

SEVAL

Schweizerische Evaluationsgesellschaft
Société suisse d'évaluation
Società svizzera di valutazione

Digital Wirkung sichtbar machen – Oder die Frage: Warum Evaluation erst dann ihren Nutzen erreicht, wenn KI auf Qualität trifft

Die SEVAL geht mit der Zeit. Mit der neuen Arbeitsgruppe *Künstliche Intelligenz (KI)* rückt sie ein hochaktuelles Themenfeld ins Zentrum, das die Evaluationspraxis grundlegend verändert. Ziel der Arbeitsgruppe ist es, Chancen und Herausforderungen von KI in der Evaluation zu beleuchten und Wege aufzuzeigen, wie diese Technologien verantwortungsvoll und wirkungsvoll eingesetzt werden können.

Im Rahmen dieser Initiative spricht Kerstin Schwarz (FMSystO GmbH), Co-Leitung der Arbeitsgruppe, mit Expertinnen und Experten an der Schnittstelle von Künstlicher Intelligenz und Evaluation. Einer von ihnen ist **Dr. Christoph Glauser** - Historiker, Politik- und Medienwissenschaftler, der sich früh mit digitaler Transformation befasst hat. Als Gründer der **ArgYou AG** und Leiter des **Instituts für angewandte Argumentenforschung (IFAA)** hat er Pionierarbeit in der computerunterstützten Inhaltsanalyse und datenbasierten Wirkungsforschung geleistet und prägt diese seit über zwei Jahrzehnten. Mit seinem klaren Fokus auf Transparenz und methodische Qualität setzt er dort an, wo viele KI-Systeme an ihre Grenzen stossen. Deshalb zeigen heute seine Entwicklungen darunter die *Find-Maschine* ihre besondere Stärken und finden breite Anwendung in der Evaluation.

Im Interview spricht Dr. Glauser über die Rolle von Künstlicher Intelligenz in der Evaluation, über Chancen und Grenzen algorithmischer Verfahren – und darüber, wie menschliche Urteilskraft und maschinelle Analyse künftig ergänzend zusammenwirken können.



Kontakt

Dr. Christoph Glauser
IFAA Institut
Marktgasse 55
CH-3011 Bern
+41313510220
info@ifaa.ch
www.ifaa.ch

Mehr Infos

[Online-Tool: Find-Maschine](#)

Dr. Christoph Glauser, IFAA Institut für Online-Forschung und Evaluation

Lieber Christoph, herzlich willkommen. Erzähle mir doch, was die ursprüngliche Idee hinter der Entwicklung der Find-Maschine war.

Christoph: Anfang der 2000er-Jahre stellten wir fest, dass wir für unsere Kunden eine Vielzahl kleiner, individueller Messungen zur Analyse von Informations- und Kommunikationsdynamiken im Internet selber durchführen mussten. Das war für uns weder effizient noch wirtschaftlich.

So kam die Idee, ein Instrument zu entwickeln, das einfache Messungen der aktiven Nachfrage im Internet den Kunden einen Zugang zu den Daten zu geben. So entstand die Find-Maschine (abrufbar unter <https://find.argyou.com>), die heute systematisch Daten aus über 14 000 Suchmaschinen, Social-Media-Plattformen und Online-Shops in 191 Ländern und 30 Sprachen integriert. Die technische Umsetzung erwies sich als anspruchsvoll, da sie eine methodisch konsistente Erfassung und Vergleichbarkeit dieser vielen Quellen in möglichst vielen Ländern und Sprachen sicherstellen musste.

Welche konkreten Herausforderungen in der Evaluation wolltest Du mit der Find-Maschine adressieren?

Christoph: Traditionelle Umfragen stoßen zunehmend an ihre Grenzen. Immer weniger Menschen nehmen daran teil, und ihre Antworten sind durch soziale Erwünschtheit verzerrt. Gleichzeitig äußern Menschen aber ihre tatsächlichen Interessen, Bedürfnisse und Einstellungen in digitalen Kanälen – durch Suchanfragen, Kommentare, Beiträge oder Käufe. Mit der Find-Maschine wollte ich ein Instrument schaffen, dass dieses reale und authentische gesellschaftliche Abbild in Form klassischer Selbstauskunft digital sichtbar macht. Das unabhängig von Befragungsbereitschaft und zur Ergänzung klassischer Evaluationsmethoden.

„Traditionelle Umfragen stoßen an ihre Grenzen – die Menschen sagen im Netz längst, was sie wirklich denken“

Dr. Christoph Glauser, IFAA Institut für Online-Forschung und Evaluation

Diese Entwicklung verändert die Art, wie wir Wirkungen erfassen und verstehen können. Die Daten werden zum direkten Spiegel gesellschaftlicher Dynamiken. Welches übergeordnete Ziel verfolgt die Find-Maschine dabei im Kontext von Evaluation und Wissenstransfer?

Christoph: Seit über 20 Jahren geht es bei unseren Wirkungsforschungen darum, objektive und nachvollziehbare, valide und reliable Indikatoren (KPI) für Wirkungsevaluationen aus dem Internet bereitzustellen – und das ganz unabhängig von der spezifischen Fragestellungen oder methodischen Vorannahmen und ohne Bias.

Ein Instrument, dass Wirkung messbar macht, kann ja in vielen Bereichen spannend sein. Wer profitiert derzeit am meisten von der Find-Maschine?

Christoph: Derzeit profitieren insbesondere Akteure, die in komplexen digitalen Umgebungen datenbasierte Entscheidungen treffen müssen. Das sind vor allem große Unternehmen mit einem hohen Anteil an digitalem Geschäft, die ein präzises Verständnis von Markt- und Nachfragedynamiken benötigen.

Dann auch politische Institutionen und wissenschaftliche Einrichtungen – etwa im Rahmen von Forschungsprogrammen wie Horizon Europe oder in den Joint Undertakings der Europäischen Kommission (z. B. Chips JU) wo das Thema Wirkung immer wichtiger wird.

Aber grundsätzlich richtet sich die Find-Maschine an alle, die es „ein bisschen genauer“ wissen wollen: also an Personen und Organisationen, die transparente, nachvollziehbare und methodisch saubere Messungen von Wirkungen bevorzugen und nicht binden Tech-Firmen vertrauen.

Wo datenbasierte Entscheidungen dominieren, rückt auch KI unweigerlich ins Zentrum. Was verstehst du unter „Künstlicher Intelligenz“ – und wo ortest Du die Find-Maschine ein?

Christoph: Ich versteh'e Künstliche Intelligenz (KI) in erster Linie als eine technologische Erweiterung. Es sind zusätzliche Kanäle, die verändern wie Menschen Inhalte suchen, verarbeiten und nutzen. Aus forschungspraktischer Sicht ist für uns relevant, wie Nutzerinnen und Nutzer dieses Kanäle neuen Kanäle tatsächlich verwenden.

Ich sehe die Find-Maschine weniger als KI, sondern als Tool, das mit datenanalytischen Methoden Muster und Wirkungen in grossen Datenmengen sichtbar macht.

KI bringt Chancen und Risiken. Uns interessieren aber vor allem die Chancen, wie neue Anwendungen empirisch beobachtet und bewertet werden können.

Also die „KI spricht mir den Menschen – die Find-Maschine hört ihnen zu“ – ein spannendes Bild. Bedeutet das, dass sie Menschen auch besser versteht? Und worin unterscheidet sich die Find-Maschine von ChatGPT zu anderen klassischen Analyse-Tools?

Christoph: Die Find-Maschine unterscheidet sich grundlegend von klassischen KI-Systemen wie ChatGPT. Während KI mit besseren Algorithmen auf bestehende Daten operieren, erhebt und analysiert die Find-Maschine reale Nachfrageaktivitäten über alle digitale Kanäle hinweg. Unser Ziel ist zu zeigen, wie Themen, Botschaften oder Akteure tatsächlich Resonanz erzeugen. Deshalb ist bei uns Machine Learning, Statistik und Computer Science, Wirkung und Qualität wichtiger als KI. Denn ohne methodische Qualität leistet KI in der Evaluation gar nichts.

„Ohne methodische Qualität leistet KI in der Evaluation gar nichts“

Dr. Christoph Glauser, IFAA Institut für Online-Forschung und Evaluation

Welche zentralen KI-Technologien und Datenverarbeitungsprozesse kommen dazu in der Find-Maschine zum Einsatz?

Christoph: Wir verwenden mehr Machine Learning und entwickeln API-Schnittstellen, unser Fokus ist die Qualität der Daten für Evaluationen und Märkte. Also robuste und vertrauenswürdige Evaluationsergebnisse.

Die Find-Maschine verbindet also datenanalytische Präzision mit KI-gestützten Verfahren – was ist dann das „Herzstück“ der Find-Maschine, und worin liegt ihre eigentliche Intelligenz?

Christoph: Die Find-Maschine hat zum, Glück selber (noch) keine Intelligenz per se. Aber das Herzstück ist die transparente Parametrisierung der Kanäle mit bis zu 120 Parametern, wie Zeitdauer, Währung Keyword, Preise, Nutzung, Nutzerzahlen pro Land, Technologie, usw. Danach geht es wie in jeder Forschung bei uns primär um Validierung und Reliabilitätstests.

In welchen Phasen des Evaluationsprozesses wird die Maschine typischerweise eingesetzt – bei der Evaluationsplanung oder Datenerhebung und -analysephase?

Christoph: Am besten sind wir von Beginn an eingebunden. So können strategische und methodische Fragen, die mit der Wirkungsmessung zusammenhängen, frühzeitig geklärt werden. Idealerweise geschieht das bereits bei der Angebotserstellung oder – aus Auftraggeberperspektive – beim Formulieren der Ausschreibung.

Kannst du ein Beispiel nenne, wo die Find-Maschine gezeigt hat, wie sie Evaluation schneller oder besseren machen kann?

Christoph: Bei digitalen Vorhaben – etwa im E-Government oder bei Online-Kampagnen – ist es immer sinnvoll, zunächst einmal zu messen „woher der digitale Wind weht“. So haben wir beispielsweise bei BELearn zunächst eine Fokusgruppe mit Personen aus Bildungsinstitutionen durchgeführt und anschliessend mit der Find-Maschine die Themen gemessen, um zu verstehen, was für die Nutzerinnen und Nutzer wirklich relevant ist. Neben Empfehlungen konnte so auch einen Geschäftsstrategie entwickelt werden.

Da sich die digitale Welt laufend verändert, sollte eine Evaluation nach einer gewissen Zeit wiederholt werden. Ein iteratives Vorgehen ist dabei nicht nur methodisch sinnvoll, sondern auch viel kostenschonender.

Die Find-Maschine liefert also nicht nur Ergebnisse für Evaluationen, sondern gleichzeitig auch Erkenntnisse, die sich direkt für die Praxis nutzen lassen. Kann sie uns Evaluatorinnen und Evaluatorien auch bei alltäglichen, zeitaufwändigen Aufgaben wie der Konstruktion von Fragebögen oder der qualitativen Datenauswertung unterstützen?

Christoph: Ja, natürlich. Wir können z. B bei der Entwicklung von Fragebögen und Interviewfragen messen, ob die Menschen die Inhalte, die abgefragt werden überhaupt verstehen, diese bereits benutzen oder nicht. Als Basis kann z.B. ein „Basket of Keywords“ dienen oder ein Konzept-Papier, das dann „gemessen“ wird und bereits aufzeigt, welches Wording von Nutzerinnen und Nutzern bereits wie aktiv gesucht wird.

Und wenn die Daten einmal erhoben sind, wie macht die Find-Maschine Wirkung sichtbar?

Christoph: Wir stellen dazu Korrelationen zwischen angebotenen Inhalten und nachgefragten Inhalten (Nutzung/Wirkung) im dreidimensionalen Raum dar. Das machen wir oft für die Sozialwissenschaften, konkreter für die Kommunikationswissenschaft und für die Marktforschung. Benchmarking und Konkurrenzanalysen sind ebenfalls ein Dauerthema; insbesondere im digitalen Bereich.

Ich möchte noch auf die Themen Ethik und Datenschutz zu sprechen kommen. Über den Einfluss von KI auf ethische Fragen weiß man noch wenig. In der Evaluation ist oft vom einem „verantwortungsvollen KI-Einsatz“ die Rede. Zudem ist in der Schweiz seit dem 1. September 2023 das revidierte Datenschutzgesetz (DSG) in Kraft. Wie stellt ihr sicher, dass die Find-Maschine ethisch und rechtlich im Einklang mit Datenschutz und SEVAL-Standards genutzt wird?

Christoph: Wir haben sehr strenge ethische Ausschlusskriterien für nicht identifizierbare Nutzergruppen. Wir sind zudem zu 100% GDPR kompatibel und verwenden keinerlei auf die Privatsphäre bezogenen persönlichen Daten.

Aber wie gelingt es euch gleichzeitig ethische Standards einzuhalten und dennoch nachvollziehbare Ergebnisse zu liefern?

Christoph: Wir legen grossen Wert auf Qualitätssicherung und methodische Strenge. Dazu gehören regelmässige Tests mit gleichbleibenden Samples, aktuelle Nutzerzahlen pro Land sowie Verfahren zur Erkennung von Ausreisern, Outlier Detection. Falls einzelne Kanäle fehlerhafte Werte liefern, führen wir Quer-Validierungen durch. Ergänzend setzen wir kryptische und statistische Verfahren ein, um die Konsistenz und Nachvollziehbarkeit der Ergebnisse sicherzustellen - und noch einiges mehr. Die Nachfrage Daten sind immer alle anonymisiert.

Also nutzt ihr zur Erkennung und Minimierung von Biases oder algorithmische Verzerrungen auch eigene Mechanismen?

Christoph: Ja. Wir arbeiten mit Daten-Visualisierungen. Diese helfen da oft weiter. Es kommt dann aber auch wieder auf die semantischen Spitzfindigkeiten an. Dafür haben wir die Analytik oder arbeiten mit den besten Statistikern, Programmierern und Analysten zusammen, die wir finden können.

Das alles sichert Qualität – aber keine Reproduzierbarkeit, wie sie in der Wissenschaft verlangt wird.

Christoph: Das stimmt. Deshalb legen wir bei jeder einzelnen Anwendung offen, wie viele Kanäle wir messen, und vor allem wie wir messen (Definitionen). Bei den wissenschaftlichen Publikationen und F&E Projekten arbeiten wir oft mit vielen anderen Disziplinen zusammen. Wir liefern die empirischen Daten in Form von Key Performance Indicators (KPIs) oder Key Exploitable Results (KERs). Die Interpretationen machen wir aus weiteren Qualitätssicherungsgründen zumeist zusammen mit den Expertinnen im Thema und den Mitautorinnen.

Was muss gegeben sein, damit Evaluatoren die Find-Maschine in Evaluationen nutzen können?

Christoph: Eigentlich nichts, aber man muss natürlich die Informationen über die Funktionsweise lesen und ein minimales statistisches Flair haben.

Das hört sich an, als ob die Find-Maschine wirklich einiges abnimmt. Aber welche Kompetenzen brauchen Evaluatorinnen und Evaluatoren, um Ergebnisse richtig interpretieren zu können und vor allem mit der rasanten digitalen Entwicklung Schritt zu halten?

Christoph: Neben den zunehmenden Methodenkompetenzen, wird in Zukunft Data Science und minimales Statistikverständnis wichtig. Mit evidenzbasierten Daten zu arbeiten ist nicht trivial, aber meiner Meinung nach je länger je mehr unumgänglich. In Zeiten von „Fake news“ und „Alternative facts“ empfiehlt es sich vorhandenes zuverlässiges Datenmaterial als Grundlage einer sorgfältigen Evaluation aktiv zu nutzen, anstatt sich mit Mutmassungen zu begnügen.

Für die Nutzung von KI empfehle ich klare Regeln. Die meisten KI Tools kann man einstellen, sodass sie nicht immer alles mitlesen. Immer kritisch bleiben, immer mehrere verschiedene Tools und Quellen nutzen und vergleichen. Das kritische selber Denken wird uns die KI wahrscheinlich nie ganz abnehmen können.

„Das kritische selber Denken kann uns die KI nicht abnehmen“

Dr. Christoph Glauser, IFAA Institut für Online-Forschung und Evaluation

Also sollte Evaluation künftig als Zusammenspiel zwischen menschlicher Urteilstkraft und maschineller Analyse verstanden werden. Was spricht dagegen, dass Evaluatorinnen und Evaluatoren eines Tages von KI ersetzt werden?

Christoph: KI erweitert nur den Horizont. Bei News z.B. hat KI ihre deutlichen Grenzen, denn sie erkennt nichts Neues. Siehe dazu auch die Studie der European Broadcasting Union.

Wo siehst du eigentlich die grössten Potenziale moderner Softwaresysteme in der Zukunft der Evaluation?

Christoph: Für schnelle Übersicht über die Nachfrage oder bei den digitalen Wirkungs KPIs und bei der Verbesserung von online Informationen hat moderne Technologie wie unsere Find-Maschine ein grosses Potenzial. Heute werden im Durchschnitt rund 52% aller publizierten Inhalte gar nie gesucht. Man bedenke die horrenden Kosten für Edition, Server, Leitungen, Grafik etc., die trotzdem investiert werden. Umgekehrt werden aber 48% der Inhalte tatsächlich mehr oder weniger häufig gesucht - hier liegt das Potenzial.

In welchen Bereichen hat dich die Leistung oder das Potenzial der Find-Maschine selbst überrascht?

Christoph: In den grossen Horizon Europe Forschungsprojekten. Sobald die Resultate aus unseren Monitorings, Begleitevaluationen und Messungen der digital impact KPIs über die Nachfrage vorliegen, finden plötzlich ganz andere Diskussionen statt als vorher. Oft mit weniger Bauchgefühl natürlich und mehr Fokus auf den Potenzialen.

Welche Risiken oder Grenzen siehst du derzeit bei der Nutzung von KI in der Evaluation?

Christoph: Ein zentrales Risiko beim Einsatz von KI in der Evaluatoin liegt im fehlenden Datenschutz und der mangeldnen Transparenz vieler Systeme. Of ist unklar, wohin die Daten fliessen, wem sie gehören und wie sie weiterverarbeitet werden. Zudem erfüllen die meisten generischen KI-Anwendungen nicht die hohen Qualitätsanforderungen, wie wir sie in unserer Find-Maschine an Objektivität, Reliabilität und Validität stellen. Da ihr Funktionsweise meist intransparent bleibt ,ist eine Ergebnissicherung kaum möglich und die Resultate sind entsprechend wenig verlässlich.

Welche Entwicklungen erwartest du für die Rolle von KI-gestützten Tools wie der Find-Maschine in den nächsten fünf Jahren?

Christoph: Ich kann mir gut vorstellen, dass wir in Zukunft mit der Find-Maschine auch Resultate aus der KI wirkungsvoller machen können, da gibt es noch viele Möglichkeiten. Die Trends und aktuelle Wirkungen können wir messen; Prognosen sind allerdings immer etwas schwierig.

Was wünscht du dir von Institutionen wie der SEVAL im Umgang mit KI in der Evaluation?

Christoph: Neue Technologien knacken oft alte Seilschaften auf. Ich kann mir durchaus vorstellen, dass die SEVAL das Potenzial hat in naher Zukunft multi-, inter-, trans-disziplinärer, etwas internationaler (vielleicht Europäischer) und offener für neue Kollaborationen und neue Methodenmixe zu werden.

Wenn du nun KI und die Find-Maschine in einem Satz zusammenfassen müsstest – was wäre deine wichtigste Botschaft an Evaluierende?

Christoph: Messen statt raten

Herzlichen Dank, lieber Christoph für den spannenden Austausch!