

Austrian Lab for AI Trust* Dossier 5

News-Recommender Systeme: KI-basierte Personalisierung im digitalen Nachrichtenangebot

Executive Summary

Mittlerweile sind das Internet und soziale Medien für viele Menschen eine zentrale Informationsquelle geworden. Um die Fülle an Informationen für einzelne Nutzer:innen zu verringern, setzen soziale Medien und Nachrichtenplattformen sogenannte News-Recommender Systeme ein. Diese filtern Informationen und Nachrichten gezielt nach den individuellen Vorlieben der Nutzer:innen, die anhand ihres bisherigen Such- und Klickverhaltens ermittelt werden. Im Idealfall erhalten Nutzer:innen dann nur Nachrichten, die ihren Interessen entsprechen. Diese Filterfunktion ist zwar bequem, sie stellt jedoch auch große Herausforderungen auf individueller und gesellschaftlicher Ebene dar. Zum einen sind Algorithmen von News-Recommender Systemen so ausgelegt, dass sie die Aufmerksamkeit der Nutzer:innen möglichst lange binden. Denn je länger sie auf einer Plattform bleiben, desto mehr Werbeeinnahmen werden generiert. Vor allem hoch emotionale und negative Nachrichten binden die Aufmerksamkeit und werden daher häufiger vorgeschlagen. Dies kann schließlich zu einem Verlangen nach mehr Nachrichten und im schlimmsten Fall zu einer Nachrichtensucht führen, ein Phänomen, das unter den jungen Menschen mittlerweile ein großes Problem darstellt und als „Doomscrolling“ bezeichnet wird. Darüber hinaus kommt es auf individueller Ebene auch zu Verzerrungen: Das Geschehen um uns herum wird durch das Filtern bestimmter Nachrichten einseitig wahrgenommen, was zudem die Verbreitung von Falschinformationen fördert. Auf gesellschaftlicher Ebene entstehen dadurch sogenannte *Echokammern* und *Filterblasen*, in denen Menschen ähnliche Nachrichten konsumieren, sich gegenseitig in ihren Meinungen bestärken und diese insgesamt verstärken. Ein Austausch unterschiedlicher Meinungen und differenzierte Informationen werden verhindert, was schließlich zur Polarisierung in der Gesellschaft führt.

Das ALAIT-Risikoradar stuft die Nutzung von Nachrichten-Empfehlungssystemen als „hohes Risiko“ (rot, siehe Grafik) ein. Gemäß dem *EU AI-Act* wird das Risiko bei der Nutzung von Nachrichten-Empfehlungssystemen im ALAIT-Risikoradar jedoch als „mittel“ (Stufe 3) eingestuft. In Bezug auf den *Autonomiegrad* sind die eingesetzten Systeme meist voll autonom, da sie unabhängig arbeiten und nur minimale oder gar keine menschlichen Eingriffe erfordern. Die Beteiligung des Menschen be-

ALAIT Risikoradar für KI-basierte Personalisierung im digitalen Nachrichtenangebot (News-Recommender Systeme)



News-Recommender Systeme

Das ALAIT Risikoradar ist ein wissenschaftlich entwickeltes Risikoanalysetool für Künstliche Intelligenz (KI), das KI-Anwendungen kontextbezogen und unter Berücksichtigung ihres technischen Autonomiegrades einstuft und so die Risikosphäre für Anwender:innen auf einen Blick sichtbar macht. Dabei gilt: Je höher das Einsatz-Risiko aus dem Anwendungskontext und je größer der Autonomiegrad des KI-Systems in Bezug auf Entscheidungen, desto riskanter ist der Einsatz einzustufen. Eine erweiterte Klammer weist auf eine Bandbreite in der Risikoeinstufung hin. Ein geringer Autonomiegrad eines KI-Systems bedeutet nicht, dass man sich zurücklehnen kann. Es erfordert eine starke Rolle der Menschen, die es anwenden. (Details zum Stufenmodell s. S. 9f).

schränkt sich auf die langfristige Aufsicht, beispielsweise durch Audits (Stufe 5).

Plattformen und Nutzer:innen können das Risiko senken, indem sie die am Ende dieses Dokuments erklärten Maßnahmen zur Risikominderung umsetzen. Grundsätzlich wäre eine größere Vielfalt an Nachrichten sowie die Entwicklung demokratischer und fairer Systeme entscheidend (siehe Seite 6). Empfehlenswert sind *transparente* und *erklärbare* Systeme sowie eine menschliche Aufsicht, um *Verzerrungen* und politische Polarisierungen zu vermeiden.



Einleitung

In den letzten Jahren sind soziale Medien und Online-Nachrichtenplattformen immer beliebter geworden.¹ Gleichzeitig sind Nutzer:innen inzwischen einer so großen Menge an Nachrichten ausgesetzt, dass es ihnen schwerfällt, den Überblick zu behalten und die für sie interessanten und relevanten Meldungen zu finden. Nachrichtenplattformen und soziale Medien haben mit personalisierten Empfehlungssystemen, sogenannte News-Recommend-Systeme, eine Lösung für dieses Problem gefunden. Die Algorithmen dieser Systeme erlernen aus Reaktionen der Nutzer:innen ihre Vorlieben und können dadurch vorhersagen, welche Nachrichten, Informationen oder Produkte und Dienstleistungen für die Person besonders interessant sind.^{3,4,5} Ein so personalisierter Nachrichtendienst kann die Bedürfnisse und Vorlieben der Nutzer:innen erkennen und die Nutzungserfahrung der Plattform optimieren.

Im Zusammenhang mit News-Recommend-Systemen ist die Unterscheidung zwischen sozialen Medien und traditionellen Nachrichtenplattformen wesentlich. Beide verwenden zwar Algorithmen, um Inhalte vorzuschlagen, arbeiten jedoch mit unterschiedlichen Anreizen, Datensätzen und regulatorischen Auflagen. Social-Media-Plattformen basieren auf einer Aufmerksamkeitsökonomie: Ihre Nachrichten-Empfehlungssysteme sind so optimiert, dass die Nutzer:innen so lange wie möglich auf der Plattform verweilen, um mehr Werbung schalten zu können. Sie priorisieren „Teilen-Wertigkeit“ und „Viralität“.³⁶ Dies belohnt oft emotionale, polarisierende oder sensationelle Inhalte, da diese die meisten Klicks und Kommentare generieren.^{36,38} Nachrichtenplattformen wollen ebenfalls Engagement erzielen, ihr primäres Ziel ist jedoch die Kundenbindung und -loyalität.³⁷ Eine Nachrichtenplattform ist auf ihren Ruf als genaue Informationsquelle angewiesen, um Abonnements zu erhalten. Ihre Empfehlungssysteme für Nachrichten werden daher oft durch redaktionelle Leitplanken eingeschränkt, um die Glaubwürdigkeit der Nachrichtenplattform nicht zu schädigen.

Viele Nachrichtenquellen und Agenturen wie CNN, die BBC und der ORF bieten über Online-Portale jederzeit und überall Zugang zu aktuellen Meldungen. Um mehr Besucher:innen zu gewinnen, setzen diese Portale zunehmend Empfehlungssysteme ein, die die Nutzererfahrung verbessern sollen. Unter „verbesserte Nutzererfahrung“ versteht man Eigenschaften wie Benutzerfreundlichkeit, Nützlichkeit, Wirksamkeit und eine zufriedenstellende Interaktion mit dem System.⁵ Moderne Nachrichten-Empfehlungssysteme passen Inhalte flexibel an den Kontext und individuelle Bedürfnisse an. Sie optimieren

Format und Länge für das jeweilige Endgerät, berücksichtigen Barrierefreiheit (z.B. Screenreader) und liefern ortsbezogene Nachrichten, wie die NZZ mit ihren spezifischen Regionalausgaben für Deutschland und die Schweiz.² Der ORF nutzt in der Videoplattform ORF ON KI-gestützte Empfehlungen, um Inhalte gezielt vorzuschlagen. Auch andere europäische Medien verfolgen einen Public-Service-Ansatz. So entwickelt der finnische Rundfunk Yle einen „Public-Service-Algorithmus“, der Personalisierung mit öffentlichem Mehrwert verbindet. Die Empfehlungen sollen persönlich relevant sein, aber zugleich vielfältige Perspektiven bieten und Nutzer:innen mit überraschenden Inhalten außerhalb der eigenen *Filterblase* konfrontieren. Zugleich legt Yle Wert auf *Transparenz* und Nutzerautonomie: Nutzer:innen können nachvollziehen, warum Inhalte empfohlen werden, und ihre Personalisierung selbst steuern.⁷

Trotz dieser positiven Beispiele gibt es auch für Nachrichtenplattformen Risiken und Probleme. Obwohl zwar Nutzer:innen von personalisierten Empfehlungen profitieren, können dadurch auch ihre Entscheidungsfreiheit beeinträchtigt und ihre Privatsphäre gefährdet werden. Werte wie Privatsphäre, Sicherheit, Verbraucherschutz, Verantwortlichkeit, Solidarität, Gleichheit, Fairness, Transparenz und demokratische Kontrolle stehen häufig im Konflikt mit den in der Plattformarchitektur angelegten ökonomischen Interessen. Da Empfehlungssysteme für Nachrichten entscheiden, welche Inhalte hervorgehoben oder ausgeblendet werden, können sie die öffentliche Meinung beeinflussen und die menschliche Autonomie einschränken.⁸ Deshalb ist es wichtig, diese und weitere Faktoren bei der Bewertung von Empfehlungssystemen für Nachrichten sorgfältig abzuwägen.

In den sozialen Medien gehen die Empfehlungssysteme noch einen Schritt weiter, da sie ein virtuelles Profil der Nutzer:innen verwenden, das sensible Daten wie Alter, Adresse und persönliche Interessen enthalten kann. Mithilfe dieses sogenannten „Social Profiling“ können Plattformen Nutzer:innen beeinflussen, ohne dass diese sich dessen bewusst sind. Diese Praxis ist in der Europäischen Union zwar verboten, jedoch ist die Kontrolle schwierig.⁶ Hinzu kommt die sogenannte Nachrichtensucht oder „Doomscrolling“, die in den letzten Jahren zu einem der größten Probleme junger Menschen geworden ist.

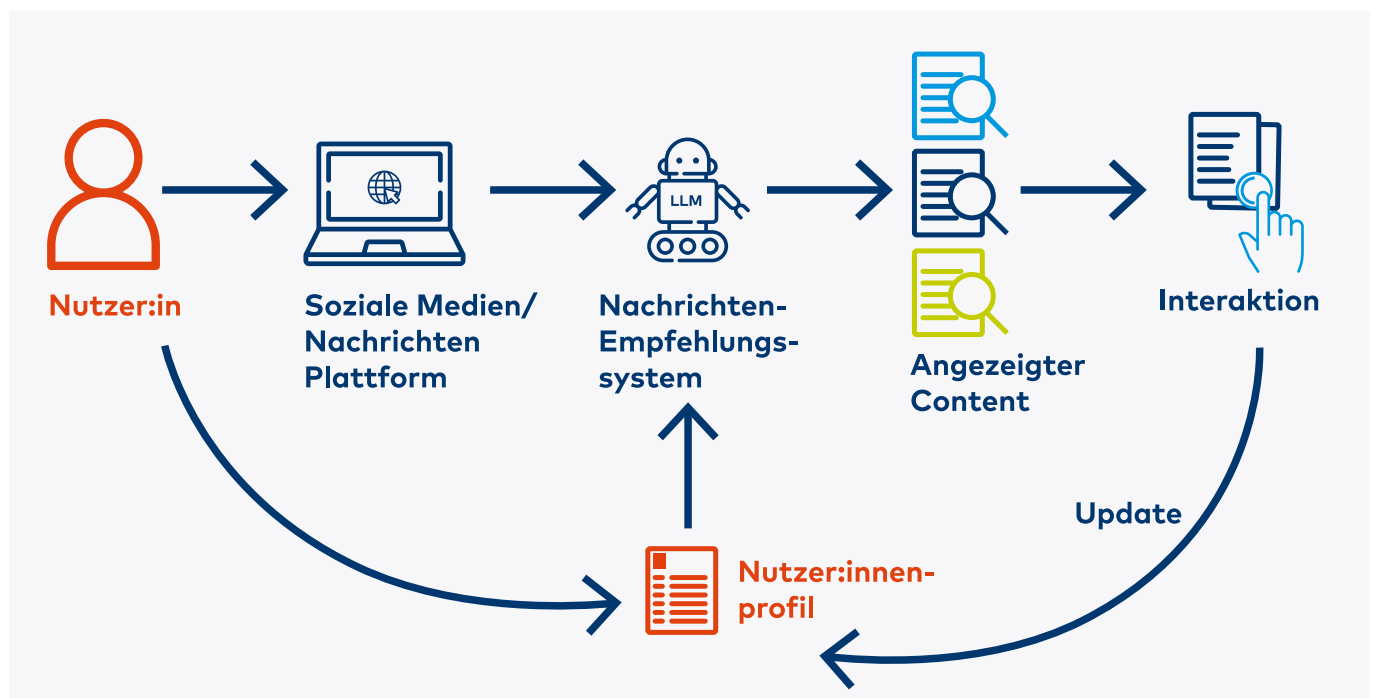
Technologiebeschreibung



Ein Nachrichten-Empfehlungssystem funktioniert so, dass es aus den vielen verfügbaren Artikeln diejenigen auswählt, die für eine bestimmte Person am interessantesten sein könnten. Öffnet die Nutzerin oder der Nutzer die sozialen Medien, nutzt das System das vorhandene Profil, um die Vorlieben und Präferenzen zu ermitteln. Falls noch kein Profil angelegt wurde, verwendet das System frühere Klicks, Lesegewohnheiten, den Standort und sogenannte „Cookies“, um ein temporäres persönliches Profil zu erstellen.^{1,5} Auf dieser Grundlage wählt das System passende Artikel aus und zeigt sie an. Wenn die Person erneut etwas anklickt oder liest, wird ihr Profil aktualisiert. Dadurch lernt das System ständig dazu und kann die Empfehlungen immer besser anpassen.

Dieses „Modeling“ von Nutzer:innen ist grundsätzlich herausfordernd.⁴ Eine der größten Schwierigkeiten von Empfehlungssystemen liegt in der dynamischen Natur der Interessen der Nutzer:innen. Ein präzises Empfehlungssystem sollte in der Lage sein, diese Veränderungen kontinuierlich zu erfassen und das Profil entsprechend anzupassen.^{5,17} Ein weiteres Problem des User Modelings ist der Mangel an direktem Feedback, da viele Nutzer:innen die Artikel nicht bewerten oder kommentieren. Das System muss deshalb aus indirekten Hinweisen wie Klicks oder Lesezeit lernen.^{4,5}

Das folgende Diagramm veranschaulicht das Grundprinzip:



Nachrichten-Empfehlungssysteme lassen sich in zwei Hauptarten unterteilen:

1. Collaborative Filtering (CF): Bei diesem Ansatz werden Rückmeldungen der Nutzer:innen, wie etwa Likes, Dislikes oder Bewertungen, genutzt, um Vorlieben zu erkennen. Auf dieser Basis werden Artikel empfohlen, die anderen Nutzer:innen mit ähnlichem Verhalten gefallen haben.
2. Inhaltsbasierte Filterung: Hierbei werden Artikel anhand ihrer Merkmale (z. B. Text oder Multimedia-Inhalte) miteinander verglichen. Das System empfiehlt Nutzer:innen Inhalte, die den bereits gelesenen oder positiv bewerteten Artikeln ähneln.

Beide Ansätze haben je nach Anwendung Vor- und Nachteile. Um diese auszugleichen, werden häufig hybride Systeme eingesetzt, die Collaborative Filtering und inhaltsbasierte Filterung kombinieren. Hybride Systeme verknüpfen die Ansätze, um die jeweiligen Limitationen auszugleichen. Durch die Verbindung beider Ansätze können sie sowohl das Problem des „Kaltstarts“ – die Schwierigkeit, für neue Nutzer:innen oder neue Inhalte sinnvolle Empfehlungen zu geben, weil noch keine oder zu wenige Nutzungsdaten vorliegen – als auch die Datenknappheit besser bewältigen. Zudem liefern sie präzisere und vielfältigere Empfehlungen, da sie die Zusammenhänge zwischen Nutzer:innen und Inhalten gleichzeitig berücksichtigen.

Die aktuellen Nachrichten-Empfehlungssysteme stützen sich auf Fortschritte im maschinellen Lernen (ML), insbesondere im Bereich der Verarbeitung natürlicher Sprache (Natural Language Processing – NLP), die sich auf Textinhalte beziehen. Im Kern umfasst diese Technologie das Training von Algorithmen anhand großer Datensätze mit gekennzeichneten Inhalten (**Überwachtes Lernen**), sodass sie lernen, Nachrichtenartikel zu kategorisieren.

Deep Learning (DL) basiert hingegen auf künstlichen neuronalen Netzen und kann wichtige Merkmale und Muster direkt aus Rohdaten lernen, ohne dass diese vorher manuell festgelegt werden müssen.

DL-Modelle gelten heute als der Stand der Technik für Nachrichten-Empfehlungssysteme, da sie

1. bei der Verarbeitung von Texten, Bildern oder Videos eine höhere **Genauigkeit** erreichen und

2. die komplexen Wechselwirkungen zwischen Nutzer:innen und Inhalten besser erfassen und
3. das zeitlich veränderliche Nutzerverhalten durch sequenzielle Modellierung berücksichtigen und
4. typische Probleme herkömmlicher Systeme wie Kaltstart oder unzureichende Bewertungsdaten mindern. DL-Modelle können diese Probleme lösen, indem sie automatisch nützliche Informationen aus Nutzer:innen- und Nachrichteninhalten ableiten und so genauere Empfehlungen ermöglichen.

Graphenbasierte Empfehlungssysteme (GRS) verbessern dies noch weiter, indem sie die komplexen Beziehungen zwischen Benutzer:innen, Artikeln und anderen Faktoren innerhalb eines multirelationalen Graphen modellieren. So können sie latente Interessen erfassen, die herkömmliche Modelle möglicherweise übersehen.³²

Chancen

Für Nutzer:innen (Leser:innen):

- Personalisierte Empfehlungen und eine bessere Nutzererfahrung: Personalisierte Empfehlungen reduzieren die Komplexität der Informationslandschaft und verbessern somit die Nutzungserfahrung. Es geht nicht darum, die Menge an Nachrichten künstlich zu begrenzen, sondern die Selektionsarbeit zu erleichtern. Ein gut abgestimmtes System fungiert als digitaler Kurator und hilft den Leser:innen, schneller zum Kern relevanter Themen vorzudringen. Dadurch wird die überwältigende Menge an Informationen verringert, die Informationskomplexität reduziert und die Zufriedenheit der Nutzer:innen gesteigert.^{3,4}
- Barrierefreiheit und Kontextsensitivität: Moderne Nachrichten-Empfehlungssysteme können sich zunehmend an die jeweiligen Nutzungssituationen und Zugangsbedürfnisse anpassen. Sie erkennen, ob Nutzer:innen ein Smartphone oder einen PC verwenden, und stellen darauf abgestimmt Inhalte länger, kürzer, textbasiert oder visuell strukturiert bereit. Ebenso lassen sich regionale Sprachvarianten, lokale Nachrichteninteressen oder Barrierefreiheitsanforderungen (z.B. Screenreader-kompatible Texte für sehbehinderte Benutzer:innen) berücksichtigen.²



Für Plattformen:

- Verbesserte Klickrate und Bindung der Nutzer:innen: Ein gut optimiertes Nachrichten-Empfehlungssystem erhöht die Relevanz der angezeigten Inhalte für die Nutzer:innen und führt somit zu einer höheren Interaktionsrate. Dies führt zu einer längeren Nutzungsdauer und einer höheren Rückkehrbereitschaft.^{3,4,15} Studien zeigen, dass Nachrichten-Empfehlungssysteme die Klickraten auf Nachrichtenportalen um mehr als 30% steigern können.^{9,15} Durch die Beobachtung des Klickverhaltens erkennen Empfehlungssysteme automatisch die Bedürfnisse und Vorlieben der Nutzer:innen, erstellen ein Vorliebenprofil und setzen diese Informationen gezielt für Werbeschaltungen ein.

Herausforderungen und Risiken



Trotz des Versprechens, einen hohen Nutzen für Nutzer:innen und Plattformbetreiber:innen zu bieten, bergen News-Recommend-Systeme eine Vielzahl an Risiken und Herausforderungen auf individueller und gesellschaftlicher Ebene.

Individuelle Ebene:

- **Datenschutz:** Die Leistungsfähigkeit eines Nachrichten-Empfehlungssystems steht in direktem Zusammenhang mit der Menge der verarbeiteten Daten. Das kann jedoch zu Datenschutzproblemen führen, da einige Nachrichtenplattformen und soziale Medien Informationen wie die geografische Lage der Nutzer:innen verwenden.^{2,4} Werden diese Daten gestohlen oder missbraucht, könnte es diesbezügliche Auswirkungen auf die Privatsphäre und Sicherheit der Betroffenen.
- **Nachrichtensucht:** Dabei handelt es sich um den zwanghaften und schädlichen Konsum von Artikeln, Newsfeeds, Radio, Podcasts, Talkshows und E-Mail-Newslettern. Betroffene überprüfen ständig bestimmte Plattformen, was auch als „Doomscrolling“ von Social-Media-Feeds bezeichnet wird. Als Teilmenge der Internet- und Technologieabhängigkeit kann die Nachrichtensucht mit der Zeit zu Veränderungen im Gehirn führen, die die Konzentrationsfähigkeit, die Prioritätensetzung, die Stimmungsregulierung und die Beziehungen zu anderen beeinträchtigen.^{27,38} Die Nachrichtensuchsysteme steigern dieses Risiko durch Personalisierung und Maximierung der Klickrate.

Gesellschaftliche Ebene:

- **Klickraten und Engagement-Maximierung:** Nachrichten-Empfehlungssysteme sind für die Plattformen nicht nur Service-Optimierungstools, sondern auch Werkzeuge einer Aufmerksamkeitsökonomie. Das heißt, sie sind profitorientierte Tools, die darauf ausgelegt sind, das Nutzer:innen-Engagement in Form von Klicks oder der Verweildauer zu maximieren.⁸ Forschungsergebnisse zeigen, dass insbesondere negative oder emotional aufgeladene Nachrichten die stärkste Reaktion hervorrufen und daher häufiger empfohlen oder geteilt werden.^{14,16,17} Dies führt zur Entstehung von systematischen Verzerrungen der Inhalte (siehe nächster Punkt). Anstatt für ausgewogene Informationen zu sorgen, fördern diese Systeme die Polarisierung der Inhalte, was zur Fragmentierung der Gesellschaft führen kann.

- **Reduktion der Vielfalt an Empfehlungen und KI-Bias:** Nachrichten-Empfehlungssysteme zeigen häufig systematische Verzerrungen (KI-Bias) auf.^{1,8,16,22,23,17} Dadurch werden die Überzeugungen und Interessen der Nutzer:innen verstärkt, was wiederum zu einer eingeschränkten Inhaltsdarstellung führt und somit die Entstehung von Echokammern und Filterblasen begünstigt.^{8,9,10,16} Der Begriff „Echokammer“ beschreibt eine Informationsblase, in der Nutzer:innen nur mit Artikeln konfrontiert werden, die ihre bestehenden Überzeugungen bestätigen. Der Begriff „Filterblase“ bezeichnet eine intellektuelle Isolation, den Zustand, von abweichenden Standpunkten und neuen Informationen abgeschnitten zu sein. Dieser Zustand entsteht durch personalisierte Suchanfragen oder Algorithmen.⁵ Dadurch werden Informationen selektiv übernommen und eingeschränkt, was zu einer sehr polarisierten Informationsumgebung führt.^{4,8,9,16} Darüber hinaus ist der digitale Journalismus anfällig für einen Prozess, der als „McDonaldisierung“ bekannt ist. Dabei wird die redaktionelle Autonomie häufig auf standardisierte Inhaltsrezepte ausgerichtet, deren gute Leistung „metrisch bestätigt“ ist. Dieser Wandel stellt Effizienz und Vorhersehbarkeit über die hohen Kosten und den ungewissen Ertrag des investigativen Journalismus und schränkt damit die Diversität und die Qualität der Nachrichten weiter ein.^{28,29}
- **Geringe Vertrauenswürdigkeit, Desinformation und Fake News:** Aufgrund ihres mathematischen Charakters sind KI-Systeme limitiert darin, den Kontext richtig zu interpretieren und die Vertrauenswürdigkeit von Informationen und Quellen abzuschätzen.^{4,5,9} In der Praxis gab es bereits einige Fälle, in denen Empfehlungssysteme für Nachrichten in sozialen Medien manipuliert wurden, um Leser:innen zu Fake News zu lenken.¹⁶ So wurde beispielsweise YouTube in Abstimmung mit der Zeitung „The Guardian“ gezielt manipuliert, um die Leser während der US-Wahlen 2016 zu reißerischen und falschen Nachrichten zu lenken.⁵ Eine Studie in 35 Ländern zeigt zudem, dass die Nutzung sozialer Medien als Hauptnachrichtenquelle mit einem geringeren Vertrauen in Nachrichten einhergeht.¹¹

Technische Ebene:

- **User Modeling von Nutzer:innen:** Eine der größten Schwierigkeiten von Empfehlungssystemen liegt in der dynamischen Natur der Nutzerinteressen. Um spezifische persönliche Empfehlungen von Nach-

richtenartikeln zu erhalten, muss ein umfassendes Nutzungsprofil angelegt werden, das die Präferenzen, Vorlieben, Abneigungen, früheren Interaktionen mit dem System, Interessen und sogar die Beziehungen zu anderen Nutzer:innen berücksichtigt.¹⁷ Mithilfe von User Modeling werden diese Vorlieben ermittelt und Veränderungen kontinuierlich erfasst und das Profil entsprechend aktualisiert.

Dies kann jedoch zu zusätzlichen **Datenschutzproblemen** führen.^{5,17} Ein weiteres Problem beim User Modeling ist der Mangel an direktem Feedback, da viele Nutzer:innen die Artikel nicht bewerten oder kommentieren. Das System muss daher aus indirekten Hinweisen wie Klicks oder Lesezeit lernen.^{4,5}

Empfehlung zum Praxiseinsatz



Für Plattformen:

1. Ausrichtung an bestehenden Richtlinien: Alle KI-Systeme müssen darauf geprüft werden, ob sie den Richtlinien der Datenschutz-Grundverordnung (DSGVO) der Europäischen Union und des **EU AI-Acts** entsprechen. Die Umsetzung und Anwendung des **EU AI-Acts** auf Unionsebene wird von Behördenstellen wie dem Europäischen Büro für Künstliche Intelligenz (AI-Office) und dem Europäischen Gremium für Künstliche Intelligenz (AI-Board) überwacht.¹⁸ Das Gesetz über digitale Dienste (GdD) führt mehrere wichtige Verpflichtungen für Social-Media-Plattformen ein. Damit sollen die Handlungsfähigkeit der Nutzer:innen erhöht und die mit der algorithmischen Kuratierung verbundenen systemischen Risiken verringert werden. So besagt beispielsweise Artikel 27 (**Transparenz** von Empfehlungssystemen), dass Plattformen die wichtigsten Parameter ihrer Empfehlungssysteme klar erläutern müssen, damit Nutzer:innen verstehen können, warum ihnen bestimmte Inhalte angezeigt werden.¹⁹ Artikel 38 (Nicht profilierte Feeds) spezifiziert weiter, dass sehr große Plattformen (VLOPs) mindestens eine Empfehlungssystemoption anbieten müssen, die nicht auf Nutzer:innenprofilen basiert.³⁰ Artikel 40 (Datenzugang für Forscher:innen) regelt darüber hinaus, dass geprüfte Forscher:innen Zugang zu Plattformdaten erhalten, um Studien durchzuführen, die zur Identifizierung systemischer Risiken beitragen (z.B. die Verbreitung illegaler Inhalte).³¹

2. Demokratische und „Diversity-aware“ Systeme: Die Möglichkeit zur Diversifizierung der Nachrichten für die Nutzer:innen ist eine wichtige Grundlage für eine demokratische Darstellung der Nachrichten. Nachrichten-Empfehlungssysteme sollten daher gezielt Diversifizierungsstrategien einsetzen. So können sie nicht nur ähnliche oder bestätigende Inhalte empfehlen, sondern auch abweichende Perspektiven, unterschiedliche Quellen und vielfältige

Themenfelder sichtbar machen. Dadurch können **Filterblasen** und **Echokammern** verringert werden.^{5,8} Zu den Faktoren, die in jedem Empfehlungssystem betrachtet werden müssen, um dieses diversifizieren zu können, gehören unter anderem die Nachrichtenthemen, Schreibstile, Tags, Perspektiven, Kontexte und Ideologien.^{5,8,12} Diversity-aware Systeme garantieren jedoch nicht, dass sie bei den Leser:innen wirken. Empirische Studien sind sich nicht einig, ob technologisches Nudging (engl. für „Stupser“), das Verhalten der Nutzenden auf vorhersehbare Weise lenkt und verschiedene Inhalte anzeigt, in Nachrichtensystemen tatsächlich wirksam ist.²⁶

3. Datenentzerrung: Um qualitativ hochwertige Nachrichtempfehlungen zu erstellen, ist es wichtig, wirksame Methoden zur Beseitigung des Einflusses verschiedener Arten von Verzerrungen auf Empfehlungsergebnisse zu entwickeln.¹ Einheitliche, neu ausbalancierte und sorgfältig ausgewählte Daten sind allerdings der Ausgangspunkt für Strategien zur Verringerung von **Verzerrungen**.²³

Für Nutzer:innen:

1. Schaffung und Stärkung von KI- und Medienwissen: Es ist wichtig, sich über KI-Technologien zu informieren, die richtigen Kompetenzen aufzubauen und seine Rechte zu kennen.²⁰ In Österreich besteht mit der Rundfunk- und Telekom Regulierungs-GmbH (RTR) eine Informations- und Anlaufstelle für die breite Öffentlichkeit. Die RTR-KI-Servicestelle unterstützt Bürger:innen bei Fragen zur Transparenz, zu Rechten im Umgang mit KI-Systemen oder zur Überprüfung von KI-Entscheidungen.²¹ Nutzer:innen können sich auch spezifisch über Methoden von Nachrichten-Empfehlungssystemen informieren. Die Informationsgruppe Algotransparency klärt

Bürger:innen darüber auf, wie sie bei einer zunächst neutralen Suche auf YouTube in jeder nachfolgenden Phase des Empfehlungszyklus zu zunehmend voreingenommenen Informationen gelenkt werden.¹³ Das European Centre for Algorithmic Transparency (ECAT) unterstützt die Europäische Kommission bei ihrer ausschließlichen Aufsichts- und Durchsetzungsfunktion hinsichtlich der systemischen Verpflichtungen, die gemäß dem Digital Services Act (DSA) für sehr große Online-Plattformen (VLOPs) und sehr große Online-Suchmaschinen (VLOSEs) gelten.²⁴ Darüber hinaus ist die Förderung der Medienkompetenz von Kindern und Jugendlichen im Kontext der sozialen Medien unerlässlich.

2. **Faktenchecks:** Obwohl die Kontrolle der Informationsqualität eine sehr komplizierte Aufgabe ist, gibt es bereits Möglichkeiten, dies zu tun. Beispielsweise können Faktenchecks zu komplexen und umstrittenen Themen hilfreich sein. So ist es beispielsweise der Austria Presse Agentur (APA) ein Anliegen, den Leser:innen durch Faktenchecks einen Einblick in die Recherchemethoden des digitalen Zeitalters zu geben. Mithilfe der Faktenchecks können sich Leser:innen auch zu komplexen und umstrittenen Themen eine Meinung bilden, die auf rationalen Zugängen und Fakten basiert.²⁵ Andere Medienhäuser wie CORRECTIV. Faktencheck, die sich unter anderem mit investigativem Journalismus beschäftigen, bieten ebenfalls Dienstleistungen im Bereich Faktencheck an.
3. **Kritisches Denken:** Die Nutzer:innen sollen die Empfehlungen des Nachrichten-Empfehlungssystems kritisch hinterfragen und aktiv nach alternativen

Sichtweisen suchen. Dazu wird vorgeschlagen, regelmäßig Inhalte aus verschiedenen Medienquellen mit unterschiedlichen politischen Perspektiven und aus internationalen Kontexten zu lesen. Viele Plattformen bieten heute Funktionen, mit denen Nutzer:innen ihre Interessen anpassen oder gezielt bestimmte Quellen hinzufügen bzw. ausschließen können. Durch solche Einstellungen kann die algorithmische Vorauswahl korrigiert und das Risiko von Filterblasen reduziert werden.^{5,8,9}

4. **Zeitbeschränkung:** Im Allgemeinen kann die Festlegung von Zeitlimits für die Nutzung sozialer Medien dazu beitragen, die mit Nachrichten-Empfehlungssystemen verbundenen Risiken insgesamt zu verringern. Insbesondere kann dadurch das Risiko einer Nachrichtenabhängigkeit und des „Doomscrollings“ verringert werden.³⁵ Studien haben einen Zusammenhang zwischen übermäßiger Nutzung sozialer Medien und Depressionen aufgezeigt.^{33,35} Daher kann die Begrenzung der Nutzung sozialer Medien das allgemeine Wohlbefinden verbessern. Es ist besonders wichtig, Kindern und Jugendlichen keinen uneingeschränkten Zugang zu sozialen Medien zu gewähren. In dieser entscheidenden Phase bilden sich digitale Verhaltensweisen wie das Knüpfen von Online-Freundschaften, das Kommentieren, Liken und Reagieren auf bestimmte Online-Inhalte heraus.³⁴ Die meisten Social-Media-Apps und Smartphones bieten derzeit Einstellungen, die Tagespläne ermöglichen. Dabei wird die Nutzung nach voreingestellten Zeiträumen pausiert, um einen bewussteren Konsum im Einklang mit persönlichen Zielen zu fördern.

Wichtige Begriffe

Datenschutz in der KI: Die Gesamtheit der Praktiken und Bedenken im Zusammenhang mit der ethischen Erfassung, Speicherung und Nutzung personenbezogener Daten durch Systeme der künstlichen Intelligenz.

Echokammer: Bezieht sich auf eine Informationsblase um einen Nutzer bzw. eine Nutzerin herum, in der nur Informationen zu finden sind, die die bestehenden Überzeugungen der Nutzerin bzw. des Nutzers bestätigen.

Erklärbare KI (XAI): Bezieht sich auf Methoden und Techniken, die es ermöglichen, die Entscheidungen und Ergebnisse von KI-Systemen besser zu verstehen und nachzuvollziehen.

Filterblase: Entspricht der intellektuellen Isolation, die durch personalisierte Suchanfragen oder Algorithmen verursacht wird, um selektiv die Informationen anzunehmen, die eine Person sehen möchte.

Gesetz der Europäischen Union über künstliche Intelligenz (EU AI-Act): Eine europäische Verordnung über künstliche Intelligenz (KI) – die erste umfassende Verordnung über KI von einer großen Regulierungsbehörde überhaupt. Sie konzentriert sich insbesondere auf KI-Systeme mit hohem Risiko.

KI-Autonomie: Die Fähigkeit eines KI-Systems, eine Reihe von Zielen unter einer Reihe von Unsicherheiten in ihrer Umgebung selbstständig und ohne externe Eingriffe zu erreichen.

KI-Genauigkeit (AI-Accuracy): Bezieht sich auf die Fähigkeit eines KI-Systems, korrekte Vorhersagen oder Entscheidungen zu treffen. Sie ist ein wichtiger Maßstab für ihre Leistung und entscheidend für die Bestimmung ihrer Wirksamkeit und Zuverlässigkeit.

Systematische Verzerrungen (Bias): Bias ist eine systematisch unterschiedliche Behandlung bestimmter Objekte, Personen oder Gruppen im Vergleich zu anderen. Behandlung ist jede Art von Handlung, einschließlich Wahrnehmung, Beobachtung, Darstellung, Vorhersage oder Entscheidung.

Transparenz: Bedeutet, dass die Funktionsweise, Entscheidungsprozesse und Einsatzbereiche eines KI-Systems nachvollziehbar, erklärbar und offen zugänglich sind – für Entwickler:innen, Nutzer:innen und andere Stakeholder.

Überwachtes Lernen (supervised learning): Ein Teilgebiet des maschinellen Lernens, bei dem die Trainingsdaten vorab gekennzeichnet werden und das System das Muster zwischen dem Inhalt und der Kennzeichnung erlernt. Die Aufgabe eines solchen KI-Systems besteht darin, eine Beziehung zu finden, die jede Eingabe des Trainingsatzes (die Daten) einer Ausgabe (der Kennzeichnung) zuordnet.

Erklärung Stufenmodell des ALAIT Risikoradars

Im ALAIT KI-Risikoradar wird die Beziehung zwischen Anwendungsrisiko und Autonomie eines KI-Systems dargestellt. Die Risikostufen stützen sich auf das EU KI-Gesetz (*EU AI-Act*), insbesondere auf Artikel 6 und Annex III, die sich mit risikoreichen Anwendungsbereichen von KI befassen. Geringere System-Autonomie und Anwendungsrisiken werden durch kältere Farben (blau) und höhere System-Autonomie und Anwendungsrisiken durch wärmere Farben (rot) dargestellt.

Der Farbwechsel vermittelt das erhöhte Risiko solcher Entscheidungen. Mithilfe dieser Farbskala lässt sich das Gesamtrisiko erkennen: Violett und dunkelrot – sehr hoch, rot und dunkelorange – hoch, hellorange und Gelbtöne – mittel, Blautöne – geringes Gesamtrisiko. Im Idealfall sollten hohe Anwendungsrisiken und System-Autonomie vermieden oder nur nach sehr sorgfältiger Abwägung eingesetzt werden.

Autonomiegrad des KI-Systems

Stufe 1: Keine Autonomie

KI ist ein passives Werkzeug; Menschen treffen alle Entscheidungen und leiten Maßnahmen ein.

Beispiel: Diagnosesysteme, die medizinische Rohdaten anzeigen oder die Daten analysieren (ohne Empfehlungen!)

Empfohlene Anwendungsfälle: Szenarien mit hohen Risiken oder bei denen ethische Entscheidungen von entscheidender Bedeutung sind (z.B. medizinische Diagnostik, Justizsystem).

Stufe 2: Geringer Autonomiegrad (Human-in-the-Loop)

Die KI gibt Empfehlungen oder Optionen, aber der Benutzer:innen bleibt für die Auswahl und Genehmigung von Maßnahmen verantwortlich.

Beispiel: KI schlägt optimale Routen für die Logistik vor oder Empfehlungssysteme im E-Commerce.

Empfohlene Anwendungsfälle: Aufgaben mittlerer Komplexität mit mäßigen Risiken (z.B. Optimierung der Lieferkette).

Stufe 3: Mittlerer Autonomiegrad (Human-on-the-Loop)

Die KI führt bestimmte Aufgaben autonom aus, wobei Menschen in Ausnahmefällen eingreifen.

Beispiel: KI-gestützte Fertigungsprozesse, bei denen das System Maschinen steuert, aber Nutzende bei Anomalien eingreifen.

Empfohlene Anwendungsfälle: Szenarien, in denen eine kontinuierliche menschliche Beteiligung nicht erforderlich ist, kritische Risiken jedoch eine menschliche Überwachung erfordern (z.B. industrielle Automatisierung, Überwachung von Finanztransaktionen).

Stufe 4: Hoher Autonomiegrad (Human in Control)

Das KI-System arbeitet weitgehend autonom, erlaubt es den Benutzern jedoch, es selbst zu übersteuern, um unerwünschte Ergebnisse zu vermeiden.

Beispiel: Autonome Fahrzeuge

Empfohlene Anwendungsfälle: Umgebungen mit geringem bis mittlerem Risiko (z.B. Logistik, einfaches Verkehrsmanagement).

Stufe 5: Vollständige Autonomie mit minimaler Aufsicht

Das KI-System arbeitet unabhängig und erfordert nur minimale oder gar keine menschliche Intervention. Die Beteiligung des Menschen beschränkt sich auf die langfristige Aufsicht (Audits).

Beispiele: Autonome landwirtschaftliche Maschinen, KI für die Stromnetzverteilung, U-Bahnen, Flughafenbahnen

Empfohlene Anwendungsfälle: Umgebungen mit geringen Sicherheits- oder ethischen Risiken und hoher Zuverlässigkeit des KI-Systems (z.B. sich wiederholende Aufgaben in kontrollierten Umgebungen).

Anwendungsbereich-Risiko

Stufe 1: Minimales Risiko

Das KI-System hat keine Auswirkungen auf die Benutzer:innen oder die Entscheidungsfindung.

Beispiele: Filter, NPCs, Empfehlungsalgorithmen ohne schwerwiegende Folgen (DeepL, andere Übersetzungssysteme)

Kriterien: Keine direkte Auswirkung auf die Hochrisikobereichen des [EU AI-Acts](#).

Stufe 2: Begrenztes Risiko

KI-Systeme, die mit Benutzer:innen interagieren, aber keine Entscheidungen mit hohen Risiken treffen. Das Risiko steigt, wenn es an Transparenz über die Beteiligung von KI mangelt.

Beispiele: Chatbots und KI-generierte Inhalte ohne Offenlegung, einfache Automatisierungsaufgaben.

Kriterien: Bereiche, die nicht in der Liste der „hohen Risiken“ des EU AI-Acts enthalten sind.

Stufe 3: Mittleres Risiko

KI-Systeme haben keine besonderen Auswirkungen auf einzelne Personen, aber sie entfalten Wirkung auf kollektiver oder gesellschaftlicher Ebene.

Beispiele: Generative KI wie ChatGPT und andere Systeme, die indirekt die Umgebung beeinflussen können, in der sie eingesetzt werden.

Kriterien: KI-Systeme, die für die öffentliche Nutzung verfügbar sind und das Potenzial haben, bestehende Gepflogenheiten zu beeinflussen und langfristig zu verändern.

Stufe 4: Hohes Risiko

Jeder Algorithmus, der in den laut EU AI-Act „Hochrisikobereichen“ angewendet wird oder direkte Auswirkungen auf einzelne Personen hat.

Beispiele: Medizin, Biometrie, kritische Infrastruktur, Bildung und Berufsausbildung, Beschäftigung, Zugang zu Dienstleistungen im öffentlichen Sektor, Strafverfolgung, Migration.

Kriterien: Zugehörigkeit zum „Hochrisikobereich“ des EU AI-Acts, nur wenn die Regeln für Transparenz und Datenqualität eingehalten werden.

Stufe 5: Extremes Risiko

Jeder Algorithmus, der in den laut EU AI-Act „Hochrisikobereichen“ angewendet wird.

Beispiele: Medizin, Biometrie, kritische Infrastruktur, Bildung und Berufsausbildung, Beschäftigung, Zugang zu Dienstleistungen im öffentlichen Sektor, Strafverfolgung, Migration.

Kriterien: Zugehörigkeit zum „Hochrisikobereich“, wenn die Regeln für Transparenz und Datenqualität NICHT eingehalten werden.

Quellen

- 1 Wu, C., Wu, F., Huang, Y., & Xie, X. (2023). Personalized News Recommendation: Methods and Challenges. *ACM Transactions on Information Systems*, 41(1), 1–50. <https://doi.org/10.1145/3530257>
- 2 Coenen, A. (2019, October 17). How The New York Times is Experimenting with Recommendation Algorithms. Medium. <https://open.nytimes.com/how-the-new-york-times-is-experimenting-with-recommendation-algorithms-562f78624d26>
- 3 Meng, X., Huo, H., Zhang, X., Wang, W., & Zhu, J. (2023). A Survey of Personalized News Recommendation. *Data Science and Engineering*, 8(4), 396–416. <https://doi.org/10.1007/s41019-023-00228-5>
- 4 Feng, C., Khan, M., Rahman, A. U., & Ahmad, A. (2020). News Recommendation Systems—Accomplishments, Challenges & Future Directions. *IEEE Access*, 8, 16702–16725. <https://doi.org/10.1109/ACCESS.2020.2967792>
- 5 Raza, S., & Ding, C. (2022). News recommender system: A review of recent progress, challenges, and opportunities. *Artificial Intelligence Review*, 55(1), 749–800. <https://doi.org/10.1007/s10462-021-10043-x>
- 6 Article 5: Prohibited AI Practices | EU Artificial Intelligence Act. (n.d.). Retrieved 12 February 2026, from <https://artificialintelligenceact.eu/article/5/>
- 7 Hilden, K.-M. (2025, May 23). Yle is developing a Public Service Algorithm to ensure the relevance and transparency of its services. *Yle Viestintä*. <https://yle.fi/aihe/a/20-10008750>
- 8 Parcu, P. L., Bird, W., Taina, B., & Mello, P. C. (2023, February). Pluralism of News and Information in Curation and Indexing Algorithms. *Forum Information & Democracy*. <https://informationdemocracy.org/publications/pluralism-of-news-and-information-in-curation-and-indexing-algorithms/>
- 9 Elahi, M., Jannach, D., Skjærven, L., Knudsen, E., Sjøvaag, H., Tolonen, K., Holmstad, Ø., Pipkin, I., Throndsen, E., Stenbom, A., Fiskerud, E., Oesch, A., Vredenberg, L., & Trattner, C. (2022). Towards responsible media recommendation. *AI and Ethics*, 2(1), 103–114. <https://doi.org/10.1007/s43681-021-00107-7>
- 10 Raza, S. (2024). Bias Reduction News Recommendation System. *Digital*, 4(1), 92–103. <https://doi.org/10.3390/digital4010003>
- 11 Kalogeropoulos, A., Suiter, J., Udris, L., & Eisenegger, M. (2019). News Media Trust and News Consumption: Factors Related to Trust in News in 35 Countries. *International Journal of Communication*, 13, 22–22.
- 12 Wang, R., Liesaputra, V., & Huang, Z. (2025). A Survey on LLM-based News Recommender Systems (No. arXiv:2502.09797). *arXiv*. <https://doi.org/10.48550/arXiv.2502.09797>
- 13 AlgoTransparency. (n.d.). AlgoTransparency. Retrieved 17 October 2025, from <https://algotransparency.org/>
- 14 Robertson, C.E., Pröllochs, N., Schwarzenegger, K. et al. Negativity drives online news consumption. *Nat Hum Behav* 7, 812–822 (2023). <https://doi.org/10.1038/s41562-023-01538-4>
- 15 Jannach, D., & Jugovac, M. (2019). Measuring the business value of recommender systems. *ACM Transactions on Management Information Systems (TMIS)*, 10(4), 1-23.
- 16 Bojic, L. (2024). AI alignment: Assessing the global impact of recommender systems. *Futures*, 160, 103383.
- 17 Goanta, C., & Spanakis, G. (2022). The commercial unfairness of recommender systems on social media. In *Artificial intelligence and the media* (pp. 148-170). Edward Elgar Publishing.
- 18 KI-Servicestelle der RTR. (n.d.). RTR. Retrieved 16 September 2025, from <https://www.rtr.at/ki-servicestelle>
- 19 Artikel 27 Transparenz der Empfehlungssysteme ~ Digital-Services-Act (DSA). (n.d.). *Digitale-Dienste-Gesetz*. Retrieved 12 November 2025, from <https://gesetz-digitale-dienste.de/dsa/artikel-27/>
- 20 Thapa, B. (2019). Predictive Analytics and AI in Governance: Data-driven government in a free society. *The European Liberal Forum*. <https://liberalforum.eu/publication/predictive-analytics-and-ai-in-governance-data-driven-government-in-a-free-society/>

- 21 KI-Servicestelle der RTR. (n.d.). RTR. Retrieved 16 September 2025, from <https://www.rtr.at/ki-servicestelle>
- 22 Tao Qi, Fangzhao Wu, Chuhan Wu, and Yongfeng Huang. 2021. PP-Rec: News recommendation with personalized user interest and time-aware news popularity. In *ACL*. 5457–5467.
- 23 Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2021). Bias and Debias in Recommender System: A Survey and Future Directions (No. arXiv:2010.03240). arXiv. <https://doi.org/10.48550/arXiv.2010.03240>
- 24 European Centre for Algorithmic Transparency—European Centre for Algorithmic Transparency. (2025, October 24). https://algorithmic-transparency.ec.europa.eu/index_en
- 25 APA. (n.d.). Faktencheck | News Verifizieren | APA - Austria Presse Agentur. Retrieved 24 November 2025, from <https://apa.at/service/faktencheck-2/>
- 26 The truth behind filter bubbles: Bursting some myths | Reuters Institute for the Study of Journalism. (2020, January 24). Retrieved 5 Dezember 2025, from <https://reutersinstitute.politics.ox.ac.uk/news/truth-behind-filter-bubbles-bursting-some-myths>
- 27 Genesung von Nachrichtensucht – Anonyme Internet- und Technologieabhängige. (n.d.). Die anonymen Internet- und Technologiesüchtigen. Retrieved 16 February 2026, from <https://internetaddictsanonymous.org/de/internet-und-technologieabhangigkeit/informationssucht/nachrichtensucht/>
- 28 Çifçi, M. E., & Ayhan, B. (2025). 'The first to report wins': The McDonaldization of digital journalism in Turkey. *Journal of Applied Journalism & Media Studies*, 14(1), 33–51. https://doi.org/10.1386/ajms_00100_1
- 29 Hastuti, H., Maulana, H. F., Lawelai, H., & Suherman, A. (2025). Algorithmic influence and media legitimacy: A systematic review of social media's impact on news production. *Frontiers in Communication*, 10. <https://doi.org/10.3389/fcomm.2025.1667471>
- 30 Article 38, the Digital Services Act (DSA). (n.d.). Retrieved 19 February 2026, from https://www.eu-digital-services-act.com/Digital_Services_Act_Article_38.html
- 31 Article 40, the Digital Services Act (DSA). (n.d.). Retrieved 19 February 2026, from https://www.eu-digital-services-act.com/Digital_Services_Act_Article_40.html
- 32 Raza, S., Rahman, M., Kamawal, S., Toroghi, A., Raval, A., Navah, F., & Kazemeini, A. (2025). A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice (arXiv:2407.13699; Version 4). arXiv. <https://doi.org/10.48550/arXiv.2407.13699>
- 33 Hunt, M., All, K., Burns, B., & Li, K. (2021). Too Much of a Good Thing: Who We Follow, What We Do, And How Much Time We Spend on Social Media Affects Well-Being. *Journal of Social and Clinical Psychology*, 40(1), 46–68. <https://doi.org/10.1521/jscp.2021.40.1.46>
- 34 Charmaraman, L., Lynch, A. D., Richer, A. M., & Grossman, J. M. (2022). Associations of early social media initiation on digital behaviors and the moderating role of limiting use. *Computers in Human Behavior*, 127, 107053. <https://doi.org/10.1016/j.chb.2021.107053>
- 35 Dr. A.Shaji George, A. S.Hovan George, Dr. T.Baskar, & Dr. M.M.Karthikeyan. (2024). Reclaiming Our Minds: Mitigating the Negative Impacts of Excessive Doomscrolling. <https://doi.org/10.5281/ZENODO.13737987>
- 36 Heitmayer, M. (2025). The Second Wave of Attention Economics. Attention as a Universal Symbolic Currency on Social Media and beyond. *Interacting with Computers*, 37(1), 18–29. <https://doi.org/10.1093/iwc/iwae035>
- 37 Goodman, E. (2016, October 26). Editors vs algorithms: Who do you want choosing your news? | Reuters Institute for the Study of Journalism. <http://reutersinstitute.politics.ox.ac.uk/news/editors-vs-algorithms-who-do-you-want-choosing-your-news>
- 38 Bhargava, V. R., & Velasquez, M. (2021). Ethics of the Attention Economy: The Problem of Social Media Addiction. *Business Ethics Quarterly*, 31(3), 321–359. <https://doi.org/10.1017/beq.2020.32>

Projekt ALAIT

Das Austrian Lab for AI Trust (ALAIT) ist ein vom österreichischen Bundesministerium für Innovation, Mobilität und Infrastruktur (BMIMI) gefördertes Forschungs- und Entwicklungs-Projekt zur Schaffung von Vertrauen durch Wissen im Bereich Künstliche Intelligenz (KI). Das Projekt ALAIT zielt darauf ab, Interessierte und wichtige gesellschaftliche Gruppen zu befähigen, KI-Technologien verantwortungsvoll zu nutzen und ethische sowie qualitativ hochwertige Standards für den Einsatz von AI zu etablieren.

Das Projekt wird von **winnovation** geleitet (Gertraud Leimüller und Brigitte Ömer-Rieder) und im Konsortium mit **leiwand.ai** (Rania Wazir und Silvia Wasserbacher-Schwarzer), **TU Wien** (Sabine Köszegi und Ilya Faynleyb) und **Austria Presse Agentur – APA** (Verena Krawarik und Sophia Marecek) umgesetzt.

Die ALAIT-Dossiers sind auf der Projekthomepage abrufbar: <https://science.apa.at/project/alait/>

Die Inhalte des Dossiers entsprechen dem aktuellen Stand der Technik und wurden sorgfältig nach wissenschaftlichen Kriterien erstellt. Sie dienen jedoch nicht als rechtsverbindliche Auskunft oder Beratung.

Impressum

Medieninhaberin und Herausgeberin:
winnovation consulting gmbh
Linke Wienzeile 42/1, Top 5
1060 Vienna

Dieses Dossier steht unter der Creative Commons Lizenz CC BY-NC-ND 4.0
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>
(Namensnennung-Nicht kommerziell-Keine Bearbeitungen 4.0 International)

Veröffentlicht 2026, gefördert durch das Bundesministerium für Innovation, Mobilität und Infrastruktur.



Danksagung:

Wir danken folgenden Expert:innen für ihr hilfreiches Feedback zu Vorversionen dieses Dossiers:
Julia Neidhart, Peter Knees.