

Wie gutefrage.net seine User vor Hate Speech schützt

Interview Kaum eine deutsche Website erreicht hierzulande mehr Internet-Nutzer als gutefrage.net. Politische Entscheidungen richten sich bisher aber nur an den globalen Plattformen aus. Das wollen die Münchner ändern. Im Anschluss an ihr erstes Parlamentarisches Frühstück im Bayerischen Landtag sprach Christian Gehl von marke 41 mit gutefrage-CEO Philipp Graf Montgelas.





Deutschlands größte Frage-Antwort-Plattform

So vielseitig wie keine andere Frage-Antwort-Plattform. Bei uns findest Du schnell neue Perspektiven - egal zu welchem Thema.

Stell Deine Frage

Jetzt anmelden



Vielseitig

Profitiere vom Wissen und der Erfahrung von 1,9 Mio. aktiven Nutzern



Schnell

Du erhältst meist in wenigen Minuten eine Antwort



Lebendig

Unsere Community gibt bis zu 30.000 Antworten täglich

Herr Montgelas, Sie kommen eben von der Gesprächsrunde mit Digital-Politikern aller Parteien im Bayerischen Landtag. Worum ging es da?

Hassreden und Gewaltandrohungen sind leider immer noch weit verbreitet im Internet. Wir haben viel darüber diskutiert, wie User sich effektiv gegen digitale Gewalt wehren können. Dabei ging es mitunter um die Online-Anzeige von Straftaten im Netz.

eine Lösung, die es Usern ermöglicht, online schnell und jederzeit Anzeige gegen digitale Kriminalität zu erstatten.

Wie könnte so eine Lösung in der Praxis aussehen?

Denkbar ist eine zentrale digitale Meldestelle, bei der User Screenshots der inkriminierten Kommentare einstellen und dort auch direkt online bereits die Anzeige erstellen. Bisher bietet zum Beispiel Baden-Württem-

ber die bayerischen Politiker. Wir hoffen, dass wir das durch diesen Austausch heute ändern könnten. Es müsste eine bundesweite Meldestelle installiert werden, um allen Bundesbürgern von überall her den Zugang für eine Online-Anzeige zu ermöglichen. Daran fehlt es aktuell noch.

Welche Themen wurden darüber hinaus bei dem Treffen besprochen?

Mit 25-30 Millionen Unique Usern im Monat gehört gutefrage.net zu den größten Internet-Plattformen in Deutschland. Bei uns gibt es Fragen und Antworten zu jedem nur erdenklichen Thema, oft diskutieren die User ausgiebig miteinander, tauschen Meinungen und Erfahrungen aus. Täglich werden 70 000 Content-Elemente neu eingestellt. Das führt natürlich zu einem hohen Administrationsaufwand. Zwar haben wir neben unseren 75 Mitarbeitern noch 13 bezahlte Moderatoren, doch wir setzen auch auf ehrenamtlich tätige User, die sicherstellen, dass auch um 3 Uhr nachts beleidigende und gewaltandrohende Beiträge aussor-

„Wir sind für eine Lösung, die es Usern ermöglicht, online schnell und jederzeit Anzeige gegen digitale Kriminalität zu erstatten.“

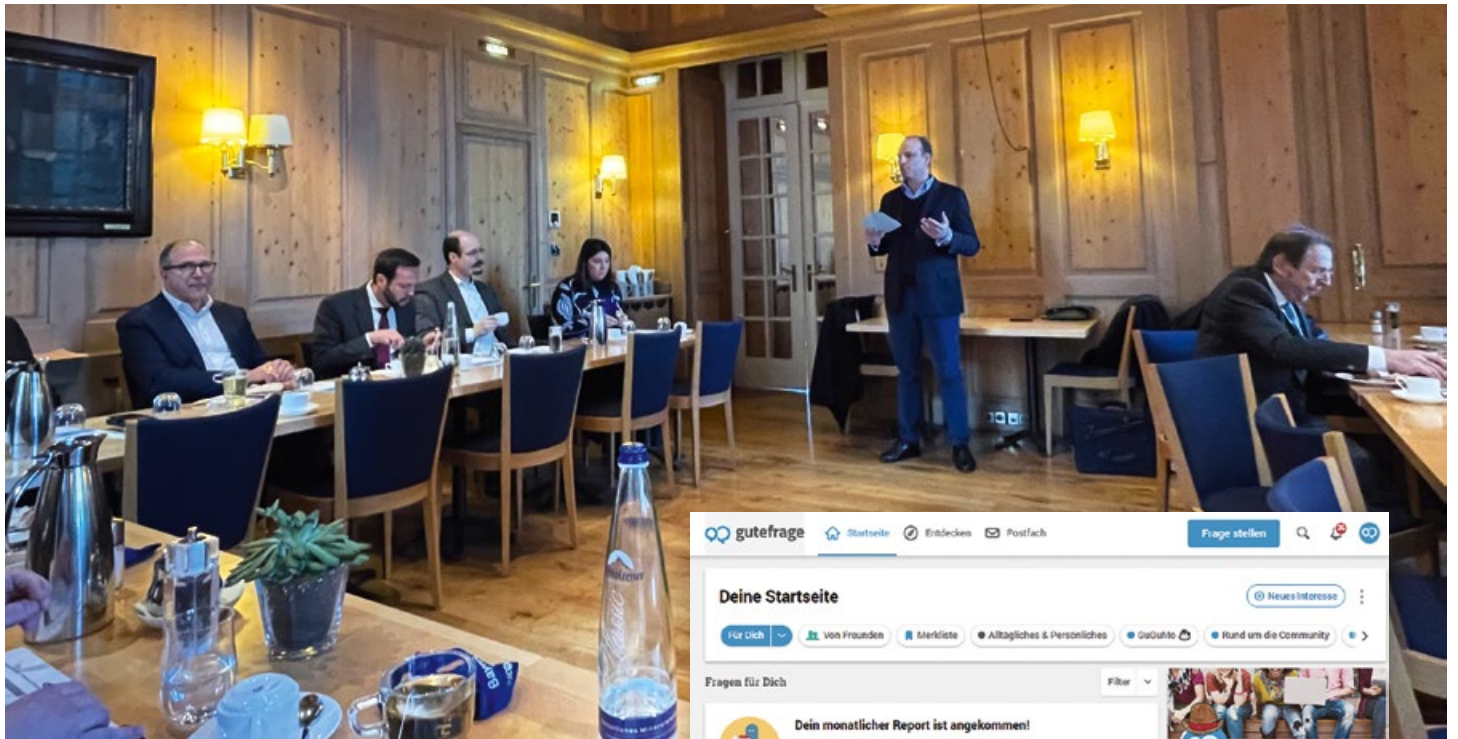
Die es weder bundesweit noch in Bayern, aber doch in Baden-Württemberg schon gibt.

Ja, und einigen Politikern war dies nicht bewusst. Allerdings reagierten Sie mit Zustimmung auf unser Anliegen, die digitale Anzeige von Straftaten im Netz möglich zu machen. Wir sind für

berg über das „Respect“-Portal so eine Möglichkeit an. Sie wird von den Bürgern und Bürgerinnen dort sehr rege genutzt.

Was spricht gegen eine ähnliche Lösung in Bayern?

Die digitale Anzeige in der von uns gewünschten Form ist derzeit noch nicht auf der Agenda



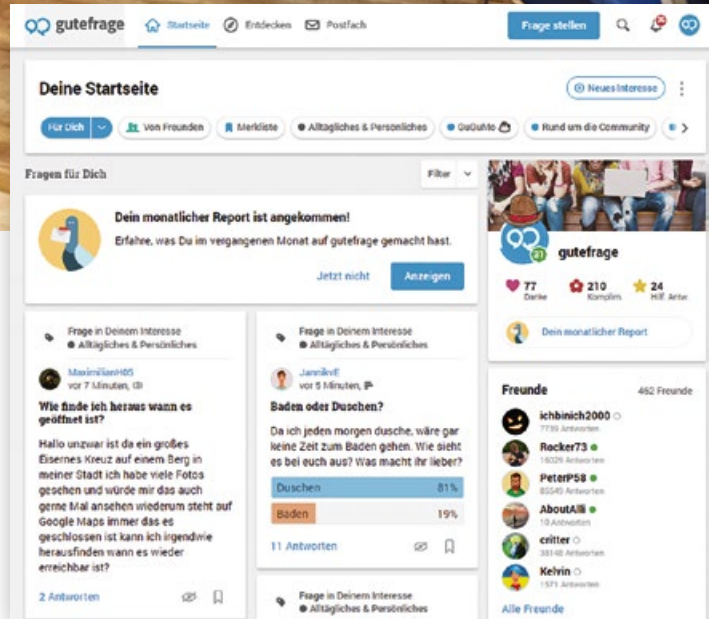
PARLAMENTARISCHES FRÜHSTÜCK

Die führenden Digitalpolitiker Bayerns diskutierten auf Anregung von gutefrage.net über Hate Speech, KI und digitale Ehrenämter.

tiert werden. Wir wollen, dass diese Mitarbeiter jenen gesellschaftlichen Stellenwert erhalten, der anderen ehrenamtlich Tätigen auch zuteil wird. Das heißt, wir haben über Möglichkeiten gesprochen, wie die Kommunen sich hier beteiligen können. Eine Option könnten etwa kostenlose Freibadbesuche, Ermäßigungen beim ÖPNV oder bei kulturellen Veranstaltungen sein.

Könnte da nicht auch KI helfen?

Doch sicher, wir haben entsprechende Lösungen im Einsatz. Alle Beiträge werden auf die Wahrscheinlichkeit hin analysiert, dass sie Hate Speech enthalten oder anderweitig gegen unsere Richtlinien verstoßen. Solche mit einer hohen Wahrscheinlichkeit gehen an unsere bezahlten Analyse-Kräfte zurück und aus deren Entscheidung lernt dann wieder das System. Basis der KI ist ein Trainingsdatenset von Elementen, die wir schon gelöscht haben, diese wendet die KI dann an.



Letzte Instanz bleibt also immer der Mensch?

Ja, aber durch den Menschen wird die KI-basierte Moderation noch besser. Gute Moderation ist das zentrale Element für Online-Communities und beim Digital Services Act, der Anfang 2024 in Kraft tritt, wird ein Fokus auf transparente und nachvollziehbare Moderation gelegt. Ein anderes Thema ist, wie weit KI das Internet grundsätzlich verändern wird. Einfache Wissensfragen zu beantworten, wird wohl eine KI-Domäne, ChatGPT scheint dafür prädestiniert zu sein. Der Kern unserer Plattform ist aber die menschliche Erfahrung, der Meinungsaustausch, die Diskussion, subjektive Einschätzungen – da hat KI Probleme. Fragen wie

„Soll ich mir das neue iPhone holen?“, das wird wohl auch in Zukunft kein Bereich für KI sein.

Wie gehen Sie denn mit reinen Wissensfragen um? Könnte ChatGPT hier zum Einsatz kommen?

Tatsächlich wollen wir ChatGPT demnächst bei Fragen einsetzen, die nach 12 Stunden noch keine Antwort haben. Das sind derzeit etwa 6 Prozent aller User-Fragen. Für bestimmte Themenfelder werden wir das ausprobieren. Jede KI-Antwort werden wir entsprechend kennzeichnen. Wir sind gespannt, wie die Nutzer darauf reagieren.

Herr Montgelas, herzlichen Dank für das Gespräch!