzerin oder des Nutzers veröffentlicht werden.
- In **Google Maps** ist es nun sehr viel einfacher, zuverlässige Informationen zu lokalen Gesundheitsressourcen abzurufen. Dazu zählen COVID-19-Testzentren, Notunterkünfte und Lebensmittelausgaben ebenso wie virtuelle Gesundheits-Angebote, sowie diese verfügbar sind. Wir nutzen zudem zuverlässige Datenquellen, um aktuelle Informationen zu den Öffnungszeiten von Geschäften während der Pandemie anzuzeigen. Darüber hinaus stellen wir Unternehmen neue Möglichkeiten bereit, wie sie Online-Informationen und -Services aktualisieren können. Dies gilt z. B. für Restaurants, die zwar geschlossen sind, ihre

⁴⁰ support.google.com/youtube/answer/9795167

⁴¹ support.google.com/youtube/answer/9004474

⁴² goo.gl/COVID19-GoogleNews

Gerichte aber ausliefern oder zum Mitnehmen anbieten.

- In **Google Play** haben wir die Prüfung und Freigabe von richtlinienkonformen Apps priorisiert, die von offiziellen Regierungsstellen und öffentlichen Gesundheitsorganisationen veröffentlicht, in Auftrag gegeben oder autorisiert werden. Autorisierte COVID-19-Apps müssen sämtliche Richtlinien für Google Play-Entwicklerinnen und -Entwickler erfüllen, darunter die Richtlinien zu Nutzerdaten⁴³, Berechtigungen⁴⁴ und irreführendem Verhalten⁴⁵. Eine neue Informationsseite im Play Store enthält Apps, mit denen sich Nutzende während der Krise auf dem Laufenden halten können. Die Seite greift auf vertrauenswürdige Quellen wie die App der Weltgesundheitsorganisation (WHO) zurück.
- Ebenfalls **neu** ist eine **Webseite**, die Ressourcen für COVID-19-Aufklärung und -Prävention bereitstellt. Zum Zeitpunkt der Veröffentlichung dieses Whitepapers ist diese Webseite unter www.google.com/COVID-19 in mehr als zwanzig Sprachen abrufbar und wird ständig weiter ausgebaut.

Maßnahmen gegen medizinische Fehlinformationen:

Neben der Förderung zuverlässiger Informationen haben wir Schritte für die Erkennung und Entfernung von COVID-19-Fehlinformationen ergriffen, die im Widerspruch zur Einschätzung von Gesundheitsbehörden stehen und konkrete Schäden verursachen können. Dabei handelt es sich um folgende Maßnahmen:

- Die **Community-Richtlinien von YouTube**⁴⁶ untersagen Inhalte, mit denen zu gefährlichen oder illegalen Praktiken ermutigt wird, die zu ernsthaften körperlichen Schäden oder zum Tode führen können. Dies betrifft nun auch bestimmte

⁴³ play.google.com/about/privacy-security-deception

⁴⁴ play.google.com/about/privacy-security-deception/permissions

⁴⁵ play.google.com/about/privacy-security-deception/malicious-behavior

⁴⁶ youtube.com/howyoutubeworks/policies/community-guidelines

Arten von medizinischen Fehlinformationen. Im Zuge der Corona-Pandemie arbeiten wir eng mit der WHO und lokalen Gesundheitsbehörden zusammen. Damit wollen wir sicherstellen, dass die Durchsetzung unserer Richtlinien einen wirksamen Beitrag dazu leistet, die Verbreitung von Fehlinformationen rund um COVID-19 zu verhindern. Unsere Richtlinien untersagen z. B. Inhalte, die die Wirksamkeit von Empfehlungen der WHO oder lokaler Gesundheitsbehörden zum Social Distancing explizit in Frage stellen und Menschen dazu veranlassen könnten, diesen Empfehlungen nicht Folge zu leisten. Wir wenden diese Richtlinien konsequent an und setzen zudem unsere 2019 begonnene Initiative⁴⁷ fort, mit der wir grenzwertige Inhalte und Videos reduzieren, die unsere Nutzerinnen und Nutzer falsch informieren und somit gefährden könnten. Seit Beginn der Pandemie haben unsere Teams mehr als **850.000 Videos** mit gefährlichen oder irreführenden medizinischen Informationen zu COVID-19 entfernt.

- Die überarbeiteten **Google Play-Richtlinien** untersagen Entwicklerinnen und Entwicklern, aus sensiblen Ereignissen Kapital zu schlagen. Unsere bewährten Inhalterichtlinien enthalten ein striktes Verbot von Apps mit gesundheitsbezogenen Inhalten oder Funktionen, die irreführend oder potenziell schädlich sind. Dies gilt auch für Inhalte oder Funktionen zu COVID-19. Apps, die diese Richtlinien verletzen, werden aus dem App Store entfernt.
- Die **Google Maps**-Richtlinien untersagen Fehlinformationen zu COVID-19-Prävention, -Übertragung und -Behandlung sowie unbewiesene Behauptungen darüber, dass sich eine Person an einem bestimmten Ort mit dem Coronavirus infiziert hat. Entsprechende Content-Beiträge werden entfernt.
- In **Google Ads** gestatten unsere Richtlinien⁴⁸ keine Werbeanzeigen, die aus einem sensiblen Ereignis wie

⁴⁷ blog.youtube/intl/de-de/inside-youtube/die-vier-stufen-der-verantwortung-teil

⁴⁸ support.google.com/adspolicy/answer/6015406

beispielsweise einem öffentlichen Gesundheitsnotstand Kapital schlagen oder dieses nicht mit der angemessenen Sensibilität behandeln. Im Zuge der Pandemie sind wir dazu übergegangen, Regierungsorganisationen, Anbietern von Gesundheitsdiensten, Nichtregierungsorganisationen, zwischenstaatlichen Organisationen, verifizierten Online-Wahlwerbern und verwalteten privaten Google-Konten, die sich erwiesenermaßen an die Richtlinien halten und der Öffentlichkeit relevante Informationen bereitstellen wollen, Plätze für Anzeigen im Zusammenhang mit Aufklärung zu COVID-19 zu reservieren. Auch diese Anzeigen müssen unseren Richtlinien⁴⁹ entsprechen, die eine Verbreitung schädlicher medizinischer oder gesundheitsbezogener Behauptungen und Praktiken ausdrücklich untersagen.⁵⁰ Darüber hinaus haben wir vorübergehende Restriktionen zum Handel mit persönlicher Schutzausrüstung in Kraft gesetzt und ergreifen weitere Schritte, um künstlich überhöhte Preise zu verhindern, die den Zugang zu anderen essentiellen Gütern im Google-Netzwerk begrenzen oder verhindern. Weitere Informationen finden Sie in der Google Ads-Hilfe⁵¹.

Im Verlauf der Corona-Pandemie überarbeiten und überprüfen wir fortlaufend unsere Richtlinien und deren Durchsetzung, um auf jede Art von Falschinformation zu COVID-19 angemessen reagieren zu können.

Unterstützung von Prüfteams:

Da unsere erweiterte Belegschaft vorübergehend nur eingeschränkt zur Verfügung steht, haben wir Beschäftigte bei Google mit neuen Aufgaben betraut, um besonders problematische Inhalte auch weiterhin konsequent ausfiltern zu können. Durch zusätzliche Hygienemaßnahmen und die Bereitstellung privater Transportmöglichkeiten können die Mitarbeiter diese Aufgaben vor Ort im Unternehmen

⁴⁹ support.google.com/adspolicy/answer/6008942

⁵⁰ support.google.com/adspolicy/answer/6015406

⁵¹ support.google.com/google-ads/answer/9811449

durchführen. Dank ihres Einsatzes sind wir in der Lage, uns um besonders dringliche Arbeitsabläufe und Meldungen trotz COVID-19 angemessen zu kümmern. Das gilt insbesondere für Meldungen, die von Trusted Flagger-Partnern und Regierungsstellen bei uns eingehen.

Damit sich möglichst wenige Mitarbeiterinnen und Mitarbeiter vor Ort im Unternehmen aufhalten müssen, setzen wir nach Möglichkeit vermehrt automatisierte Systeme ein. Um der Gefahr von Fehlentscheidungen bei der automatisierten Verarbeitung zu begegnen (d.h. der fälschlichen Löschung legitimer Inhalte), können unsere Creator Einspruch gegen eine Verwarnung einlegen.

Unterrichtung der Nutzerinnen und Nutzer über unsere Aktivitäten:

Im Verlauf der Corona-Pandemie haben wir eine Fülle an öffentlich zugänglichen Informationen bereitgestellt, um Nutzerinnen und Nutzer, Zivilgesellschaft und andere Interessierte über unsere Aktivitäten zu unterrichten. Diese Informationen veröffentlichen wir auf speziellen Seiten im Google-Blog⁵² und in der YouTube-Hilfe⁵³.

⁵² blog.google/inside-google/company-announcements/googles-response-covid-19

⁵³ support.google.com/youtube/answer/9777243

Zusammenfassung

Informationsqualität und der verantwortungsvolle Umgang mit Inhalten sind von zentraler Bedeutung für unser Selbstverständnis. Darin kommt unsere Selbstverpflichtung zum Ausdruck, Nutzerinnen und Nutzern zuverlässige und nützliche Informationen bereitzustellen, die ihren Erfordernissen entsprechen und sie nicht gefährden. An dieser Selbstverpflichtung werden wir jeden Tag und bei jeder Nutzerinteraktion gemessen.

Informationsqualität und der Umgang mit Online-Inhalten bergen jedoch auch besondere und nicht unerhebliche Herausforderungen. Selbst objektive Betrachter können ganz unterschiedliche Vorstellungen davon haben, was wünschenswert ist. Zugleich sind gerade böswillige Akteure besonders hartnäckig und kreativ. Dazu kommt, dass die schiere Menge an Online-Inhalten es unmöglich macht, jedem Inhalt dasselbe Maß an Aufmerksamkeit, menschlicher Urteilskraft und Sorgfalt zukommen zu lassen. Wir werden uns also noch lange mit diesem Thema beschäftigen müssen.

Noch nie in unserer kurzen Geschichte kam es so sehr auf unsere Arbeit an wie heute. Aus diesem Grund investieren wir auch weiterhin in die Entwicklung und Optimierung der Richtlinien, Produkte, Tools, Prozesse und Teams, die sich mit Informationsqualität und Verantwortung auf unseren Plattformen befassen. Dass uns dies gelingt, ist von größter Bedeutung für unser Unternehmen und die Gesellschaften, in denen wir tätig sind.

Wir sind zuversichtlich, dass der hier beschriebene Ansatz aktuellen wie künftigen Herausforderungen gerecht werden kann. Um hierfür gewappnet zu sein und die Informationsqualität über alle Produkte und Dienste hinweg sicherzustellen, wenden wir die folgenden vier Instrumente an:

- **Entfernen:** Wir definieren verantwortungsbewusste Richtlinien für alle unsere Produkte und Dienste und leiten Maßnahmen gegen Inhalte und Verhaltensweisen ein, die gegen diese Richtlinien verstoßen. Wir kommen außerdem den gesetzlichen Verpflichtungen zur Löschung von Inhalten nach.

- **Sichtbar machen:** Wir stufen hochwertige Inhalte und zuverlässige Quellen hoch, wo und wenn es darauf ankommt.
- **Reduzieren:** Wir schränken die Verbreitung von potenziell schädlichen Informationen dort ein, wo wir Inhalte vorstellen oder empfehlen.
- **Belohnen:** Wir setzen hohe Qualitäts- und Zuverlässigkeitsstandards für Inhalteerstellerinnen oder -ersteller, die ihre Inhalte monetarisieren oder bewerben möchten.

Wir werden diese Instrumente auch künftig auf unsere Produkte und Dienste anwenden, daraus lernen und sie fortentwickeln und zugleich mit unseren Partnern in Wirtschaft, Zivilgesellschaft und Politik zusammenarbeiten. Gemeinsam werden wir eine Zukunft formen, die die Nutzerrechte schützt, die Werte eines offenen Internets bewahrt und dabei für alle eine bessere und sicherere Online-Umgebung schafft. All diese Anstrengungen stehen in Einklang mit den Gründungswerten von Google, mit den Bedürfnissen unserer Nutzerinnen und Nutzer und den Erfordernissen der Gesellschaften, in denen wir tätig sind.

Ihr Feedback zu diesem Ansatz und den erzielten Fortschritten ist uns jederzeit willkommen.

