



Fachhochschule Graubünden  
University of Applied Sciences

# Churer Schriften zur Informationswissenschaft

Herausgegeben von  
Wolfgang Semar, Bernard Bekavac, Ivo Macek, Armando Schär

---

Arbeitsbereich Bachelor of Science  
in Digital Business Management

**Schrift 176**

## **Governance von Social-Media-Algorithmen**

Analyse der Aufsicht und Regulation von

**Rade Jevdenic**

---

Chur 2024



# **Churer Schriften zur Informationswissenschaft**

Herausgegeben von Wolfgang Semar,  
Bernard Bekavac, Ivo Macek, Armando Schär

Schrift 176

## **Governance von Social-Media-Algorithmen**

Analyse der Aufsicht und Regulation von

**Rade Jevdenic**

Diese Publikation entstand im Rahmen einer Thesis zum Bachelor of Science in Digital Business Management.

Referent: Müller Elham

Korreferent: Aschwanden Michael

**Verlag:** Fachhochschule Graubünden

**ISSN:** 1660-945X

**Ort, Datum:** Chur, März 2024



## Abstract

Im Februar 2024 treten mit dem Digital Services Act neue Bestimmungen für Social-Media-Unternehmungen, die auf dem europäischen Markt ihre Onlinedienste anbieten, in Kraft. Der Digital Services Act sieht dabei die intensivste Aufsicht und Regulation für die marktführenden Dienste Facebook, TikTok, Twitter, Instagram und Youtube vor. Im Mittelpunkt der Regulation stehen unterschiedliche Aspekte des Konsumentenschutzes, darunter neue Richtlinien für die algorithmischen Systeme der Dienste. Doch kann durch die Governance das Nutzererlebnis verbessert werden? Um diese Frage zu beantworten, wurden theoretische und empirische Erkenntnisse aus UX- und Usability-Forschung, Informatik, Datenwissenschaften, der Policy- und Governance-Forschung mit Soziologie und Psychologie kombiniert, um einen interdisziplinären theoretischen Rahmen zu erarbeiten. Dieser theoretische Rahmen unterteilt das algorithmische Nutzererlebnis in: Gerechtigkeit, Haftbarkeit, Transparenz, Erklärbarkeit und Definierbarkeit. Anhand dieser fünf theoretischen Dimensionen wurden die nutzerbezogenen Vorschriften im Digital Services Act, insbesondere mit Schwerpunkt der Empfehlungssysteme, analysiert. In einer repräsentativen Umfrage wurden Governance-Ansätze aus dem Digital Services Act sowie aus der wissenschaftlichen Literatur gegenübergestellt. Die Umfrage ergab mit wenigen Ausnahmen eine Übereinstimmung mit den Ansätzen des Digital Services Act, was eine Verbesserung des algorithmischen Nutzererlebnisses nahelegt. Die Ergebnisse zeigen, dass Governance nach Digital Services Act für die Schweiz ein gangbarer Weg ist. Künftige Forschung sollte sich darauf fokussieren, welche Kompetenzen Nutzerinnen und Nutzer benötigen, um algorithmische Systeme selbst zu konfigurieren, sowie welche Bedürfnisse sie insgesamt an entsprechende Parametrierungen stellen.

## Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretische Modelle .....	5
2.1	Algorithmic eXperience.....	5
2.2	ResQue .....	7
2.3	FATE-Faktoren.....	8
2.4	Theoretischer Rahmen FATED .....	9
3	Literaturanalyse .....	11
3.1	Digital Services Act.....	11
3.1.1	Aufbau und Inhalt des Digital Services Act.....	11
3.1.2	Digital Services Act aus Nutzerperspektive .....	13
3.1.3	Empfehlungssysteme im Digital Services Act.....	15
3.1.4	Aufsichtsrechtliche Befugnisse und Sanktionen .....	18
3.2	Governance und Regulation von algorithmischen Systemen .....	18
3.2.1	Governance-Theorie.....	18
3.2.2	Social Media, AI und Governance .....	19
3.3	Social Media, Recommender Systeme und maschinelles Lernen .....	21
3.3.1	Abgrenzung der Technikbegriffe.....	21
3.3.2	Empfehlungssysteme .....	22
3.3.3	Collaborative Filtering.....	24
3.3.4	Social Recommender Systems.....	25
3.3.5	Maschinelles Lernen und Social Recommender Systems.....	26
3.4	Empirischer Forschungsstand .....	27
3.4.1	Typisierung von Social-Media-Algorithmen .....	27
3.4.2	Risiken von sozialen Medien .....	28
3.4.3	FATED-Faktoren .....	29
4	Methodik.....	37
4.1	Hypothese und Operationalisierung .....	38
4.1.1	Gerechtigkeit .....	38
4.1.2	Haftbarkeit.....	39
4.1.3	Transparenz .....	41
4.1.4	Erklärbarkeit .....	42
4.1.5	Definierbarkeit .....	42
4.2	Datenerhebung.....	43

---

4.3	Struktur der Stichprobe.....	44
5	Ergebnisse.....	47
5.1	Einstellung zu Aufsicht, Regulation und den algorithmischen Systemen.....	47
5.2	Einstellung zu Risiken auf sozialen Medien.....	48
5.3	Einstellung zu Haftbarkeit und Aufsichtsinstanz .....	49
5.4	Einstellung zu Transparenzpflichten von algorithmischen Systemen .....	51
5.5	Hypothesentest.....	51
5.6	Teilmengenanalyse.....	56
5.6.1	Einstellung zu Governance.....	56
5.6.2	Einstellung zum Digital Services Act .....	58
6	Diskussion .....	61
6.1	Interpretation der Ergebnisse .....	61
6.2	Kontextuelle Limitationen.....	63
6.3	Methodische Limitationen.....	64
6.4	Implikationen .....	65
7	Fazit.....	67
8	Literaturverzeichnis .....	71
9	Anhang.....	81
9.1	Anhang 1: Fragebogen.....	81
9.2	Anhang 2: Gesamtergebnisse Umfrage .....	88

## Abbildungsverzeichnis

Abbildung 1: Algorithmic eXperience in Social Media (Alvarado und Waern, 2018).....	5
Abbildung 2: ResQue Structural Model (Pu et al., 2011).....	7
Abbildung 3: FATED-Rahmen.....	10
Abbildung 4: Grafische Darstellung der Technikbegriffe .....	22
Abbildung 5: Grafische Darstellung des methodischen Vorgehens.....	37



## Tabellenverzeichnis

Tabelle 1: Tabelle zum Aufbau des Digital Services Acts .....	12
Tabelle 2: Struktur der Stichprobe.....	44
Tabelle 3: Statistische Werte.....	45
Tabelle 4: Ergebnisse zur Einstellung zu Aufsicht, Regulation und Algorithmen.....	47
Tabelle 5: Ergebnisse zur Einstellung zu Risiken auf sozialen Medien .....	48
Tabelle 6: Ergebnisse zur Einstellung zu Haftbarkeit und Aufsichtsinstanz (1 von 2) ..	49
Tabelle 7: Ergebnisse zur Einstellung zu Haftbarkeit und Aufsichtsinstanz (2 von 2) ..	50
Tabelle 8: Ergebnisse zur Einstellung zu Transparenzpflichten .....	51
Tabelle 9: Tabelle der Hypothesentests .....	52
Tabelle 10: Ergebnisse nach Hypothesen (ohne Grafik) .....	54
Tabelle 11: Ergebnisse nach Hypothesen (mit Grafik) .....	55
Tabelle 12: Ergebnisse der Teilmengenanalyse zu allen Governance-Ansätzen.....	56
Tabelle 13: Ergebnisse der Teilmengenanalyse zu DSA-Ansätzen .....	58

## Abkürzungsverzeichnis

Art.	Artikel (Gesetz)
Abs.	Absatz (Gesetz)
ADM	Automated Decision-Making, Automatisierte Entscheidungssysteme
AGB	Allgemeine Geschäftsbedingungen
AS	Algorithmisches System
AX	Algorithmic eXperience
CBF	Content-Based Filtering
CF	Collaborative Filtering
CSR	Corporate Social Responsibility
DSA	Digital Services Act
EU	Europäische Union
KI	Künstliche Intelligenz
ML	Maschinelles Lernen
RL	Reinforcement Learning
RS	Recommender-System, Empfehlungssystem
UCD	User-Centered Design
UX	User eXperience
VLOP	Very Large Online Plattform
XAI	Explainable AI

# 1 Einleitung

*«The best products don't win. The ones everyone use win.»*

- Andrew Bosworth, Chief Technology Officer von Meta Inc., Auszug aus einem internen Memo aus dem Jahr 2016 (Mac et al., 2018)

Die Art und Weise, wie Menschen kommunizieren, Information austauschen und am öffentlichen Diskurs teilnehmen, wurde in den letzten Jahrzehnten zunehmend von sozialen Medien beeinflusst. Während diese Dienste uns eine nie dagewesene Konnektivität ermöglichten, entstanden durch neuartige, algorithmische Systeme soziale und politische Risiken. Diese algorithmischen Systeme, die als zentrales Element von Social-Media-Plattformen personalisiert Inhalte selektieren, priorisieren und empfehlen, erschaffen in den Köpfen der Nutzerinnen und Nutzer neue Realitäten. Beiträge auf Plattformen, die mit hoher Wahrscheinlichkeit Interaktion generieren, werden ohne Berücksichtigung der Inhalte Nutzenden empfohlen. Dies kann unter anderem bedeuten, dass Nutzenden der Plattformen gewisse Weltanschauungen, die Unwahrheiten oder sogar Verschwörungstheorien transportieren, algorithmisch selektiert und empfohlen werden, was insbesondere seit der Corona-Pandemie in weiten Teilen der Bevölkerung für Unmut sorgt. Die Plattformen wurden aber auch zur Durchführung von politischen Kampagnen, die als manipulativ und intransparent bewertet werden, benutzt, darunter 2016 im Kontext der US-amerikanischen Präsidentschaftswahlen sowie der Brexit-Abstimmung im Vereinigten Königreich. Obschon die Dienste weiterhin populär sind, haben diese Vorkommnisse eine Vertrauenskrise gegenüber sozialen Medien und algorithmischen Systemen verursacht. Die Befürchtung, dass Social-Media-Plattformen ohne Rücksicht auf das Wohl der Nutzenden bedingungsloses Wachstum anstreben, haben Konsumentinnen und Konsumenten der Plattformen schon lange. In gewissem Ausmass wurden diese Befürchtungen durch Aussagen von Whistleblowern bestätigt, unter anderem in einer medial vielbeachteten Befragung der ehemaligen Facebook-Mitarbeiterin und Datenwissenschaftlerin Frances Haugen im US-Kongress im Jahr 2021.

Daran, dass Social-Media-Unternehmungen aus gutem Willen auf potenziell schädliche Geschäftspraktiken verzichten werden, scheinen mittlerweile nur noch wenige Menschen zu glauben. Diese Skepsis ebnete jüngst den Weg für Debatten zu Regulation und Aufsicht von Onlineplattformen in der Europäischen Union. Der 2020 vorgestellte Digital Services Act wurde am 19. Oktober 2022 verabschiedet und tritt nach Ablauf der Übergangsfrist im Februar 2024 in Kraft. Die EU verspricht sich durch die darin enthaltenen Neuerungen ein sichereres Online-Umfeld, verbesserten Verbraucherschutz und hochwertige Onlinedienste. Ein besonderer Fokus des Gesetzes liegt auf den algorithmischen

Systemen der Plattformen, die nach DSA neuen Transparenzpflichten unterliegen und Anbieter für Schäden, die durch diese Systeme entstehen, künftig haftbar gemacht werden können.

Die vorliegende Bachelor-Thesis beschäftigt sich mit der Fragestellung, ob Governance-Strategien nach Digital Services Act für die schweizerische Legislative einen gangbaren Weg darstellen. Dazu wird zunächst anhand von User-eXperience-Modellen eruiert, welche der DSA-Gesetzesartikel für eine nutzerzentrierte Analyse ausschlaggebend sind, da sie sich direkt oder indirekt auf das Nutzererlebnis auswirken. Im Anschluss werden diese Artikel mit bestehender Literatur verglichen, um alternative Umsetzungsmethoden zu identifizieren, die sich auf die User eXperience besser auswirken könnten. In einer Umfrage werden die DSA-Ansätze diesen Alternativen gegenübergestellt. Diese Bachelor-Thesis fokussiert sich auf die Zielgruppe der jungen Erwachsenen, die in ihrem Alltag auf digitale Governance angewiesen sind, da die Dienste ihre primäre Unterhaltungs- und Informationsquelle darstellen oder ihr bevorzugter Kommunikationskanal sind. Der Konsumverzicht scheint, trotz Misstrauen gegenüber den Diensten, keine Option zu sein. Die Forschungsfrage lautet dabei: «befürwortet eine Mehrheit von jungen Erwachsenen in der Schweiz die Governance von Social-Media-Algorithmen nach Digital Services Act?»

Diese Bachelor-Thesis folgt einem Mixed-Method-Approach. Sie kombiniert die vertiefte Analyse der Gesetzesreform mit einer wissenschaftlichen Literaturanalyse für die Formulierung der Forschungshypothesen. Diese werden mit einer Meinungsbefragung überprüft.

Die Arbeit ist in einen empirischen Aufbau gegliedert. Auf die Einleitung folgt das zweite Kapitel «theoretische Modelle», in dem die Grundlagenmodelle aus den Bereichen User eXperience vorgestellt werden und der theoretische Rahmen definiert wird. Im dritten Kapitel «Literaturanalyse» wird der Wissensstand zum Digital Services Act, Governance von sozialen Medien und algorithmischen Systemen, die Systematik und Funktionsweise von Empfehlungssystemen aufgearbeitet. Das Kapitel wird mit einem Exkurs in empirisch gesichertes Wissen zu Social-Media-Risiken und der Typisierung von Social-Media-Algorithmen abgerundet. Das vierte Kapitel «Methodik» erläutert das Vorgehen bezüglich der Datenerhebung und den statistischen Berechnungen. Es beinhaltet ausserdem einen Überblick der Hypothesen und der Operationalisierung. Im fünften Kapitel «Ergebnisse» werden die Umfrageresultate in unterschiedlichen Analysen dargelegt und die Hypothesentests durchgeführt. Im sechsten Kapitel «Diskussion» findet sich eine Interpretation der Ergebnisse sowie eine kritische Würdigung der kontextuellen und methodische Limitationen der Arbeit sowie eine Auflistung künftiger Implikationen für die Legislative,

Forschende und andere Stakeholder. Das siebte und letzte Kapitel «Fazit» resümiert die gesamte Bachelor-Thesis.



## 2 Theoretische Modelle

### 2.1 Algorithmic eXperience

Bevor Algorithmic eXperience (AX) eingeordnet wird, müssen User eXperience (UX) und User-Centered Design (UCD) abgegrenzt werden. Es gibt mehrere Definitionen von UX. UX beschäftigt sich mit der Gestaltung von befriedigenden und zufriedenstellenden Nutzererlebnissen. Laut Hassenzahl und Tractinsky (2006) ist UX eine Gestaltungsmethode, die die Nutzung von Technologie neben praktischen Faktoren als ein persönliches Erleben der Konsumentinnen und Konsumenten versteht. UX resultiert aus dem Aufeinandertreffen von mehreren Faktoren: Haltungen und Erwartungen der Nutzenden, Handhabung des Systems und dem Interaktionskontext der Nutzung (S. 95). UX bildet seit Jahren die Gestaltungsgrundlage digitaler und interaktiver Produkte, priorisiert Erlebens- und Erfahrungsqualitäten gegenüber technischen Produkteigenschaften und betrachtet Interaktion als eine Begegnung, bei der Werte, Bedürfnisse, Wünsche und Ziele von Nutzenden im Zentrum stehen (Hassenzahl et al., 2010, S. 353–354). Das Konzept User-Centered Design (UCD) bedeutet nach Garrett (2011, S. 17), dass künftige Nutzende von Beginn an bei jedem Schritt des Produktentwicklungsprozesses berücksichtigt werden. Im UCD-Prozess wird die UX aus verschiedenen Sichtweisen betrachtet, um unterschiedliche Auswirkungen der Entscheidungen im Produktentwicklungsprozess im Vorverein kritisch zu hinterfragen.

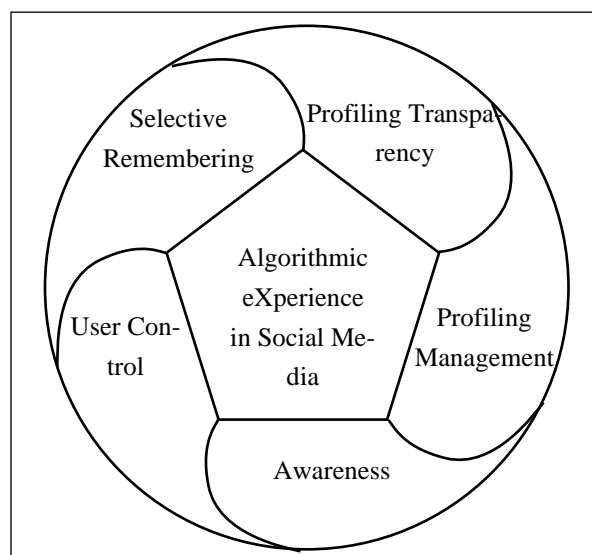


Abbildung 1: Algorithmic eXperience in Social Media (Alvarado und Waern, 2018)

Das Modell der Algorithmic eXperience (AX) als UCD-Konzept wurde erstmals in einem Konferenzpapier von Alvarado und Waern (2018) ausformuliert, obschon der Begriff, bereits vorher in der Literatur verwendet wurde, um Aspekte des Benutzererlebnisses von

algorithmischen Systemen zu beschreiben (Rossiter & Zehle, 2015). Algorithmic eXperience nach Alvarado und Waern (2018) wurde in mehreren Publikationen positiv rezipiert (Klumbyte et al., 2020, S. 2; Shin, 2020a, S. 2; Sundar, 2020, S. 82; Freeman et al., 2023, S. 1; Eg et al., 2023, S. 4–11).

Laut Alvarado und Waern (2018) wirken sich bei sozialen Medien die zugrundeliegenden Algorithmen direkt auf die Nutzererfahrung aus (S. 1). Die für Nutzerinnen und Nutzer bedeutendsten algorithmischen Systeme (AS) sind die, die selbstständig Prozesse einleiten oder Entscheidungen treffen (S. 2). Um diese Nutzererfahrung zu modellieren, wird die beste Praxis nach AX in unterschiedliche Teilbereiche geteilt, wie nachfolgend tabellarisch dargestellt ist.

<b>Faktor</b>	<b>Beschreibung</b>
<b>«profiling transparency»</b>	Nutzende haben Zugriff auf die Personenprofile, die den Empfehlungen zugrunde liegen. Hierbei sollte sowohl internes Profiling innerhalb des Dienstes wie auch externes Profiling, z. B. durch Browser-Tracking, offengelegt werden. Die Information sollte im Interface leicht zugänglich sein (S. 6).
<b>«profiling management»</b>	Nutzende können ihr algorithmisches Profil, als eine Darstellung ihrer Person, selbst verwalten oder bearbeiten (S. 6–7).
<b>«awareness»</b>	Nutzende verstehen die algorithmische Beeinflussung sowie das Profiling und kennen Steuermöglichkeiten, die sie in Bezug auf algorithmische Prozesse selbst haben (S.7).
<b>«user control»</b>	Nutzende können AS steuern, indem sie AS gänzlich deaktivieren oder bestimmte Datenquellen, die den Profilierungsalgorithmus beeinflussen, z. B. die Positionsbestimmung, selektiv ausschalten. Ausserdem sollte im Fall von unpassenden Empfehlungen den AS eine explizite negative Rückmeldung erteilt werden können (S. 7).
<b>«selective remembering»</b>	Einfluss der Nutzenden darauf, welche Daten die AS für die Entscheidungsfindung nutzen, was AS über sie lernen, sowohl bei der Datenermittlung, wie auch später, um AS bestimmte Informationen vergessen zu lassen, wenn sich beispielsweise die persönliche Situation der Nutzenden ändert (S. 7).

Gemäss Klumbyte et al. (2020, S. 2) ist AX als Antwort auf die Zunahme von Machine-Learning-Systemen zu verstehen. ML hat mittels massenhafter Verarbeitung natürlicher Sprache, Bilderkennung, Benutzermodellierung und weiteren Innovation im Allgemeinen die UX von digitalen Produkten erheblich verbessert. Entgegen traditionellen UCD-Methoden, die sich auf Einfachheit und Zugänglichkeit fokussieren, befasst sich AX explizit mit der Interaktion mit Algorithmen in allen Facetten, die auch über das Interface hinausgehen.



## 2.2 ResQue

Das etablierte Framework ResQue (Pu et al., 2011, S. 158), kurz für «Recommender systems Quality of user experience», definiert die elementaren UX-Qualitätsmerkmale von Empfehlungssystemen und schlägt einen Fragebogen für die UX-Evaluation vor.

Das Framework besteht aus 15 Faktoren, die in vier Dimensionen unterteilt sind: den wahrgenommenen Systemqualitäten («perceived system qualities»), den Benutzerüberzeugungen als Ergebnis dieser Qualitäten («users' beliefs»), ihren subjektiven Einstellungen («subjective attitudes») und ihren Verhaltensabsichten («behavioral intentions»). Die Pfeile im Modell implizieren eine positive Beeinflussung zwischen Faktoren.

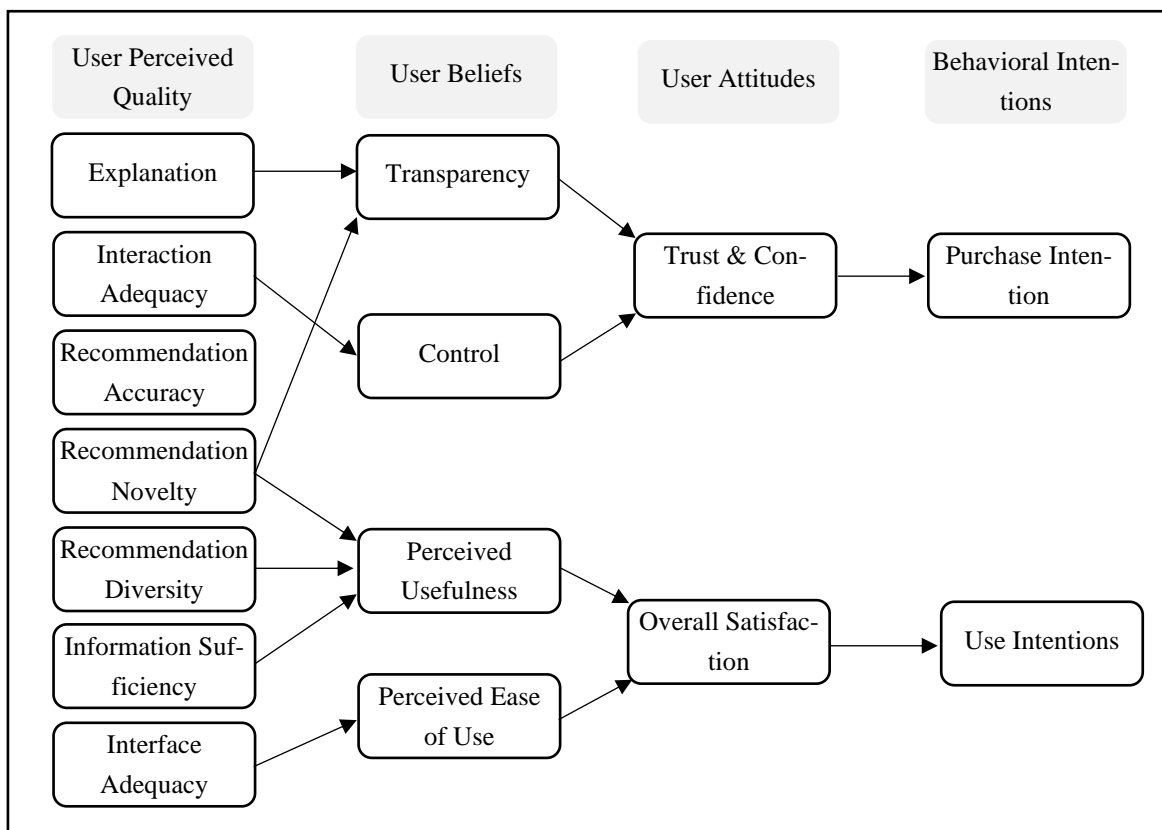


Abbildung 2: ResQue Structural Model (Pu et al., 2011)

Die wahrgenommenen Systemqualitäten («perceived system qualities») beschreiben die unmittelbar wahrgenommenen Qualitäten des Systems, etwa die Empfehlungsgenauigkeit, den Neuheitswert und die Attraktivität der Empfehlungen, sowie die Angemessenheit der Schnittstelle. Dies bezieht sich auf die Sichtbarkeit bedeutsamer Informationen, die Nachvollziehbarkeit der Beschriftungen, der Angemessenheit und Verständlichkeit des Layouts und die Möglichkeit, negative Feedbacks für unpassende Empfehlungen zu geben (S. 158 – 159).

Die zweite Dimension, die Benutzerüberzeugungen («user beliefs»), resultiert aus den wahrgenommenen Qualitäten der ersten Dimension. Darunter fallen die Benutzerfreundlichkeit, Benutzerkontrolle über die Konfiguration des Systems, Transparenz zur inneren Logik des Systems und die Entscheidungsqualität, die bedeutet, dass sich Entscheidungen im Nachhinein als die richtige Wahl erweisen. Die dritte Dimension, die subjektive Einstellung («user attitudes»), resultiert wiederum aus der vorherigen Dimension und beschreibt idealerweise die generelle Zufriedenheit der Nutzerinnen und Nutzer sowie Vertrauen in das Empfehlungssystem. Die vierte Dimension, die Verhaltensabsichten («behavioral intentions») gegenüber dem Empfehlungssystem, umfasst, ob das System Nutzende dazu bewegt, es langfristig zu nutzen sowie die Bereitschaft, das System weiteren Personen zu empfehlen und selbst den erhaltenen Empfehlungen zu folgen (S. 159–160)

### 2.3 FATE-Faktoren

Das Akronym FATE steht für Gerechtigkeit («Fairness»), Haftbarkeit («Accountability»), Transparenz («Transparency») und Ethik («Ethics») oder Erklärbarkeit («Explainability»). FATE ist nicht abschliessend und kein eigenständiges Modell, sondern beschreibt Kernfaktoren, die das Empfinden gegenüber algorithmischen Systemen (AS), insbesondere maschinellem Lernen (ML), positiv beeinflussen (Klumbyte et al., 2020, S. 11; Robles Carrillo, 2020, S. 3; Shin, 2020b, S. 542; Sigfrids et al., 2023, S. 3; Wang et al., 2023, S. 2).

Es ist für die Wissenschaft eine Herausforderung, subjektive Variablen wie Transparenz und Fairness zu messen, da sie von der Wahrnehmung der Nutzenden abhängen. Die Bedeutungen solcher Variablen können als sozial ausgehandelt und gewissermassen konstruiert betrachtet werden (Shin et al., 2022, S. 11). Vertrauensfördernde Effekte der FATE-Faktoren sind nicht gänzlich ergründet und werden im wissenschaftlichen Kontext teilweise kritisch diskutiert. In der Praxis ist die Relevanz von vertrauensbildenden Massnahmen in der Mensch-Maschinen-Interaktion generell anerkannt und wird bei der Entwicklung nutzerzentrierter Algorithmen berücksichtigt, was insbesondere wegen der aktuellen Vertrauenskrise gegenüber KI-Technologien bedeutsam ist (Shin, 2020b, S. 544).

Gerechtigkeit beschreibt aus Nutzerperspektive ausgewogene Entscheidungsprozesse, die insbesondere nicht durch Voreingenommenheit oder Vorurteile beeinflusst werden (Shin et al., 2022, S. 7). Gerechtigkeit ist als eine zentrale gesellschaftliche Wertvorstellung, die die Gleichbehandlung von Menschen und die Einhaltung von Regeln verlangt, zu verstehen (Wang et al., 2023, S. 3). Im Detail lässt sich Gerechtigkeit im Kontext von

AS jedoch kaum definieren, denn sie hängt vom spezifischen Kontext der Applikation und ihrer Nutzenden ab (Shin, 2020b, S. 546).

Das Konzept der Haftbarkeit ergibt sich aus der Tatsache, dass Algorithmen nicht strafrechtlich zur Rechenschaft gezogen werden können und somit die menschliche Rechenschaftspflicht, beispielsweise in Form der Anbieter der digitalen Dienstleistung, definiert werden muss (Shin, 2020b, S. 546).

Transparenz bedeutet, dass AS getestet werden, dass Informationen darüber, warum Entscheidungen getroffen wurden, zugänglich sind (Shin et al., 2022, S. 7) und die innere Funktionsweise der AS wahrheitsgetreu in Beschreibungen wiedergegeben wird (Y. Zhang et al., 2020, S. 296).

Das E steht für Ethik oder Erklärbarkeit. Ethik beschreibt, dass moralische Abwägungen zum Einsatz der AS stattfinden (Robles Carrillo, 2020, S. 2). Somit ist eine grosse Schnittmenge mit dem Konzept der Gerechtigkeit vorhanden. Erklärbarkeit beschreibt, dass das verwendete ML-Modell oder bestimmte Entscheidungsparameter Nutzenden vermittelt werden (Lopes et al., 2022, S. 3). Man spricht bei Informationen zu einem einzelnen Ergebnis von lokalen Erklärungen, oder wenn sie sich auf die allgemeine Logik des Modells beziehen, von globalen Erklärungen (Y. Zhang et al., 2020, S. 296).

## 2.4 Theoretischer Rahmen FATED

Den theoretischen Rahmen der Arbeit bilden die FATE-Kernfaktoren: «Fairness», «Accountability», «Transparency», «Ethics» und «Explainability». «Fairness» und «Ethics» werden zusammengefasst, da sie Vorgaben für gerechte AS, die an Mensch und Gesellschaft keinen Schaden verursachen, formulieren. Aus dem AX-Modell nach Alvarado und Waern (2018) wird ein fünfter Faktor «Definability» abstrahiert. Ihm liegen die AX-Dimensionen «profiling management», «user control» und «selective remembering» zugrunde. Diese Dimensionen verlangen eine Einstellbarkeit der algorithmischen Systeme. Der Aspekt der «Awareness» aus dem AX-Modell wird als Bestandteil von «Explainability» nach FATE betrachtet, da Verständnis die Erklärbarkeit bedingt und inhaltlich kaum davon getrennt werden kann. Daraus ergibt sich der FATED-Rahmen:

- |  |  |
|--|--|
| <ul style="list-style-type: none"><li>• Fairness (and Ethics),</li><li>• Accountability,</li><li>• Transparency,</li><li>• Explainability (and Awareness), and</li><li>• Definability.</li></ul> | <ul style="list-style-type: none"><li>• Gerechtigkeit (und Ethik),</li><li>• Haftbarkeit,</li><li>• Transparenz,</li><li>• Erklärbarkeit (und Verständnis), und</li><li>• Definierbarkeit.</li></ul> |
|--|--|

Obschon das Akronym auf den englischen Wörtern basiert, werden in der Bachelor-Thesis zur besseren Lesbarkeit die deutschen Begriffe verwendet. Der FATED-Rahmen wird verwendet, um nutzerbezogene Indikatoren für die Analyse in Dimensionen zu strukturieren. FATED beschreibt Faktoren, die gewährleistet werden sollen, dass Nutzende den AS von sozialen Medien vertrauen. Anhand FATED wird folglich analysiert, inwiefern sich der DSA auf diese Dimensionen auswirkt und entsprechende Hypothesen zu Indikatoren formuliert.

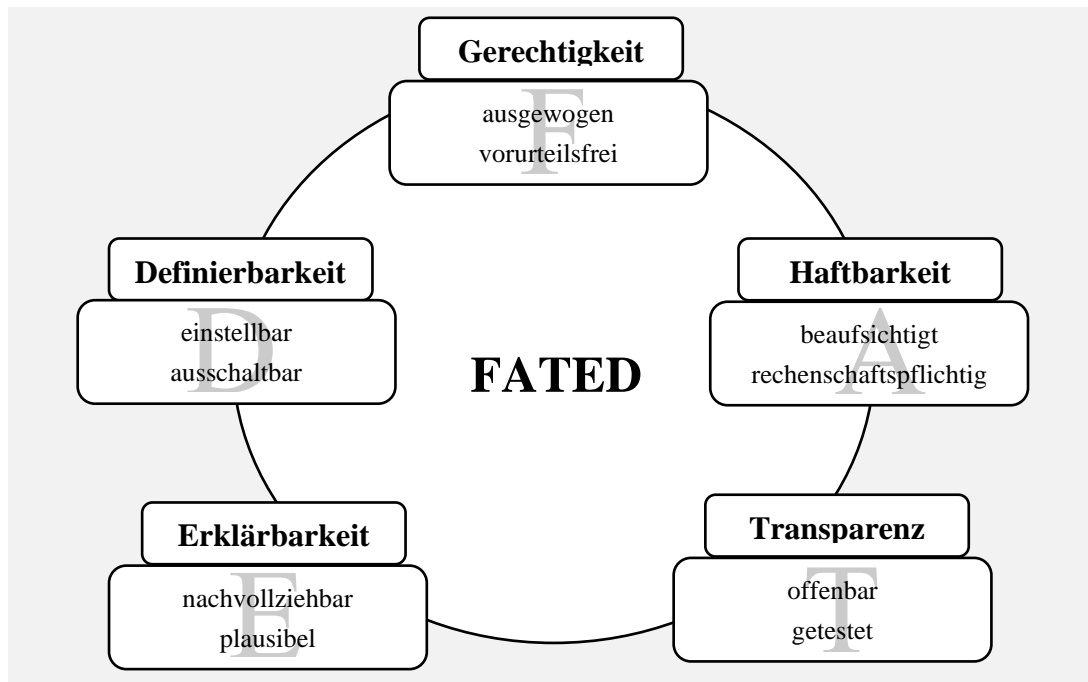


Abbildung 3: FATED-Rahmen

### 3 Literaturanalyse

Im nachfolgenden Kapitel wird der aktuelle Wissens- und Forschungsstand zu den Themenbereichen zusammengefasst, mit dem Ziel, Hypothesen zur Auswirkung des Digital Services Act auf die FATED-Faktoren zu formulieren.

#### 3.1 Digital Services Act

Das aktuelle Jahrzehnt wurde von der Europäischen Kommission zur «Digital Decade», also zum «Jahrzehnt der Digitalisierung» erklärt. Die «Digital Decade» stellt eine Strategie zur technologischen Weiterentwicklung der Europäischen Union (EU) dar und beinhaltet insbesondere Ziele für die digitale Souveränität. Die Entwicklungen im Bereich der Regulation von Online-Plattformen sind im Kontext dieser digitalen Strategie zu sehen. Nachdem die Europäische Kommission im Jahr 2018 eine Beobachtungsstelle für Online-Plattformen geschaffen hatte, wurden 2019 initial Regulationsverordnungen im Plattform-to-Business-Sektor (P2B) auf dem EU-Gebiet erlassen und schlussendlich 2020 erstmals der Digital Services Act (DSA) vorgestellt (Turillazzi et al., 2023, S. 84).

Der DSA ergänzt bestehende EU-Rechtsinstrumente bezüglich audiovisuellen Medien, dem Verbraucherschutz und bezweckt die Aktualisierung bestehender Vorschriften, insbesondere der Richtlinie 2000/31/EG zum elektronischen Geschäftsverkehr im Binnenmarkt (Cauffman & Goanta, 2021, S. 760). Die Probleme, die mit dem DSA adressiert werden, lassen sich laut Turillazzi et al. (2023, S. 89) in fünf Hauptkategorien unterteilen: mangelnde Aufsicht von digitalen Diensten, Gefahren für die Rechte und Freiheiten von Bürgerinnen und Bürgern, Missbrauch der Marktstellung durch Online-Plattformen, mangelnde Transparenz und fehlende Beschwerdeinstrumente für Nutzende.

Die Vorlage wurde am 19. Oktober 2022 vom Europäischen Parlament und Rat verabschiedet, am 27. Oktober 2022 im Amtsblatt publiziert und tritt nach Ablauf einer Übergangsfrist am 17. Februar 2024 in Kraft (Amtsblatt der EU, 2022).

##### 3.1.1 Aufbau und Inhalt des Digital Services Act

Der DSA ist in fünf Kapitel gegliedert, wovon das zweite und dritte Kapitel jeweils in sechs Abschnitte unterteilt sind. Vor dem ersten Kapitel wird in einer Präambel, die insgesamt 156 Bemerkungen beinhaltet, die Relevanz der Gesetzesvorlage erläutert und die Entstehungsgründe dargelegt.

Im ersten Kapitel «Allgemeine Bestimmungen» werden Gegenstand, Geltungsbereich und Begriffe definiert. Im zweiten Kapitel «Haftung der Anbieter von Vermittlungsdiensten» sind Haftungsbestimmungen zu Vermittlungsdiensten aufgeführt. Im dritten Kapitel «Sorgfaltspflichten für ein transparentes und sicheres Online-Umfeld» befinden sich zahlreiche Bestimmungen zu Moderation, Sanktionen gegenüber Nutzenden, Beschwerdeverfahren, Transparenz, Empfehlungssystemen, Interface-Gestaltung und digitalem Marketing. Dieses Kapitel ist dadurch für diese Bachelor-Thesis zentral. Im vierten Kapitel «Umsetzung, Zusammenarbeit, Sanktionen und Durchsetzung» werden den EU-Mitgliedsstaaten Pflichten und Rechte zur Durchführung der Kontroll-, Aufsichts- und Vollzugstätigkeit verliehen. Das fünfte und letzte Kapitel «Schlussbestimmungen» definiert Auswirkungen auf andere Verordnungen und präzisiert den Zeitpunkt der Inkraftsetzung.

<b>Kapitel und Abschnitt</b>	<b>Artikel</b>
1. Allgemeine Bestimmungen	1 – 3
2. Haftung der Anbieter von Vermittlungsdiensten	4 – 10
3. Sorgfaltspflichten für ein transparentes und sicheres Online-Umfeld	11 – 48
1. Bestimmungen für alle Anbieter von Vermittlungsdiensten	11 – 15
2. Zusätzliche Bestimmungen für Hostingdiensteanbieter, einschliesslich Online-Plattformen	16 – 18
3. Zusätzliche Bestimmungen für Anbieter von Online-Plattformen	19 – 28
4. Bestimmungen für Anbieter von Online-Plattformen, die Verbrauchern den Abschluss von Fernabsatzverträgen mit Unternehmern ermöglichen	29 – 32
5. Zusätzliche Verpflichtungen in Bezug auf den Umgang mit systemischen Risiken für Anbieter von sehr grossen Online-Plattformen und sehr grossen Online-Suchmaschinen	33 – 43
6. Sonstige Bestimmungen über Sorgfaltspflichten	44 – 48
4. Umsetzung, Zusammenarbeit, Sanktionen und Durchsetzung	49 – 88
1. Zuständige Behörden und Koordinatoren für digitale Dienste	49 – 55
2. Zuständigkeit, koordinierte Untersuchungen und Kohärenzmechanismen	56 – 60
3. Europäisches Gremium für digitale Dienste	61 – 63
4. Beaufsichtigung, Untersuchung, Durchsetzung und Überwachung in Bezug auf Anbieter sehr grosser Online-Plattformen und sehr grosser Online-Suchmaschinen	64 – 83
5. Gemeinsame Durchsetzungsbestimmungen	84 – 86
6. Delegierte Rechtsakte und Durchführungsrechtsakte	87 – 88
5. Schlussbestimmungen	89 – 93

Tabelle 1: Tabelle zum Aufbau des Digital Services Acts

Der DSA spezifiziert unterschiedliche Kategorien von digitalen Serviceanbietern, für die unterschiedliche Regulationsintensitäten gelten. Die Kategorien gestalten sich wie folgt:

- Vermittlungsdienste: Dienste, die Hosting, Caching oder Durchleitungen anbieten, beispielsweise um das Nutzerverhalten seitenübergreifend aufzuzeichnen.
- Online-Plattformen: Dienste, die im Auftrag von Nutzenden Informationen speichern und öffentlich verbreiten. Hierzu zählen Social-Media-Dienste.
- Als «Very Large Online Plattform» (VLOP), auf Deutsch «sehr grosse Online-Plattformen», gelten Dienste, die im EU-Gebiet mindestens 45 Millionen regelmässige Nutzende erreichen. Für VLOP gelten strengste Auflagen. Die Schwelle von 45 Millionen Nutzenden stellt zehn Prozent der EU-Bevölkerung dar. Im DSA ist definiert, dass die Zahl angepasst werden kann, wenn sich die Bevölkerungszahl der EU massgeblich verändert.

### 3.1.2 Digital Services Act aus Nutzerperspektive

Im nachfolgenden Abschnitt werden die möglichen Auswirkungen des DSA auf Nutzende von digitalen Plattformen erläutert. Es gibt keine abschliessende Systematisierung der nutzerbezogenen Effekte des DSA. Gemäss einer Auflistung auf der Webseite (Europäische Kommission, 2023) der Europäischen Kommission werden die folgenden positiven Auswirkungen auf Nutzende erwartet:

Thema	Auswirkungen
Ein sichereres Online-Umfeld	Harmonisierte Verfahren zur Meldung illegaler Inhalte, Sorgfaltspflichten für Plattformen, Befähigung der Behörden zur Kontrolle der Plattformen
Besserer Verbraucherschutz	Information im Falle von Löschung von Inhalten, Zugang zu Beschwerdemechanismen im Domizilland der Nutzenden, Risikobewertungen durch Plattformen, Reaktionsmechanismen im Falle von öffentlichen Krisen, Schutzbestimmungen für Minderjährige, Regulierung von Werbung, Datenzugänge für Wissenschaftlerinnen und Forscher beispielsweise zur Ermittlung von Risiken
Mündige Nutzende, Bürgerinnen und Bürger	Schutz vor Manipulation durch Empfehlungssysteme und dem Missbrauch von Werbesystemen
Hochwertige digitale Dienste zu niedrigeren Preisen	Schutz vor marktbeherrschender Stellung einiger weniger Konzerne und unfairen Servicebedingungen.

Unterschiedliche Betrachtungsweisen auf das Gesetz können zu abweichenden Rückschlüssen dazu führen, welche Aspekte des DSA für Nutzende von Belangen sind. Dies

spiegelt sich in den wissenschaftlichen Publikationen wieder, die relevante Änderungen für Konsumentinnen und Konsumenten diskutieren. Nachfolgend wird in einer nicht abschliessenden Aufzählung eine Auswahl an Publikationen resümiert.

Gemäss Carvalho et al. (2021, S. 96–98) beziehen sich die für Konsumentinnen und Konsumenten wesentlichen Änderungen auf die Rückverfolgbarkeit der Produkthanbieter (S. 96), vorvertragliche Informationen, darunter Produktsicherheitsinformationen (S. 97), Regulation und Transparenz bezüglich Werbung und Empfehlungssystemen sowie Haftungsnormen von Online-Plattformen (S. 98).

Laut Cauffman & Goanta (2021, S. 759) sind aus Nutzerperspektive wichtig: optimierte Kohärenz innerhalb des EU-Rechts, die Haftung von Vermittlungsdiensten, die Regulation von Ausgliederungen an private Drittanbieter und die digitalen Mechanismen zur Einsprache und Beschwerdeführung gegen Dienste.

Nach Senftleben (2021, S. 1–3) sind die Transparenzpflichten in Bezug auf personalisierte Marketing-Automatismen wichtig. Gemäss Hacker (2021, S. 25–26) besteht aus Nutzerperspektive eine besondere Relevanz in den Massnahmen zum Schutz der Konsumentinnen und Konsumenten vor algorithmischer Manipulation in den Bereichen Empfehlungssysteme, digitales Marketing und Beeinträchtigung der freien Entscheidungsfindung der Nutzenden.

Gemäss Leerssen (2023, S. 5) sind Transparenzpflichten in Bezug auf Reichweitenbeschränkung für Nutzende massgebend. Damit gehe die Transparenz in Bezug auf Empfehlungssysteme einher, da diese bei der Reichweitenbeschränkung eine Rolle spielen, wenn sie bestimmte Inhalte beispielsweise automatisch depriorisieren, was im Internet-Jargon auch als Downranking bezeichnet wird.

Laut Zódi (2022) sind als besondere Mehrwerte für Nutzende der Schutz vor illegalen Inhalten (S. 102), Vorgaben gegenüber der Gestaltung von Dienstleistungsverträgen (S. 103), Transparenz zu Moderationsentscheidungen gegenüber Nutzenden und ihren Konten (S. 103), Transparenzpflichten zu Empfehlungsalgorithmen (S. 104), Beschwerdemechanismen (S. 105) sowie die generellen, öffentlichen Transparenzregeln für die Plattformen (S. 106) zu nennen.

Als wiederkehrende Konzepte, die aus Nutzerperspektive als signifikant eingestuft werden, lassen sich also unterschiedlichste Formen von Transparenzpflichten erkennen, die unter anderem die Empfehlungssysteme, Werbung, Kontobeschränkungen- sowie Moderationsmassnahmen und Offenlegung aller beteiligten Produkthanbieter bedingen. Ferner werden die neuen Instrumente der Beschwerdeführung, Schutzpflichten vor illegalen Inhalten sowie Haftungspflichten der Online-Dienste als relevant für Nutzende betrachtet.



Zu jedem der genannten Themenbereiche liesse sich eine eigenständige Analyse über die Auswirkungen auf Nutzende anstellen. Diese Bachelor-Thesis wird sich auf die Regulation der algorithmischen Systeme, insbesondere der Empfehlungssysteme, fokussieren.

### 3.1.3 Empfehlungssysteme im Digital Services Act

Im DSA sind die Recommender-Systeme (RS) in den Art. 27 und 38 reguliert. In den Art. 34 und 35 werden Pflichten zur Risikobewertung und Risikominderung spezifiziert, die ebenfalls insbesondere für RS gültig sind. In den Art. 40 (Datenzugang für öffentliche Instanzen) und Art. 44 (Normierungen) finden RS beiläufig Erwähnung. Die betreffenden Gesetzesstellen sind nachfolgend in tabellarischer Darstellung mit zusammengefasstem Inhalt aufgeführt und werden mit Verweis auf wissenschaftliche Literatur kommentiert.

Art.	Inhalt
27	Darlegung der Parameter von Empfehlungssystemen und Anpassungsmöglichkeiten in den AGBs (Abs. 1); Erläuterung, warum dem Nutzer bestimmte Informationen vorgeschlagen werden, darunter a) die Kriterien für die Selektion von Information und b) Gründe für die relative Bedeutung dieser Parameter (Abs. 2). Stehen mehrere Parameter mit Einfluss auf die Informationsbereitstellung zur Verfügung, muss es Nutzern ermöglicht werden, die bevorzugte Option jederzeit auszuwählen und zu ändern (Abs. 3).
38	Zusätzlich zu Art. 27 müssen sehr grosse Online-Plattformen, die Empfehlungssysteme verwenden, eine Konfiguration der Empfehlungssysteme anbieten, die nicht auf Profiling der Nutzenden beruhen.

Die in Art. 27 Abs. 1 und Abs. 2 dargestellten Transparenzpflichten beziehen sich zunächst auf die Offenlegung der Empfehlungsmethoden in den allgemeinen Geschäftsbedingungen (AGB). In Abs. 3 findet sich die Ergänzung, dass, sofern mehrere Parametrierungsvarianten verfügbar sind, auf der Plattform eine Funktion zugänglich gemacht werden muss, mit der Parametrierungsoptionen angepasst werden können. Aber es besteht keine Pflicht zur Bereitstellung mehrerer Optionen, ausser für VLOPS, die ein Profiling-freies Empfehlungssystem (RS) bereitstellen müssen.

Zu betonen ist, dass gemäss gängiger Interpretation des DSA die verwendeten Parameter nur in den AGB und nicht bei der angezeigten Information abgebildet werden müssen (Lodder & Carvalho, 2022, S. 552; Schwemer, 2021, S. 64; Zödi, 2022, S. 104). Dies steht in Kontrast zum Art. 26 «Werbung auf Online-Plattformen», der für bezahlte Beiträge und Werbung vorschreibt, dass Informationen zu den verwendeten Parametern in Echtzeit beim Werbebeitrag angezeigt werden müssen (Art. 26 Abs. 1).

Im Art. 34 werden sehr grosse Online-Plattformen (VLOP) verpflichtet, Risikobewertungen selbst vorzunehmen sowie im Art. 35 ebendiese Risiken zu mindern. Im Art. 37 ist festgehalten, dass unabhängige Prüfgesellschaften einmal jährlich die Einhaltung der Sorgfaltspflichten überprüfen und in einem Bericht dokumentieren sollen. Im Art. 41 ist zudem festgelegt, dass eine Compliance-Abteilung eingerichtet werden muss, die die Einhaltung des DSA innerhalb der VLOP sicherstellt. Im Art. 42 ist niedergeschrieben, dass regelmässig Transparenzberichte veröffentlicht werden müssen. Dies betrifft Berichte zu Moderations- und Beschwerdeprozessen (Art. 15), der Risikobewertung (Art. 34) sowie den Massnahmen zur Risikominderung (Art. 35) und den unabhängigen Prüfbericht (Art. 37). Während den Behörden der unabhängige Prüfbericht vollständig zugestellt werden muss, dürfen vertrauliche oder sicherheitsrelevante Informationen aus der öffentlichen Version entfernt werden.

Art	Inhalt
34	<p>Sehr grosse Online-Plattformen ermitteln, analysieren und bewerten sorgfältig alle systemischen Risiken in der EU, die sich aus der Konzeption, dem Betrieb sowie der Nutzung ihrer algorithmischen Systeme ergeben. Sie führen die Risikobewertungen zum Anwendungsbeginn und danach jährlich, sowie vor der Einführung von kritischen Funktionen durch (Abs. 1). Die Risikobewertung erfolgt verhältnismässig zu den systemischen Risiken, ihrer Schwere und Wahrscheinlichkeit. Sie umfasst die folgenden systemischen Risiken: Verbreitung rechtswidriger Inhalte über ihre Dienste (Abs. 1a); Auswirkungen auf die Ausübung der Grundrechte, insbesondere Achtung der Menschenwürde, Achtung des Privat- und Familienlebens, Schutz personenbezogener Daten, Meinungs- und Informationsfreiheit, Recht auf Nichtdiskriminierung, Rechte des Kindes und den Verbraucherschutz (Abs. 1b); Auswirkungen auf die gesellschaftliche Debatte, Wahlprozesse und die öffentliche Sicherheit (Abs. 1c); Auswirkungen auf geschlechtsspezifische Gewalt, Schutz der öffentlichen Gesundheit, Schutz von Minderjährigen sowie schwere nachteilige Folgen für das körperliche und geistige Wohlbefinden einer Person (Abs. 1d).</p> <p>Bei der Risikobewertung berücksichtigen sehr grosse Online-Plattformen, wie die folgenden Faktoren diese systemischen Risiken beeinflussen (Abs. 2): die Gestaltung ihrer Empfehlungssysteme und anderer algorithmischer Systeme (Abs. 2a).</p>
35	<p>Sehr grosse Online-Plattformen ergreifen angemessene, verhältnismässige und wirksame Risikominderungsmassnahmen, die auf die ermittelten systemischen Risiken zugeschnitten sind, darunter (Abs. 1): die Erprobung und Anpassung ihrer algorithmischen Systeme, einschliesslich ihrer Empfehlungssysteme (Abs. 1d).</p>

Bezogen auf Art. 34 und Art. 35 finden in der Literatur Ansätze Kritik, die nur schwere Schäden begrenzen wollen, da technologische Systeme in etlichen Fällen subtilere Schäden an Menschen und Gemeinschaften verursachen können (Salgado-Criado &

Fernandez-Aller, 2021, S. 61). Der Ansatz der Risikoanalyse und Risikominderung ist laut Turillazzi (2023, S. 100) als «Ex-Ante» zu bezeichnen, also auf die Vorhersage zukünftiger Risiken ausgerichtet und wird als Ansatz von sanfter Ethik kategorisiert. Das bedeutet, die Abwägung von ethischen Risiken durch Beauftragte selbst wird priorisiert, anstatt von harter Ethik, also klaren moralischen und ethischen Vorgaben durch Politik, Aufsichtsbehörden oder das Gesetz. Der DSA-Ansatz wird ausserdem als eine Form von teilweiser Selbstregulation betrachtet (Buchheim, 2022, S. 259–260), als Due Diligence beziehungsweise an Sorgfaltspflichten orientiert (Schwemer, 2021, S. 63–65) und als prozedural-formal (Zödi, 2022, S. 107) klassifiziert, da der Fokus auf der Schaffung von Prozessen der Selbstregulation anstatt auf definierten Grenzwerten und taxativer Kontrolle durch Behörden liegt. Das Einbinden von unabhängigen Prüfinstitutionen, die sogenannte regulatorische Intermediation, kann zwar die Glaubwürdigkeit politischer Selbstverpflichtungen erhöhen und effizient sein, doch kann es vorkommen, dass die Regulierungsakteure nicht unabhängig von denen sind, die sie ernannt haben (Medzini & Levi-Faur, 2023, S. 2). Laut Cauffmann und Goanta (2021, S. 770–771) ergibt das Zuweisen der Durchführung von Risikoanalysen an die Dienstleister Sinn, weil die Unternehmen selbst als Einzige vollen Zugriff auf die fraglichen Daten haben und die Prozesse auf ihren Plattformen am zuverlässigsten kennen. Dies bedingt jedoch, um eine reine Selbstregulation zu vermeiden, das Einbinden unabhängiger Prüfgesellschaften.

Insgesamt sind im DSA vorwiegend prozessuale Vorgaben zur Bewertung und Minderung von Risiken verschriftlicht, wodurch die Onlinedienste in die Verantwortung gestellt werden, die Risiken auf ihren Plattformen selbstständig zu erkennen und Sorgfaltspflichten nach bestem Wissen und Gewissen zu befolgen.

Art	Inhalt
40	Sehr grosse Online-Plattformen gewähren Zugang zu den Daten, die für die Überwachung und Bewertung der Einhaltung dieser Verordnung erforderlich sind (Abs. 1). Für die Zwecke des Absatzes 1 erläutern sehr grosse Online-Plattformen die Logik, die Funktionsweise und die Tests ihrer algorithmischen Systeme einschliesslich ihrer Empfehlungssysteme (Abs. 3).
44	Die EU fördert die Entwicklung und Umsetzung freiwilliger Normen, die einschlägige europäische und internationale Normungsgremien zumindest in Bezug auf folgende Bereiche festlegen (Abs. 1): Auswahlchnittstellen und Darstellung von Informationen über die Hauptparameter verschiedener Arten von Empfehlungssystemen gemäss den Artikeln 27 und 38 (Abs. 1i).

Im Art. 40 wird der Zugang zu Daten für die Aufsichtstätigkeit der EU-Instanzen bezüglich Empfehlungssystemen geregelt. Im Art. 44 verpflichtet sich die Europäische

Kommission, internationale Normen für «Auswahlschnittstellen und Darstellung von Informationen über die Hauptparameter verschiedener Arten von Empfehlungssystemen» zu erstellen. Diese institutionell ausgerichteten Artikel haben keinen direkten Bezug zu Nutzenden oder der Gestaltung der Onlinedienste.

#### **3.1.4 Aufsichtsrechtliche Befugnisse und Sanktionen**

Die Befugnisse, mit denen die überwachenden Behörden ausgestattet werden, sind aus Perspektive der Haftbarkeit äusserst bedeutsam. Dazu gehören Untersuchungsbefugnisse im Art. 51. Sie beinhalten im ersten Absatz Auskunftspflichten seitens der Anbieter und in Verbindung stehenden Geschäftspartnern oder Mitarbeitern, das Recht auf Untersuchungen am Unternehmenssitz sowie weitere Befragungsbefugnisse. Im zweiten Absatz finden sich Befugnisse, bei Verstössen entsprechende Geldbussen zu erlassen, sowie schädliche Geschäftsmechanismen zu verbieten. Der dritte Absatz bezieht sich auf schwerwiegende Schäden, die Dienste verursachen könnten. In diesem Fall kann die vollständige Sperrung des Dienstes in der EU veranlasst werden. Geldbussen können sich in einem empfindlichen Bereich von bis zu sechs Prozent des weltweiten Jahresumsatzes bewegen. Sanktionsmechanismen gegen Einzelpersonen, beispielsweise technischen Verantwortlichen oder Mitglieder des Managements, sind nicht vorgesehen.

Ferner ist anzumerken, dass Privatpersonen nun bei der zuständigen Behörde Beschwerden gegen die Dienste einreichen können (Art. 53). Ansprüche privater Personen gegenüber den Diensten, beispielsweise im Schadenfall, sind gemäss DSA nach nationalem Recht zu behandeln (Art. 54).

### **3.2 Governance und Regulation von algorithmischen Systemen**

#### **3.2.1 Governance-Theorie**

Als Governance bezeichnet man laut Steurer (2013) die Steuerung und Regulierung von Organisationen, Prozessen und Menschen anhand Regeln und Normen. Diese Regeln und Normen werden als Regulation bezeichnet. Regulation kann sowohl freiwillig als auch verbindlich sein und wird durch staatliche, wirtschaftliche und gesellschaftliche Akteure sich selbst oder anderen auferlegt. Governance bezeichnet in diesem Sinne die Formulierung, Bekanntmachung, Einführung und Durchsetzung von Regulation (S. 388). Governance wird nach Steurer (2013) häufig in der Dichotomie «staatlicher» versus «nicht staatlicher» Regulation typisiert, was die Komplexität von Governance-Formen kaum abzubilden vermag (S. 389). Steuer schlägt für die Betrachtung von Governance

eine Venn-Klassifizierung vor, bei der als Governance-Akteur zwischen Staat («Government»), Gesellschaft («Civil») und Privatwirtschaft («Business») differenziert wird (S. 398). Es existieren Mischformen, bei denen Staat, Gesellschaft und Privatwirtschaft zu zweit oder zu dritt am Governance-Prozess teilnehmen. Bei staatlicher Regulation («Government Regulation») differenziert Steurer (2013) weiter zwischen harter Regulation («Hard Regulation»), die verpflichtend ist und sanfter Regulation («Soft Regulation»), die freiwillig ist (S. 393). Bei privatwirtschaftlicher Regulation findet zudem eine Differenzierung zwischen kollektiver Industrie-Selbstregulation und individueller Betriebs-Selbstregulation statt (S. 394–395). Individuelle Selbstregulation, die nicht auf rechtlichen Pflichten basiert und auf Ebene einer Organisation, einer Institution oder eines Betriebes stattfindet, wird als Corporate Social Responsibility (CSR) bezeichnet (S. 395–396).

CSR beschreibt selbst auferlegte, öffentlich kommunizierte Geschäftsprinzipien, mit dem Ziel, Stakeholder wie Aktionäre und Investoren vom Willen zur Transparenz und Rechenschaft zu überzeugen (Albareda, 2008, S. 433; Gill, 2008, S. 453). Eine Hauptkritik an selbstregulativen Ansätzen am Beispiel von sozialen Medien ist, dass die Organisationen in die Verantwortung gestellt werden, gegen ihre eigenen wirtschaftlichen Interessen zu handeln und die gründliche Umsetzung der Massnahmen den Geschäftszielen widersprechen kann (Medzini, 2022, S. 4; Medzini & Levi-Faur, 2023, S. 2–5).

### 3.2.2 Social Media, AI und Governance

Nach Latzer und Festic (2019) umfasst Governance durch Algorithmen («Governance by algorithms»), auch als algorithmische Governance («Algorithmic Governance») bezeichnet, die automatisierten Steuerungseffekte algorithmischer Systeme auf Menschen im Alltag (S. 2). Governance von Algorithmen («Governance of Algorithms») hingegen stellt die politischen Entscheidungen und Prozesse zur Steuerung und Regulierung von Algorithmen dar (S. 8–11). Nachfolgend wird näher auf die Governance von Algorithmen eingegangen.

Zahlreiche wissenschaftliche Disziplinen beschäftigen sich mit unterschiedlichen Teilaspekten der Social-Media-Regulation: von Datenschutz (Arogyaswamy, 2020, S. 836–837), über Content-Regulation von verbotenen Inhalten und Desinformation (Medzini, 2022, S. 1–3), bis hin zu Konsumentenschutz oder dem Kartell- und Wettbewerbsrecht (Budzinski & Mendelsohn, 2021, S. 2–5). Aus diesem Grund findet sich in der wissenschaftlichen Literatur keine einheitliche Empfehlung dazu, wie Social-Media-Anbieter reguliert werden sollen. Welche Ansätze Sinn ergeben, hängt von den konkreten Risiken, der Eintrittswahrscheinlichkeit und dem Schadenpotenzial ab.

Laut Krönke (2020) benutzen Social-Media-Plattformen Verfahren maschinellen Lernens (ML), die auch als künstliche Intelligenz (KI) bezeichnet werden, beispielsweise zur Moderation von Inhalten oder für die Generierung von personalisierten Empfehlungen sowie der Kuratation von Inhalten (S. 150). Die Regulation von ML-basierten Social-Media-Algorithmen kann protektiv, also auf Kundenschutz ausgerichtet sein, etwa wenn sie die Verbreitung schädlicher Inhalte von Drittparteien adressiert (S. 152–153). Sie kann aber auch fördernd sein, wenn sie ergebnisorientierte Normen und Standards formuliert, die eingehalten werden sollten. Ein mögliches Beispiel wäre die Förderung von Meinungsvielfalt durch strengere Regulation von Empfehlungssystemen, die teils in der Kritik stehen, durch automatisierte Kuratation von Inhalten die Entstehung von unausgeglichenen Informationsangeboten zu begünstigen (S. 167–168). Während protektive Regulation die Anspruchsgruppen vor bestimmten Risiken und Gefahren schützen will, normiert fördernde Regulation die Konfiguration der Systeme und Applikationen (S. 169). Festzuhalten ist, dass ein generelles Verbot von KI-gestützten Informations- und Kommunikationssystemen verfassungsrechtlich kritisch ist. Im Kontext von Empfehlungssystemen können Nutzende primär regulatorisch vor manipulativen Filtermassnahmen, die durch selektive Informationsvermittlung auch Verzerrungseffekte fördern, geschützt werden (S.170).

Laut Sigfrids et al. (2023) hebt eine Mehrheit der wissenschaftlichen AI-Governance-Konzepte die Beteiligung, Kooperation und Zusammenarbeit mit Stakeholdern durch öffentliche Debatten, Konsultationen, Gremien oder Ausschüsse als wesentliche Komponenten der KI-Governance hervor. Bestehende Konzepte gehen aber nicht detailliert darauf ein, wie diese Anspruchsgruppen in der Praxis einbezogen werden sollen und formulieren keine konkreten Empfehlungen zu notwendigen neuen Instrumenten der Debatte und Entscheidungsfindung (S. 5). Ebenfalls solle in der AI-Governance-Diskussion anstatt der Begrenzung auf unerwünschte Effekte vermehrt ein Fokus daraufgelegt werden, welches Ergebnis erwünscht ist (S. 6).

Selbstregulative Ansätze werden im Kontext von KI, spezifisch im Zusammenhang mit sozialen Medien, sowohl wissenschaftlich, medial wie politisch kritisch rezipiert und werden von zahlreichen Akteurinnen als ungenügend bewertet (Medzini, 2022, S. 21; Medzini & Levi-Faur, 2023, S. 2; B. Zhang & Dafoe, 2020, S. 190–191). Dies mag an der skandalreichen Unternehmensgeschichte von Facebook liegen, die den Eindruck vermittelt, dass ohne klare Regeln oder externe Aufsicht keine effektive Governance möglich sei. Dem Ruf nach staatlicher Regulation ist indes entgegenzuhalten, dass staatliche Akteure nicht zwangsläufig im Interesse der Konsumentinnen und Konsumenten handeln

und gegenüber der Beeinflussung seitens der Industrie nicht immun sind (Chomanski, 2021, S. 262). Darum ist die Integrität von regulierenden Institutionen ausschlaggebend. Als Kompromiss zwischen Selbstregulation und staatlicher Regulation werden Ansätze der hybriden Regulation diskutiert, beispielsweise unter Einsatz von unabhängigen privaten Prüfgesellschaften. In der Literatur wird zudem aufgrund der weltweiten Verfügbarkeit digitaler Dienstleistungen insbesondere die Notwendigkeit von transnationalen oder globalen Regulationsinstrumenten betont (Hoffmann-Riem, 2020, S. 20–22). Ausserdem finden in der jüngeren Literatur Governance-Methoden gewisse Beachtung, an denen die Nutzerinnen und Nutzer der Plattform selbst auch partizipieren können (Engelmann et al., 2020; Fan & Zhang, 2020, S. 1–2).

### **3.3 Social Media, Recommender Systeme und maschinelles Lernen**

#### **3.3.1 Abgrenzung der Technikbegriffe**

In der Literaturrecherche wurden Publikationen ausgewertet, die sich unter anderem mit Recommender-Systemen (RS), maschinellem Lernen (ML), künstlicher Intelligenz (KI), algorithmischer Kuration und automatischen Entscheidungssystemen (ADM) beschäftigen. Die Begriffe haben Überschneidungen, welche nachfolgend erläutert und in Kontext gesetzt werden.

Algorithmische Kuration beschreibt die Auswahl und Selektion von Inhalten zur Personalisierung eines Social-Media-Nutzererlebnisses, bezieht sich meist auf den News-Feed von Plattformen (Bandy & Diakopoulos, 2021, S. 79; Rader & Gray, 2015, S. 174) und wird mit Empfehlungssystemen gleichgesetzt (Freeman et al., 2023, S. 2). Für diese Arbeit wird algorithmische Kuration als eine Methode betrachtet, wie RS in digitale Dienste eingebettet werden können.

ADM sind Algorithmen, die selbstständig Fälle einem Handlungsszenario zuweisen und darauf aufbauend Arbeitsvorgänge umsetzen. Ob alle RS als ADM klassifizierbar sind, ist in der Literatur nicht einheitlich definiert. RS können jedoch im Kontext von sozialen Medien als ADM betrachtet werden, da sie automatisiert Entscheidungen zur Auswahl von Inhalten treffen (Jung & Nardelli, 2020, S. 1–2). Da diese Bachelor-Thesis sich mit sozialen Medien beschäftigt und zum Zeitpunkt, an dem ein Beitrag auf dem News-Feed von Nutzenden erscheint, keine selbstbestimmte Entscheidung mehr möglich ist dazu, ob ein Inhalt konsumiert werden möchte, werden RS den ADM-Systemen gleichgesetzt.

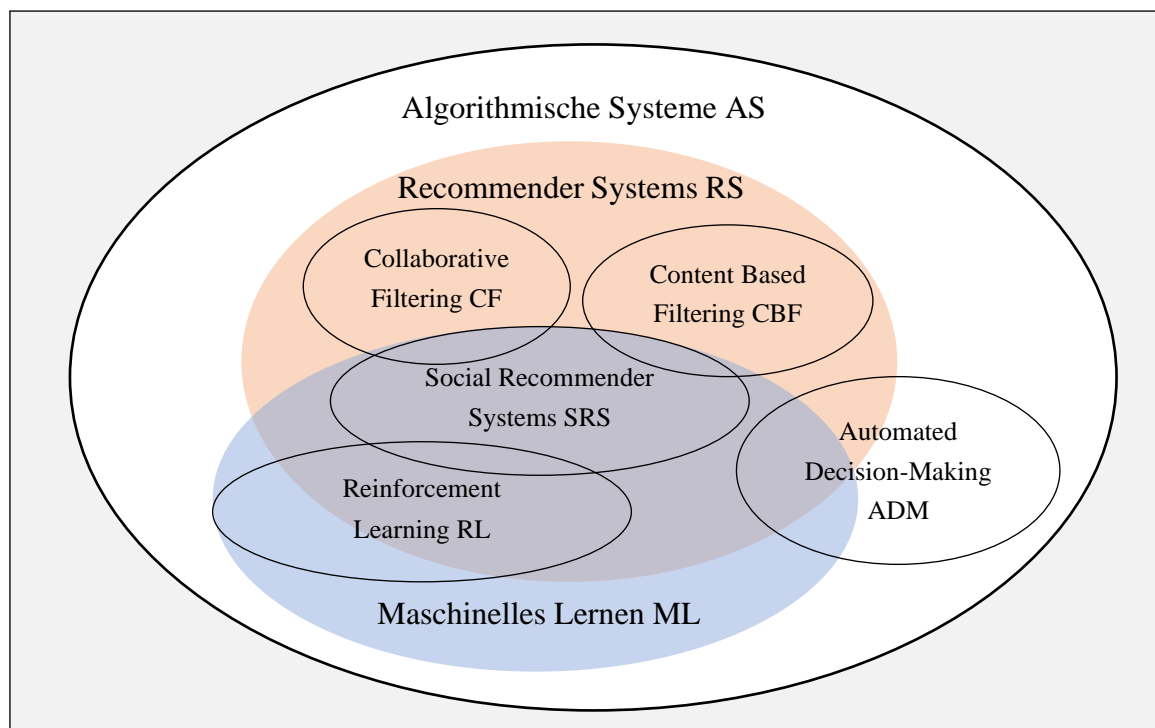


Abbildung 4: Grafische Darstellung der Technikbegriffe

KI ist ein Schlagwort, das insbesondere seit den späten 2010er Jahren medial, politisch und gesellschaftlich verwendet wird, um moderne ML-Verfahren zu beschreiben. Aus Sicht der Informatik ist KI nicht einer bestimmten Technologie zuordenbar, weshalb der Begriff insbesondere von Laien verwendet wird, um die Technologien aus Nutzerperspektive zu beschreiben. KI umfasst Verfahren, mit denen Tätigkeiten die traditionellerweise von Menschen durchgeführt werden, mit technischen Hilfsmitteln wie Robotik und Computer-Software realisiert werden (Hoffmann-Riem, 2020, S. 2–3). ML-basierte RS werden als KI bezeichnet, da das Auswählen und Empfehlen von Informationen und Inhalten nicht mehr durch die Chefredaktion einer Zeitung, den Verkäufer im Plattengeschäft oder die Bibliothekarin an der Theke bewerkstelligt wird, sondern durch ein Computersystem.

### 3.3.2 Empfehlungssysteme

Obschon der Begriff «Algorithmische Systeme» im DSA mehrfach Erwähnung findet, wird er im Art. 3 «Begriffsbestimmungen» nicht definiert. Ein Algorithmus ist nach Cormen (2009, S. 5) die Methode zur Lösung eines Berechnungsproblems und beschreibt das zur Realisation einer Input-Output-Beziehung verwendete Lösungsverfahren. Dasgupta et al. (2006, S. 11–12) definieren Algorithmen als präzise, eindeutige, mechanische, effiziente und korrekte Verfahren, die heute zur Lösung aller Arten von Problemen und zur Entwicklung neuer Anwendungen eingesetzt werden. Im DSA wird der Begriff



«Algorithmische Systeme» jeweils im Art. 35 «Risikominderung» und Art. 40 «Datenzugang» mit «Empfehlungssystemen» zusammen genannt. «Empfehlungssysteme» werden im Art. 3 des DSA wie folgt definiert:

*«Vollständig oder teilweise automatisiertes System, das von einer Online-Plattform verwendet wird, um auf ihrer Online-Schnittstelle den Nutzern bestimmte Informationen vorzuschlagen oder diese Informationen zu priorisieren, auch infolge einer vom Nutzer veranlassten Suche, oder das auf andere Weise die relative Reihenfolge oder Hervorhebung der angezeigten Informationen bestimmt.»*

Diese Definition deckt sich grösstenteils mit gängigen RS-Definitionen aus der wissenschaftlichen Literatur.

Beispielsweise definieren Shokeen & Rana (2020, S. 633–634) RS als einen Subtyp von Information-Retrieval-Systemen, die automatisiert auf Grundlage von hinterlegten Präferenzen die individuelle Relevanz bestimmter Inhalte für einzelne Nutzende bestimmen. Durch die Reduktion der angezeigten Informationen wird unter Berücksichtigung der individuellen Interessen der Nutzenden die Komplexität des Inhaltsangebotes verringert. Dadurch soll nicht zuletzt das Nutzererlebnis verbessert werden (S. 638). Laut Schwemer (2021, S. 62) handelt es sich bei Empfehlungssystemen um automatisierte Systeme, die zur abfragelosen Empfehlung von Inhalten wie News, Produkten oder Unterhaltungsmedien eingesetzt werden sowie die Auswahl und Priorisierung von Informationen auf einzelne Nutzende personalisiert durchführen. Gemäss Pu et al. (2011, S. 157), die mit ResQue ein vielbeachtetes RS-Evaluationsverfahren kreiert haben, ist ein RS ein Computersystem, das Nutzenden auf der Grundlage ihres Verhaltens oder im System hinterlegten Präferenzen selbstständig Inhalte vorschlägt.

Die drei Definitionen aus der Literatur zeigen allesamt ein hohes Ausmass an Übereinstimmung mit der Definition im DSA. In der vorliegenden Arbeit werden RS als eine spezifische Art von AS betrachtet, da im DSA keine präzisere Abgrenzung erfolgt.

Mit dem «Recommender Systems Handbook» haben Ricci, Rokach, Shapira und Kantor (2011) eines der umfassendsten RS-Standardwerke geschaffen. Für die Systematisierung von RS wird deshalb nachfolgend auf die etablierten Konzepte aus diesem Fachbuch Bezug genommen. Demnach bestehen RS aus einer Empfehlungsfunktion (S. 4–5), den Daten- und Wissensquellen (S. 7–8) und einer Empfehlungstechnik (S. 10). Die Empfehlungsfunktion beschreibt den direkten Mehrwert für die Anbietende sowie Nutzende des Systems (S. 6–7), während die Daten- und Wissensquellen die empirische Grundlage für die Empfehlungsgenerierung bilden und allgemein hin abgegrenzt werden

können als gespeicherte Informationen zu Items, Nutzenden und bisherigen Transaktionen (S. 7–10). Inhalte, Artikel oder Dokumente werden in der englischsprachigen Literatur als Item zusammengefasst und stellen die zu empfehlende Einheit im Datenbestand dar (S. 1–2). Bei umfangreichen Datenbeständen kommen ML-Verfahren zum Einsatz, die Kandidaten generieren. Dabei handelt es sich um Items, deren Relevanz mittels ML ermittelt wurde (S. 63–65). Items oder Kandidaten werden vom Empfehlungssystem ausserdem rangiert oder auch – aus dem Englischen eingedeutscht – «geranked» (S. 2). Die Empfehlungstechnik beschreibt, nach welchem Vorgehensmuster die Vorhersagen über die individuelle Relevanz eines Inhaltes für einzelne Nutzende getroffen werden (S. 10–11). Die Empfehlungstechniken werden wie folgt systematisiert:

- «Content-based Filtering» oder inhaltsbasierte RS (CBF) empfehlen Inhalte, die denjenigen ähnlich sind, die in der Vergangenheit für dieselben Nutzenden von Interesse waren (S. 11).
- «Collaborative Filtering»-basierte RS (CF) empfehlen Nutzenden Inhalte, die für andere Benutzer mit vergleichbaren Interessen oder ähnlichem Nutzerverhalten in der Vergangenheit bedeutsam waren (S. 11).
- Demographische RS empfehlen Inhalte auf der Grundlage des demographischen Profils der Nutzenden (S. 12).
- Wissensbasierte RS empfehlen Inhalte auf der Grundlage von strukturiert hinterlegtem Fachwissen zu Inhalten oder Artikeln und Angaben dazu, wie diese Merkmale den Bedürfnissen von Nutzerinnen und Nutzern entsprechen (S. 12).
- «Community-based» RS empfehlen Inhalte auf der Grundlage der Interessen und Vorlieben von Menschen im sozialen Umfeld der Nutzenden (S. 13).
- Hybride RS basieren auf einer Kombination mehrerer Techniken (S. 13).

### 3.3.3 Collaborative Filtering

Die Hauptkategorien der Collaborative-Filtering-Methoden (CF) in RS werden als speicherbasiert («Memory-based»), modellbasiert («Model-based») und hybrid eingegrenzt (Papadakis et al., 2022, S. 37; Su & Khoshgoftaar, 2009, S. 3).

Sie lassen sich wie folgt beschreiben:

- Speicherbasiertes CF heisst, dass jedes Mal, wenn eine neue Empfehlung benötigt wird, Berechnungen auf Grundlage früherer Interaktionen sowie Daten anderer Nutzenden durchgeführt werden (Papadakis et al., 2022, S. 37). Speicherbasierte Systeme sind leicht zu implementieren, nehmen neue Daten unmittelbar auf, sind nicht von der Analyse des Inhalts abhängig, aber auf menschliches Feedback sowie

riesige Datenmengen angewiesen und können ohne Daten keine zuverlässigen Empfehlungen generieren (Su & Khoshgoftaar, 2009, S. 3).

- Modellbasiertes CF bedeutet, dass anhand Daten früherer Interaktionen sowie Daten anderer Nutzenden ein Nutzermodell generiert wird, das zur Erstellung von Empfehlungen für aktive Nutzende verwendet wird, ohne auf einzelne Datensätze zuzugreifen (Papadakis et al., 2022, S. 37). Modellbasierte Systeme sind skalierbar, haben in der Regel eine befriedigende Voraussagegenauigkeit, beruhen auf einer nachvollziehbaren Logik, sind aber von teuren Modellbildungstechnologien abhängig und müssen auf Kosten der Komplexitätsreduzierung auch Informationsverluste verkraften (Su & Khoshgoftaar, 2009, S. 3).
- Hybrides CF bedeutet, dass CF mit anderen Methoden, darunter CBF-Methoden, kombiniert wird (Su & Khoshgoftaar, 2009, S. 11). Hybride CF-Systeme überwinden Einschränkungen von reinen CF-Methoden und können unter Umständen eine optimierte Leistung und Vorhersagegenauigkeit bewirken. Dazu benötigen sie möglicherweise Informationen aus externen Quellen, sind komplexer im Aufbau und kostenintensiv in der Umsetzung (Su & Khoshgoftaar, 2009, S. 3).

### 3.3.4 Social Recommender Systems

Soziale Medien sind digitale Dienste, auf denen nutzergenerierte Inhalte bereitgestellt, konsumiert und weiterverbreitet werden. Sie werden von Menschen für den Konsum von unterschiedlichsten multimedialen Inhalten verwendet: von Nachrichten und Unterhaltung hin zu privaten Inhalten von Familie und Angehörigen. Sie dienen somit der Befriedigung von informationellen wie emotionalen Bedürfnissen, zum Teilen von persönlichen Meinungen und Erfahrungen sowie zum Erteilen von Bewertungen oder Empfehlungen zu Produkten und Dienstleistungen (Kapoor et al., 2018, S. 531–532). Laut Shokeen & Rana (2020) sind Social Recommender Systems (SRS) Empfehlungssysteme, die für die Generierung von Empfehlungen Daten über soziale Beziehungen nutzen. In SRS werden CF-Algorithmen mit anderen Empfehlungstechniken kombiniert (S. 638). Diese können unter anderem sein:

- Deep-Learning-basierte RS verwenden mehrschichtige Relevanzberechnungsverfahren und werden als ML und KI klassifiziert. Dazu zählen insbesondere künstliche neuronale Netze (S. 639–641).
- Fuzzy-basierte RS sind auf der Fuzzylogik nach Zadeh begründet und eignen sich zur Analyse von linguistischen oder verbalen Daten (S. 645–647).

- Clustering-basierte RS detektierten in der Datengrundlage Ähnlichkeiten zwischen Nutzenden oder Inhalten. Dazu zählt auch der gängige K-Nearest-Neighbor-Algorithmus, mit dem ähnliche Elemente gruppiert werden können (S. 647).
- Semantische RS werten soziale und behaviorale Informationen aus, wie etwa Geodaten oder Interaktionslogs von Nutzenden, um deren Tagesablauf zu modellieren und die vielfältigen Beziehungsebenen zu anderen Nutzenden auszuwerten (S. 650–651).

### 3.3.5 Maschinelles Lernen und Social Recommender Systems

Die heute gängige Definition von maschinellem Lernen von Koza et al. (1996, S. 153) entspringt einer Paraphrasierung von Samuel (1959) und lässt sich auf Deutsch wiedergeben als «Fachgebiet, das Computern die Fähigkeit verleiht, zu lernen, ohne ausdrücklich programmiert zu werden». Laut Batrinca & Treleaven (2015, S. 103–104) ist der Zweck von ML umfassende Datenmengen mit unzähligen Variablen zu analysieren. ML wird weiter unterteilt in:

- Überwachtes Lernen («Supervised Learning»), das sowohl auf Input- als auch Output-Trainingsdaten basiert und für Klassifikations- und Regressionsaufgaben verwendet wird.
- Unüberwachtes Lernen («Unsupervised Learning»), das aufgrund fehlender Trainingsdaten ausschliesslich auf Input-Daten basiert und für Clustering-Aufgaben verwendet wird.

Aktuelle Forschung fokussiert sich im Zusammenhang mit dem Einsatz von ML für CF-Methoden vor allem auf Deep Learning. Dies bezeichnet ML-Algorithmen, die zwischen Input- und Output-Schicht eine hohe Anzahl dazwischenliegende versteckte Berechnungsschichten haben (Bobadilla et al., 2020, S. 1–2). Seichte, auch «Shallow Learning Models» genannt, werden im Rahmen dieser Bachelor-Thesis nicht bearbeitet.

Für SRS ist zudem das Konzept Reinforcement Learning (RL) von Relevanz (Arogyaswamy, 2020, S. 834; Pérez & Fernández-Aller, 2022, S. 7). Gemäss Alpaydin (2010) kommt RL in der Regel in unüberwachten ML-Architekturen vor (S. 471–472). RL beschreibt das stetige Lernen von registrierten Feedbacks von Nutzenden, um das ML-Modell laufend anzupassen (S. 447–448). Das Auswerten der Nutzerinteraktionen ermöglicht eine schnelle Bestimmung von Relevanz und beschleunigt allgemein den Lernprozess des Systems (S. 471).

Sowohl modellbasierte wie speicherbasierte CF-Methoden können mittels ML-Verfahren umgesetzt werden. Bei speicherbasierten CF-Systemen hat der Einsatz von K-Nearest-Neighbor-Algorithmen zur Identifikation ähnlicher Inhalte und Nutzenden bereits Tradition, während in jüngerer Vergangenheit bei modellbasierten Systemen zur Relevanzbestimmung oft Matrix-Faktorisierung eingesetzt wird, wobei es sich um ein Verfahren der Dimensionsreduktion handelt (Bobadilla et al., 2020, S. 1–2).

### 3.4 Empirischer Forschungsstand

#### 3.4.1 Typisierung von Social-Media-Algorithmen

Wie bereits erläutert gelten die strengsten Transparenzpflichten im DSA für Onlineplattformen, die im Unionsgebiet über 45 Millionen Nutzende erreichen und «Very Large Online Platforms», kurz VLOP, genannt werden. Nach aktuellen Zahlen betrifft dies insbesondere Youtube mit 401.7 Millionen Nutzenden (Google, 2023), Facebook mit 255 Millionen Nutzenden sowie Instagram mit 250 Millionen Nutzenden (Meta, 2023), TikTok mit 150 Millionen Nutzenden (TikTok, 2023) und zuletzt Twitter mit rund 59.8 Millionen Nutzenden (Twitter, 2023b).

Social-Media-Plattformen haben in der Vergangenheit kaum Einblick in die Funktionsweise ihrer RS gewährt. Was über die Algorithmen bekannt ist, entstammt sporadischen Beiträgen von Angestellten der Konzerne an Fachkonferenzen sowie Blogposts auf den eigenen Webseiten. Entsprechend lassen sich die Angaben nicht unabhängig überprüfen und der Mangel an gesichertem Wissen zu den exakten Mechanismen führt zu einer gewissen Unschärfe in der Erforschung von SRS. Im nachfolgenden Abschnitt werden dennoch die Informationen, die öffentlich zugänglich sind, aufgelistet.

Drei Google-Angestellte haben im Jahr 2016 an der RecSys, einer RS-Fachkonferenz, die Architektur des Youtube-Empfehlungssystems vorgestellt (Covington et al., 2016). Das Empfehlungssystem verwendet zur Generierung von Kandidaten ReLU, ein künstliches neuronales Netz, um das Nutzungsverhalten, darunter Suchverlauf oder abgespielte Videos, auszuwerten. Unter Berücksichtigung von demographischen Daten identifiziert es anschliessend mittels Approximations-Algorithmus nach K-Nearest-Neighbor-Prinzip ähnliche Nutzende. Beim Ranking der generierten Kandidaten werden die Eigenschaften der Videos, beispielsweise die Inhaltssprache, höher gewichtet.

In einem Blogbeitrag aus dem Jahr 2019 wird der Instagram-Recommendier, der Inhalte für die Startseite der Applikation selektiert, erläutert. Zur Approximation und Kandidatengenerierung verwendet Instagram ein selbst entwickeltes künstliches neuronales Netz,

das sich «ig2vec» nennt und auf dem NLP-Algorithmus «Word2vec» basiert. Anschließend kommen für den Rankingprozess weitere künstliche neuronale Netze zum Einsatz, die gemeinsam als «Distillation Model» bezeichnet werden. Nach Angaben von Meta werden so jeweils von 500 generierten Kandidaten nur die 25 passendsten Beiträge ausgewählt (Medvedev et al., 2019).

Eine Gruppe von Meta-Vertretern hat im Jahr 2020 an der Konferenz «Symposium on High Performance Computer Architecture HPCA» Details zur Architektur des Facebook-Recommendors offengelegt (Gupta et al., 2020). Daraus geht hervor, dass das RS auf einem künstlichen neuronalen Netz unter Verwendung von ReLU-Algorithmen basiert.

In einem Blogpost aus dem Jahr 2023 hat der Onlinedienst Twitter die Funktionsweise ihres Tweet-Recommendors dargestellt. Ein künstliches neuronales Netz erstellt Kandidaten, die zu 50% aus Inhalten aus dem Netzwerk der Nutzenden bestehen und 50% aus Inhalten von ausserhalb des persönlichen Netzwerks. Dabei werden graphenbasierte ML-Verfahren wie RealGraph, SocialGraph sowie der Twitter-eigene GraphJet eingesetzt (Twitter, 2023a).

TikTok hat die Funktionsweise oder Architektur des RS nie selbst offengelegt. Im Dezember 2021 veröffentlichte jedoch die New York Times, basierend auf durchgesickerten internen TikTok-Dokumenten, einen Artikel über die algorithmische Architektur des TikTok-RS. Das darin beschriebene RS basiert ebenfalls auf ML, wird aber sonst nicht präzisiert (Smith, 2021).

Werden die öffentlich zugänglichen Informationen über die RS von Facebook, Instagram, Twitter oder Youtube resümiert, lässt sich ableiten, dass schwergewichtig ML, namentlich künstliche neuronale Netze, zum Einsatz kommen.

### **3.4.2 Risiken von sozialen Medien**

Die Erforschung der psychologischen und gesellschaftlichen Risiken von sozialen Medien ist ein aktuelles Phänomen in der Social-Media-Forschung. Ein roter Faden, der sich von den frühen 2010er bis in die 2020er Jahre durch Social-Media-Forschung zieht, ist zwar die Erforschung der psychologischen Risiken sozialer Medien sowohl bei Erwachsenen wie bei Jugendlichen (Gonzales & Hancock, 2011, S. 79–80; Vogel et al., 2014, S. 206–208; Wolfers & Utz, 2022, S. 1–2; Popat & Tarrant, 2023, S. 323–324). Davon abgesehen fokussierten sich wissenschaftliche Analysen von Social-Media-Risiken in den 2010er Jahren häufig auf Fragestellungen rund um Privatsphäre und Online-Betrug (Kapoor et al., 2018, S. 546–547) oder Cyber-Mobbing und Trolling (Baccarella et al., 2018, S. 432). Gegen Ende der Dekade und mit Beginn der 2020er-Jahre rückten im

wissenschaftlichen und öffentlichen Diskurs zunehmend systemische, gesellschaftliche und politische Risiken in den Vordergrund. Dies wird mit Desinformation und manipulativen Social-Media-Kampagnen, beispielsweise während der US-amerikanischen Wahlen im Jahr 2016 sowie der Brexit-Abstimmung im Vereinigten Königreich in Verbindung gebracht (Bastick, 2021, S. 1–2; McKay & Tenove, 2021, S. 1–3). In Bezug auf RS wird spezifisch das Risiko von Filterblasen diskutiert. Dabei handelt es sich um einen Effekt der abnehmenden Vielfalt von empfohlenen Inhalten infolgedessen, dass die RS Inhalte priorisieren, die eine hohe Interaktionswahrscheinlichkeit haben. Das sind überdurchschnittlich häufig Inhalte, die die Weltanschauung der Nutzenden bestätigen und polarisierte Themen betreffen (Alvarado & Waern, 2018, S. 2; Latzer & Festic, 2019, S. 9; Rader, 2017, S. 73).

### **3.4.3 FATED-Faktoren**

#### **3.4.3.1 Gerechtigkeit und Ethik**

Das philosophische Teilgebiet der Ethik beschäftigt sich mit moralischen Bewertungen von menschlichem Verhalten (Robles Carrillo, 2020, S. 2). Moralische Bewertungen und daraus abgeleitete ethische Modelle haben sich historisch betrachtet stets gewandelt und in verschiedenen Gesellschaften, Kulturen und Zivilisationen unterschiedliche Konnotationen gehabt (S. 2). Es existieren frappante Unterschiede bezüglich Moralvorstellungen der westlich-europäischen Welt und der asiatischen, amerikanischen, afrikanischen oder islamischen Welt, obschon eine ständige gegenseitige Beeinflussung der Kulturen, etwa durch Eroberung oder Kolonisierung, stattgefunden hat (S. 3).

Obwohl die Bedeutung von Gerechtigkeit von KI-Anbietern häufig hervorgehoben wird, gibt es keine allgemein anerkannte Definition der algorithmischen Gerechtigkeit. Bei der Entwicklung von KI-Systemen befindet sich das unerwünschte Reproduzieren von Vorurteilen im Zentrum der Debatte (Shin, 2020b, S. 546). Der Grundsatz der «Fairness» kann in diesem Sinne als unparteiische Behandlung der Nutzerinnen und Nutzer durch KI-Systeme definiert werden (Fjeld et al., 2020, S. 49), beispielsweise in Bezug auf marginalisierende Merkmale wie Ethnie oder Geschlecht (Salgado-Criado & Fernandez-Aller, 2021, S. 58). Auch KI-getriebene Empfehlungssysteme bieten erwiesenermaßen eine unterschiedliche Empfehlungsqualität für verschiedene Nutzende je nach Geschlecht, Alter oder ethnischer Zugehörigkeit (Di Noia et al., 2022, S. 70).

Gerechtigkeit kann auf der individuellen Ebene betrachtet werden, in Bezug darauf, ob ein KI-System jede einzelne Person in Übereinstimmung mit dem Zweck des Systems behandelt, sowie ob eine Individualperson gleich behandelt wird wie die anderen

Nutzenden. Auf einer kollektiven Ebene bedeutet Gerechtigkeit, dass Personengruppen, die marginalisierende Merkmale teilen, von den Systemen gleich behandelt werden wie Personengruppen ohne diese Merkmale (Di Noia et al., 2022, S. 70; Salgado-Criado & Fernandez-Aller, 2021, S. 58). Die Brisanz von kollektiver Gerechtigkeit bezüglich marginalisierten Gruppen wird in unzähligen Arbeiten diskutiert (Fjeld et al., 2020, S. 51).

Die Dienstleistungen marktführender Technologiekonzerne stehen unter Verdacht, Ungerechtigkeiten zu amplifizieren. Da Menschen in ihrer Entscheidungsfindung voreingenommen sind, ist es durchaus möglich, dass Entwicklerinnen unbeabsichtigt bei der Entwicklung von KI-Systemen ihre eigene Voreingenommenheit in Applikationen einfließen lassen. Diskriminierungen und Vorurteile können, auch wenn die Algorithmen aus mathematischer Sicht korrekt sind, in den Systemen auftreten. Darum sind vertiefte Analysen dazu erforderlich, welche langfristigen Auswirkungen Algorithmen auf Gerechtigkeit haben können, damit nicht riskanterweise nur KI-Projekte mit Potenzial zur Gewinnmaximierung realisiert werden (Salgado-Criado & Fernandez-Aller, 2021, S. 56).

Algorithmische Verzerrungen basieren auf systematischen Unter- oder Überschätzungen von Wahrscheinlichkeiten für bestimmte Personengruppen. Sie können ihre Wurzel in unrepräsentativen, fehlerhaften oder verzerrten Daten haben, die im Modell amplifiziert werden oder aus der Tatsache resultieren, dass die Daten die Realität ungenügend abbilden (Di Noia et al., 2022, S. 70; Fjeld et al., 2020, S. 47).

Anstatt dass andere Anspruchsgruppen wie Content Creators oder Plattformanbieter selbst einbezogen werden, wird Gerechtigkeit meist aus der Perspektive der privaten Endnutzerinnen und -nutzer betrachtet. Eine umfassende Gerechtigkeit aus rechtlicher Sicht bedingt aber auch Regeln zur Inhalts- und Datenqualität sowie wirksame Mechanismen zum Testen der AS auf Einhaltung der Rechtsvorschriften (Di Noia et al., 2022, S. 71). Ein weiterer Lösungsansatz wäre, dass ML-basierte Anwendungen stets anhand von Daten trainiert werden müssen, was insbesondere eine strenge Regulation oder ein Verbot von Systemen des unüberwachten ML und RL bedingen würde (Salgado-Criado & Fernandez-Aller, 2021, S. 57).

Gemäss Wang et al. (Wang et al., 2023) gibt es keine einheitliche Meinung dazu, ob die Öffentlichkeit AS und RS als unfair bewertet, da Meinungsbefragungen gemischte Resultate zeigen. Sowohl automatisierte Entscheidungsprozesse wie auch menschengemachte Entscheidungen können als gerecht oder ungerecht eingestuft werden. Im Einzelfall hängt die Wahrnehmung von der Komplexität der Aufgabe, Tragweite der Entscheidungen, den Fähigkeiten des zuständigen Systems beziehungsweise des



Menschen ab (S. 2–4) und wird durch die Vertrautheit mit der entscheidenden Institution bestimmt (S. 8).

Im DSA wird Gerechtigkeit insbesondere mit der Pflicht zu Risikobewertung (Art. 34) und Risikominderung (Art. 35) adressiert, ohne konkrete ethische Regeln zu erlassen.

### **3.4.3.2 Haftbarkeit**

Die Grundsätze des Themas Haftbarkeit und Rechenschaftspflicht beziehen sich auf drei Phasen der KI-Implementation: die Entwurfsphase vor der Einführung, Überwachungsphase während der Verwendung und Rechtshilfe sowie Strafverfolgung beim Auftreten von Schäden (Fjeld et al., 2020, S. 29).

In der Entwurfsphase können Verifizierungs- und Replizierungsprozesse sicherstellen, dass KI-Systeme so funktionieren wie beabsichtigt, während Folgenabschätzungen für die Identifizierung, Verhinderung und Abschwächung von KI-Folgen für Mensch und Umwelt notwendig sind (Fjeld et al., 2020, S. 29–31). Die Vorhersage möglicher unbeabsichtigter Folgen ist ein essenzieller Baustein, um algorithmische Verantwortung und somit Haftbarkeit herzustellen (Shin, 2020b, S. 546).

Damit in der Überwachungsphase Rechenschaft und Haftbarkeit gewährleistet sind, müssen die Systeme überprüfbar und insofern verbesserungsfähig sein, dass fehlerhafte Entscheidungen zur Verbesserung in das System zurückgeführt werden können (Fjeld et al., 2020, S. 31). Ein Auditsystem, beispielsweise durch neu konzipierte Überwachungsorgane, ist erforderlich, sowohl um KI-Praktiken zu beaufsichtigen, wie auch um KI-geschädigten Personen entsprechende Einsprache- und Beschwerdemöglichkeiten zu gewähren (Fjeld et al., 2020, S. 32; Salgado-Criado & Fernandez-Aller, 2021, S. 63)

Die Phasen Rechtshilfe sowie Strafverfolgung beschreiben Prozesse nachdem an Menschen, Gesellschaft oder Umwelt ein systembedingter Schaden entstanden ist. Um Haftbarkeit sicherzustellen, müssten automatisierte Handlungen ebenso wie die Folgen menschlicher Handlungen strafrechtlich belangbar sein, da die Systeme in sensiblen Umgebungen eingesetzt werden und ihre Resultate tatsächliche Auswirkungen haben. Dies verlangt auf juristischer Ebene neue, innovative Ansätze, um für einen Schaden, der durch Computer-Systeme entsteht, die verantwortlichen Personen oder Einrichtungen zu identifizieren und zu belangen (Fjeld et al., 2020, S. 33–34). In bestehenden Policy-Vorschlägen wird diskutiert, welche Instanzen im Schadenfall zur Verantwortung gezogen werden sollten, darunter die Regierung, Unternehmen, Forscher, Entwickler oder in bestimmten Fällen sogar Nutzende (Fjeld et al., 2020, S. 35–36).

Da Gerichte reaktive Institutionen sind und erst im Schadenfall über die Haftung entscheiden, bedarf es einer Regulierungsbehörde, die präventiv eingreift. Auch die Self-Governance-Massnahmen der Industrie ersetzen externe Überwachung zur Sicherstellung von Haftbarkeit nicht. Um Haftbarkeit herzustellen, können aber Konzepte der gesetzlichen Haftung und der Self-Governance kombiniert werden (Roski et al., 2021, S. 1586).

Berücksichtigt werden muss, dass das Vertrauen in staatliche Akteure nicht zwangsläufig ausgeprägter als in Privatunternehmen ist und die Glaubwürdigkeit der Governance-Akteure eine gravierende Rolle spielt. So zeigt eine Befragung, dass US-Amerikanerinnen und US-Amerikaner beim Management von KI-Technologie mehr Vertrauen in Technologieunternehmen und NGOs haben als in ihre eigene Regierung (B. Zhang & Dafoe, 2020, S. 191).

Im DSA wird den Dienstleistern in Form der juristischen Person die Haftbarkeit zugewiesen, während Privatpersonen wie Manager oder Entwicklerinnen nicht direkt belangt werden können. Anlaufstellen für Beschwerden sollen national gewährleistet werden. Ferner sieht der DSA eine aufgeteilte Regulation zwischen Staat, unabhängigen Prüfgesellschaften und den Dienstleistern selbst vor. Die exakte Aufgabenteilung zwischen nationalen Koordinationsstellen und der zentralen EU-Stelle wird im DSA nicht abschliessend geklärt.

### **3.4.3.3 Transparenz**

Transparenz bedeutet, dass die innere Funktionsweise des AS dargelegt wird und dass offensichtlich ist, welche Daten verwendet werden. Grundsätzlich sollten die verschiedenen Schritte von der Eingabe bis zur Ausgabe für Nutzende ersichtlich und verständlich sein, während Fachleute in der Lage sein sollten, zu entscheiden, ob die Anwendung statistisch und mathematisch für die jeweilige Aufgabe geeignet ist und ob das Ziel des AS gültig und vertretbar ist (Haque et al., 2023, S. 4; Shin, 2020b, S. 547). Transparenz ist eng mit Gerechtigkeit verbunden: Wenn AS gerecht sind, werden sie als transparent empfunden und umgekehrt. Während Transparenz eher das AS als Ganzes beschreibt, steht Gerechtigkeit in Bezug mit den Nutzenden und damit, was ihnen widerfährt (Shin et al., 2022, S. 12). In manchen Untersuchungen wird hervorgehoben, dass Transparenz ein Schlüsselfaktor für die Schaffung von Vertrauen ist und dass sich die Nutzenden bei intransparenten Systemen beispielsweise überwacht fühlen (Eg et al., 2023, S. 10).

ML-basierte Systeme werden als Blackbox-Algorithmen bezeichnet und daher im Allgemeinen als weniger transparent wahrgenommen als regelbasierte Systeme. Dies stellt eine besondere Herausforderung für die Schaffung von Transparenz von ML-basierten RS dar (Shin, 2020b, S. 547; Wang et al., 2023, S. 8). Aus Sicht der Nutzenden mangelt

es ML-Systemen vor allem an Transparenz dazu, welche Kriterien bei der automatischen Kuration verwendet werden und technische Information wird in der Regel so abgebildet, dass sie für Laien nicht verständlich ist (Shin, 2020b, S. 547–548). Gemäss Sigfrids (2023, S. 4) bedingt Vertrauen, dass Nutzende ein Grundverständnis der Funktionsweise des KI-Systems haben und nachvollziehen können, wie die Systeme Entscheidungen treffen und auch die Anbieter darlegen können, warum gewisse Resultate generiert werden. Ferner bedeutet Transparenz, dass Nutzende Informationen dazu erhalten, wie unerwünschte Wirkungen erkannt, angehalten und verhindert werden.

Im DSA wird eine Erklärung der Kriterien der Empfehlungsfunktion und der verwendeten Parameter in den allgemeinen Geschäftsbedingungen (Art. 27) vorgeschrieben.

#### **3.4.3.4 Erklärbarkeit und Verständnis**

Da sowohl Input und Output von ML-Anwendungen im Dunkeln bleiben, wissen die Nutzenden meist nicht, warum sich das System für einen Lösungsweg entschieden hat. Erklärbarkeit von KI wird eXplainableAI (XAI) genannt und bedeutet, den Nutzenden verständliche, präzise, ausreichende und massgebende Informationen dazu zu geben, warum das System eine bestimmte Entscheidung getroffen hat (Di Noia et al., 2022, S. 73; Haque et al., 2023, S. 2–3; Salgado-Criado & Fernandez-Aller, 2021, S. 58; Shin, 2020b, S. 547–548; Shin et al., 2022, S. 4). «User Awareness» beschreibt auf der Gegenseite das Verständnis der Nutzenden, wie und warum eine bestimmte Empfehlung für sie erstellt wurde (Shin, 2020b, S. 548). Das mentale Modell, mit dem Nutzerinnen und Nutzer sich AS vorstellen, wird als «algorithmic imaginary» bezeichnet, was nicht heisst, dass es imaginär oder unecht ist, sondern dass Nutzende es mittels ihrer Vorstellungskraft produzieren (Bucher, 2017, S. 31).

Wie beim Faktor Transparenz ist im Bereich der Erklärbarkeit eine Unterscheidung zwischen datengesteuerten Systemen und regelbasierten Systemen angebracht, da die Erklärbarkeit bei regelbasierten ADM einfacher ist, als im Fall von datenbasierten Systemen, die ihre autonomen Entscheide oft nicht offenlegen (Wang et al., 2023, S. 3). ML-gesteuerte Systeme haben durch Applikationen, die wir täglich nutzen, zwar ihren Weg in unser Leben gefunden, aber geben etwa Nachrichten-Websites, soziale Medien und Onlineshops ihren Nutzenden selten Aufschluss darüber, was ein bestimmtes Ergebnis, z. B. eine Empfehlung verursacht hat (Eg et al., 2023, S. 1; Haque et al., 2023, S. 1).

Die Nutzenden neigen in diesen Situationen dazu, AS und getroffene Entscheidungen auf der Grundlage ihres eigenen, womöglich limitierten Wissens über diese Anwendungen zu bewerten (Shin, 2020a, S. 2) und einige Studien haben sogar ergeben, dass nur wenige Nutzende überhaupt von der Tragweite algorithmischer Auswahl- und

Entscheidungsprozesse im Internet wussten (Alvarado et al., 2020, S. 4; Bucher, 2017, S. 31–32; Eg et al., 2023, S. 10; Klumbyte et al., 2020, S. 2; Sundar, 2020, S. 80).

Darum ist es von Bedeutung, die Anspruchsgruppen von XAI im Rahmen von ihren Fähigkeiten und Kenntnissen zu betrachten, um die unterschiedlichen Anforderungen zwischen Fachpersonal, Regulierungsbehörden und Nutzenden zu berücksichtigen (Haque et al., 2023, S. 5; Liao et al., 2022, S. 148). Laut Haque et al. (2023) kann eine hybride Erklärung sowohl Informationen zu den verwendeten Parametern und ihren Abhängigkeiten wie auch Angaben zur Funktionsweise des Systems enthalten. Unter Abhängigkeiten versteht man die Zusammenhänge zwischen Input und Output des AS. Meistens bevorzugen Nutzende hybride Erklärungen, die textliche und visuelle Erklärungen kombinieren (S. 5). Ferner können kontextbezogene Angaben belangvoll sein, die fach- oder anwendungsspezifisch sind (S. 6). Nutzende profitieren dabei zugleich von Erklärungen, die personalisierte Angaben zur Priorisierung sowie einordnende Informationen zur erwarteten Genauigkeit enthalten (S. 8). Das bedeutet, im Idealfall wird Nutzenden eine automatische oder abrufbare XAI-Erklärung zu jedem Output angeboten (S. 9).

Im DSA beschränkt sich die Erklärung der Empfehlungsfunktion auf eine globale, theoretische Beschreibung in den AGB (Art. 27), was eher der Verbesserung von Transparenz dient als der Optimierung von Erklärbarkeit. Lokale Erklärungen nach XAI sind laut DSA einzig bei Werbungen vorgesehen.

#### **3.4.3.5 Definierbarkeit**

Als Definierbarkeit wird im Rahmen dieser Bachelor-Thesis die Kontrolle von Nutzenden über die algorithmischen Systeme bezeichnet. Nutzerkontrolle beschreibt sowohl Einfluss auf den Empfehlungs- als auch auf den Datenerfassungsprozess (Shin et al., 2022, S. 2).

Wie Bucher (2017) feststellte, verursacht die Tatsache, dass Facebook-Algorithmen beeinflussen, was sie zu sehen bekommen, bei Nutzenden ein Unwohlsein (S. 39), was dazu führte, dass manche sogar Strategien zur Datenverschleierung entwickelten, indem sie versuchten, das AS zu verwirren und eine gewisse Kontrolle über die Vorschläge zu erringen (S. 41). Dies zeigt, dass die Nutzenden teilweise einen inhärenten und intuitiven Drang verspüren, die AS zu steuern.

Nach Alvarado & Waern (2018) existieren verschiedene Möglichkeiten, wie Nutzende Kontrolle über das System ausüben können. Durch algorithmisches Profilmanagement können Nutzende algorithmisch erstellte Profile überprüfen und das Modell, das das System von ihnen erstellt hat, bearbeiten (S. 6). Algorithmische Benutzerkontrolle bedeutet

überdies, dass die Applikation die Möglichkeit bietet, AS auszuschalten, ihnen explizit negative Feedbacks zu erteilen, die Datenquellen, auf deren Grundlage Empfehlungen generiert werden zu beeinflussen, sowie selektiv unerwünscht Datenquellen vollständig auszuschalten, sowohl bevor als auch nach der Datensammlung (S. 7)

Es ist anzumerken, dass Nutzerkontrolle auch durch die Verteilung der Handlungsfähigkeit zwischen Nutzer und System definiert wird. Dabei spielt es eine Rolle, wie und durch wen Aktivität oder Reaktivität provoziert werden und zu welchem Grad eine Zusammenarbeit zwischen Nutzenden und System zur Erfüllung einer Aufgabe erforderlich ist. Während bei Personen mit hoher Nutzungsintensität und technikversierten Nutzenden die algorithmische Kuration von Inhalten unbeliebt zu sein scheint, gibt es unter Techniklaien und Gelegenheitsnutzenden eine befürwortende Tendenz gegenüber algorithmischer Personalisierung (Sundar, 2020, S. 76).

Laut Eg et al. (2023) besteht die Herausforderung bei der Verbesserung der Nutzerkontrolle darin, abzugrenzen, welche Informationen die Nutzenden zum Ausüben von Kontrolle benötigen. Dieser Mangel an etablierten Standards zur Kontrolle, Konfiguration und Automatisierung führe indes dazu, dass Plattformen den Nutzenden unverständliche Regelwerke zur Verfügung stellen, die die Verantwortung letztendlich vom Betreiber auf Nutzende verlagern. Dass Nutzende sozialer Medien nicht gänzlich wissen, wie Algorithmen funktionieren und unter Umständen zusätzlich nicht motiviert sind, sich darüber zu informieren, kann die Bereitschaft etwaige Konfiguration vorzunehmen senken, weshalb Forschende dafür plädieren, dass Dienste Designlösungen implementieren müssen, die Konsumentinnen und Konsumenten zur Konfiguration animieren (S. 11). AS sind nur ein Faktor der beeinflusst, was Nutzende sehen und diverse Faktoren können nicht kontrolliert werden, darunter etwa, was andere Nutzende teilen (S. 12).

Digitale, personalisierte Nudges könnten ferner eingesetzt werden, um Nutzende zu motivieren, ihre Einstellungen aktiv zu konfigurieren. Es muss jedoch angemerkt werden, dass manipulative Designelemente, auch wenn sie im Interesse der Nutzenden sind, die Wahrnehmung von mangelnder Kontrolle verstärken (Bongard-Blanchy et al., 2021, S. 773).

In Bezug auf RS besteht für VLOP die Pflicht auf die Bereitstellung einer nichtpersonalisierten Variante (Art. 38), was zwar personalisierte Empfehlungen verhindert, aber mangels Personalisierung die Nutzerfahrung verschlechtern kann. Es fehlt in diesem Sinne eine dynamische Einstellbarkeit, die es Nutzenden ermöglicht, in ihrem eigenen Interesse Parameter zu gewichten oder auszuschliessen.



## 4 Methodik

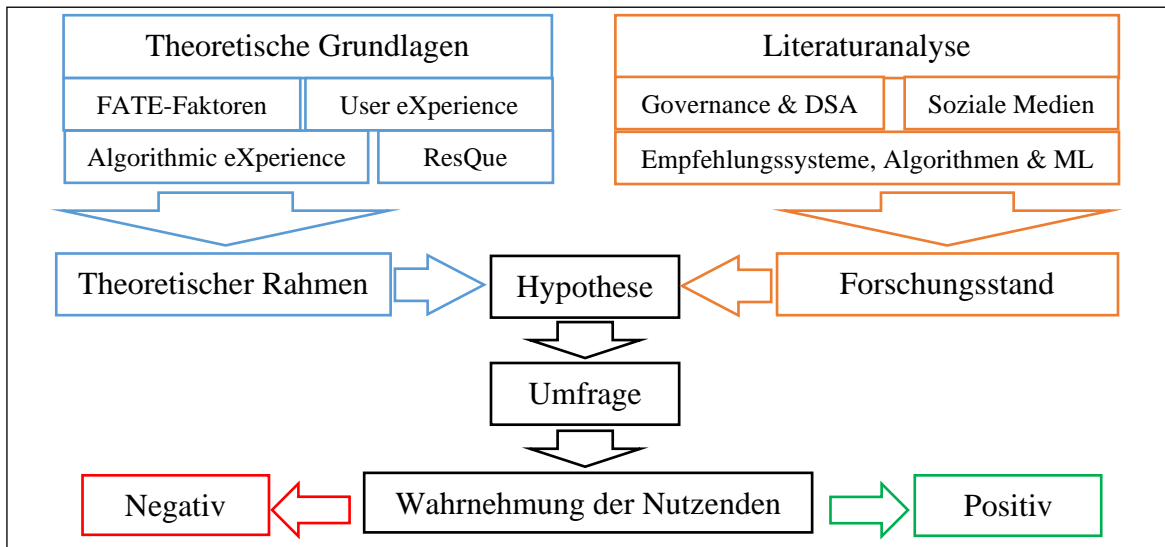


Abbildung 5: Grafische Darstellung des methodischen Vorgehens

Die Methodik der Thesis kann als Mixed-Method-Approach, also als eine Herangehensweise, basierend auf mehreren Methoden, beschrieben werden. Zum einen kann die vertiefte Analyse einer Gesetzesreform als Policy-Analyse bezeichnet werden. Dabei handelt es sich um Case-Studies aus dem Gebiet der Politik- und Rechtswissenschaften. Die Forschungshypothesen werden auf der inhaltlichen Analyse von Literatur gebildet. Die Hypothesen werden mit einer Meinungsbefragung verifiziert oder falsifiziert.

Damit die Umfrage aussagekräftig ist, ist zu berücksichtigen, dass verschiedene Generationen unterschiedliche Sichtweisen auf soziale Medien haben (Prakash Yadav & Rai, 2017, S. 10–11). Unter Berufung auf frühere Forschung werden in Prakash Yadav & Rai (2017) Generationen wie folgt definiert: Geburtsjahrgänge 1961 bis 1980 bilden die Generation X, gefolgt von den Geburtsjahrgängen 1981 bis 1990, die die Generation Y darstellen, sowie der Generation Z ab Geburtsjahrgang 1991. Manchmal wird der Anfang der Generation Z auch mit Geburtsjahrgängen Mitte der 1990er Jahre angegeben (Bongard-Blanchy et al., 2021, S. 766). Zu betonen ist, dass die Einteilung von Menschen in Generationen keiner exakten Wissenschaft entspricht und auf der Annahme basiert, dass bestimmte Altersgruppen universell ähnliche Erfahrungen machen. Der Einfluss anderer Faktoren, die auf das Heranwachsen von Menschen prägende Auswirkungen haben, wie die sozioökonomische Stellung, geographische oder ethnische Herkunft oder die öffentlichen, politischen Verhältnisse im Heimatland, wird mitunter ausgeklammert.

In der Befragung liegt ein Fokus auf jungen Erwachsenen, die ungefähr in die Generation Z fallen und aktuell mit einem Alter von 18 bis 35 Jahren definiert werden. Diese

Altersgruppe ist interessant, da ein reger Gebrauch von Social-Media-Plattformen zu beobachten ist, obwohl eine kritische Grundhaltung gegenüber den Diensten verbreitet ist. Diese Benutzerinnen und Benutzer, die eine gewisse Verdrossenheit oder gar Resignation gegenüber Social-Media-Plattformen empfinden, aber dennoch tägliche Nutzende sind, sind aus regulatorischer Perspektive interessant. Beim Betrachten der Nutzerzahlen wird ersichtlich, dass es für diese Personengruppe keine Option ist, Social-Media-Dienste aufgrund der potenziellen Risiken nicht zu benutzen. Diese Dienste stellen mittlerweile ihre primäre Unterhaltungs- und Informationsquelle dar oder sind ihr bevorzugter Kommunikationskanal. Die Dienste sind in das Leben junger Menschen derart eingewoben, dass der Konsumverzicht keine Option darstellt und sie in ihrem digitalen Alltag auf die Governance der Dienste angewiesen sind. Als geographische Einschränkung gilt die Schweiz, da die Umfrage evaluieren soll, ob die Governance nach DSA ein Modell darstellt, das schweizerische Nutzerinnen und Nutzer positiv bewerten. Da die Umfrage auf Deutsch durchgeführt wird, werden primär Personen aus deutschschweizerischen Regionen befragt.

#### 4.1 Hypothese und Operationalisierung

Da in der Forschungsliteratur alternative Formen von Governance und Verbesserung der algorithmic eXperience (AX) diskutiert werden, geht diese Forschungsarbeit von der These aus, dass der Governance-Ansatz nach DSA nicht den optimalen Lösungsweg darstellt und herauszufordern ist. Die Variablen der Umfrage sind AX- und Governance-Ansätze (H) sowie der Zustimmungswert des Samples ( $\mu$ ).

Nr.	Leithypothese	Test
H <sub>0</sub>	Die Mehrheit von jungen Erwachsenen in der Schweiz befürwortet Governance von Social-Media-Algorithmen nach Digital Services Act.	H <sub>0</sub> : $\mu = \mu_0$
H <sub>A</sub>	Die Minderheit von jungen Erwachsenen in der Schweiz befürwortet Governance von Social-Media-Algorithmen nach Digital Services Act.	H <sub>A</sub> : $\mu = \mu_A$

Da die Leithypothese nicht als Ganzes überprüfbar ist, wurde sie in Teilhypothesen A bis I zerlegt. Die Nullhypothese jeder Teilhypothese repräsentiert den DSA-Governance-Ansatz, während die anderen Hypothesen Alternativen aus wissenschaftlichen Quellen darstellen.

##### 4.1.1 Gerechtigkeit

Gerechtigkeit existiert im DSA in Form der Pflicht, Personen vor Schäden (z. B. Diskriminierung) zu schützen und die Risiken für Nutzende zu evaluieren, zu bewerten und zu



mindern. Die Teilhypothesen A beziehen sich auf die Risikobewertung und Risikominderung durch Plattformbetreiber (Art. 34, Art. 35), denen die Alternative von allgemein gültigen ethischen Richtlinien gegenübergestellt wird.

Nr.	Hypothese	Quelle	Test	
A	H <sub>0A</sub>	Die Befragten befürworten Governance nach Risikobewertung und -minderung.	Art. 34 DSA, Art. 35 DSA	H <sub>0A</sub> : $\mu = \mu_{0A}$
	H <sub>1A</sub>	Die Befragten befürworten Governance nach klaren ethischen Richtlinien.	(Turillazzi et al., 2023, S. 100)	H <sub>1A</sub> : $\mu = \mu_{1A}$

Die Teilhypothesen B beziehen sich auf die Schadenminderungspflicht, die sich gemäss DSA (Art. 35, Art. 35) explizit auf schwere Schäden an Mensch, Gesellschaft und Demokratie beschränkt. Dem ist gegenüberzustellen, dass bereits subtilere Formen von Schäden verhindert werden können. Für den Einbezug von subtileren Formen von Schäden an Mensch und Gesellschaft werden den fünf Schadentypen nach DSA weitere fünf Schadentypen entgegengestellt, die nicht Risikominderungspflichten unterliegen, aber in der Literatur erwähnt werden. Dazu wurden ausgewählt: Fake News, Polarisierung, Radikalisierung, Schlafstörungen, «FOMO», was im weiteren Sinne die Angst, etwas zu verpassen, beschreibt. Diese fünf Faktoren bilden die Hypothese H<sub>1B</sub>.

Nr.	Hypothese	Quelle	Test	
B	H <sub>0B</sub>	Die Befragten befürworten Governance zur Verhinderung «schwerer Schäden».	Art. 34 DSA, Art. 35 DSA	H <sub>0B</sub> : $\mu = \mu_{0B}$
	H <sub>1B</sub>	Die Befragten befürworten Governance zur Verhinderung «aller Schäden».	(Salgado-Criado & Fernandez- Aller, 2021, S. 61)	H <sub>1B</sub> : $\mu = \mu_{1B}$

#### 4.1.2 Haftbarkeit

Fragen rund um die Haftbarkeit und durchsetzende Instanzen werden im DSA ausgiebig definiert. Die Teilhypothesen C beziehen sich auf die Frage nach der haftbaren Instanz im Schadenfall, die im Art. 51 mit der Unternehmung selbst angegeben ist und der andere Varianten aus der Literatur (Programmentwickelnde, Entscheidungstragende) gegenübergestellt werden.

Nr.	Hypothese	Quelle	Test	
C	H <sub>0C</sub>	Die Befragten befürworten Governance mit Haftbarkeit der Unternehmung.	Art. 51 DSA	H <sub>0C</sub> : $\mu = \mu_{0C}$
	H <sub>1C</sub>	Die Befragten befürworten Governance mit Haftbarkeit des Entwicklers bzw. der Entwicklerin.	(Fjeld et al., 2020, S. 35–	H <sub>1C</sub> : $\mu = \mu_{1C}$

H <sub>2C</sub>	Die Befragten befürworten Governance mit Haftbarkeit des Managements.	36)	H <sub>2C</sub> : $\mu = \mu_{2C}$
-----------------	---	-----	------------------------------------

Die Teilhypothesen D beziehen sich auf die Governance mittels regulierter Selbstregulation und unabhängigen Prüfinstitutionen im Vergleich zu Ansätzen der reinen Selbstkontrolle, reinen staatlichen Kontrolle oder demokratischen Governance durch Nutzende.

Nr.	Hypothese	Quelle	Test	
D	H <sub>0D</sub>	Die Befragten befürworten Governance mittels regulierter Selbstregulation und Kontrolle durch unabhängige Prüfgesellschaften.	Art. 37	H <sub>0D</sub> : $\mu = \mu_{0D}$
	H <sub>1D</sub>	Die Befragten befürworten Governance mittels Kontrolle durch staatliche Behörden.	(Steurer, 2013, S. 395–398)	H <sub>1D</sub> : $\mu = \mu_{1D}$
	H <sub>2D</sub>	Die Befragten befürworten Governance mittels unregulierter Selbstkontrolle der Unternehmung.	(Albareda, 2008, S. 433; Gill, 2008, S. 453)	H <sub>2D</sub> : $\mu = \mu_{2D}$
	H <sub>3D</sub>	Die Befragten befürworten demokratische Governance durch Nutzerinnen und Nutzer durch Abstimmungen zu Governance-Themen.	(Engelmann et al., 2020; Fan & Zhang, 2020, S. 1–2)	H <sub>3D</sub> : $\mu = \mu_{3D}$

Die Teilhypothesen E präsentieren die Aufsichtszuständigkeit nach DSA Art. 49 (Nationale Koordinationsstelle) gegenüber einer schweizerischen Koordinationsstelle. Dieser Vorschlag wird in der Theorie nicht erwähnt, sondern wird inkludiert, weil die Umfrageteilnehmenden darüber Auskunft erteilen sollen, ob sie schweizerische Behörden mit dieser Tätigkeit grundsätzlich beauftragen würden.

Nr.	Hypothese	Quelle	Test	
E	H <sub>0E</sub>	Die Befragten befürworten Governance durch Nationale Koordinationsstelle im Land der Dienstleister.	Art. 49	H <sub>0E</sub> : $\mu = \mu_{0E}$
	H <sub>1E</sub>	Die Befragten befürworten Governance durch eine zentrale Stelle der Europäischen Union.		H <sub>1E</sub> : $\mu = \mu_{1E}$
	H <sub>2E</sub>	Die Befragten befürworten Governance durch schweizerische Aufsichtsbehörde.	-	H <sub>2E</sub> : $\mu = \mu_{2E}$

Die Teilhypothesen F beziehen sich auf empirische Ergebnisse, die eklatante Abweichungen zwischen dem Vertrauen, welches Nutzende in verschiedene Institutionen haben, zeigen. Die Frage soll dazu dienen, die Ergebnisse der Teilhypothesen E besser interpretieren zu können.

Nr.	Hypothese	Quelle	Test	
F	H <sub>0F</sub>	Die Befragten vertrauen in die Aufsicht von amerikanischen und chinesischen Diensten durch EU-Behörden.	Art. 49 – Art. 51	H <sub>0F</sub> : $\mu = \mu_{0F}$
	H <sub>1F</sub>	Die Befragten vertrauen in die Aufsicht von chinesischen Diensten (z. B. TikTok) durch chinesische Behörden.	(B. Zhang & Dafoe, 2020, S. 191)	H <sub>1F</sub> : $\mu = \mu_{1F}$
	H <sub>2F</sub>	Die Befragten vertrauen in die Aufsicht von amerikanischen Diensten (z. B. Facebook, Instagram, Twitter) durch amerikanische Behörden.		H <sub>2F</sub> : $\mu = \mu_{2F}$
	H <sub>3F</sub>	Die Befragten vertrauen in die Aufsicht von amerikanischen und chinesischen Diensten durch schweizerische Behörden.		H <sub>3F</sub> : $\mu = \mu_{3F}$

### 4.1.3 Transparenz

Die Transparenz der algorithmischen Systeme wird im DSA mit der Offenlegung der Empfehlungsmechanismen in den AGB reguliert. Die Teilhypothesen G beziehen sich auf die Offenlegung von Empfehlungsmechanismen und Parametern (Art. 27) anstelle einer Verpflichtung, die Mechanismen direkt in der Plattformoberfläche zu beschreiben.

Nr.	Hypothese	Quelle	Test	
G	H <sub>0G</sub>	Die Befragten befürworten die Offenlegung der Empfehlungsmechanismen in den AGB.	Art. 27	H <sub>0G</sub> : $\mu = \mu_{0G}$
	H <sub>1G</sub>	Die Befragten befürworten die Offenlegung der Empfehlungsmechanismen auf der Seite.	(Shin, 2020b, S. 547)	H <sub>1G</sub> : $\mu = \mu_{1G}$

#### 4.1.4 Erklärbarkeit

Im DSA sind Vorgaben zur Erklärbarkeit der algorithmischen Systeme durch die Offenlegung der verwendeten Datenparameter definiert. Die Teilhypothesen H stellen die generalisierte Offenlegung der Empfehlungsparameter in einer globalen Erklärung der gängigen XAI-Empfehlung, die Parameter direkt bei der Empfehlung selbst anzuzeigen, gegenüber.

Nr.	Hypothese	Quelle	Test	
H	H <sub>0H</sub>	Die Befragten befürworten die Offenlegung der Empfehlungsparameter als globale Erklärung des Systems.	Art. 27	H <sub>0H</sub> : $\mu = \mu_{0H}$
	H <sub>1H</sub>	Die Befragten befürworten die Offenlegung der Empfehlungsparameter als lokale Erklärung bei der Empfehlung selbst.	(Di Noia et al., 2022, S. 73; Haque et al., 2023, S. 2–3; Salgado-Criado & Fernandez-Aller, 2021, S. 58; Shin, 2020b, S. 547–548; Shin et al., 2022, S. 4)	H <sub>1H</sub> : $\mu = \mu_{1H}$

#### 4.1.5 Definierbarkeit

Die Definierbarkeit der Empfehlungsmechanismen wird im DSA einzig mit der Pflicht zur Bereitstellung von Profiling-freien Algorithmen adressiert. Die Teilhypothesen I beziehen sich auf diese Pflicht gegenüber vielfach in der Literatur diskutierten Möglichkeiten, dass Nutzende selbst Konfigurationen vornehmen können.

Nr.	Hypothese	Quelle	Test	
I	H <sub>0I</sub>	Die Befragten befürworten die Bereitstellung Profiling-freier Recommender-Systeme.	Art. 38	H <sub>0I</sub> : $\mu = \mu_{0I}$
	H <sub>1I</sub>	Die Befragten befürworten die dynamische Einstellbarkeit von Recommender-Systemen.	(Alvarado & Waern, 2018, S. 6–7; Bucher, 2017, S. 39–41; Shin et al., 2022, S. 2)	H <sub>1I</sub> : $\mu = \mu_{1I}$

## 4.2 Datenerhebung

Basierend auf der Operationalisierung wurde ein Online-Fragebogen konzipiert. Der Aufbau des Fragebogens und die Reihenfolge der Frageblöcke gestaltet sich wie folgt:

- Erhebung der demographischen Daten
- Fragen zum Nutzererlebnis und zur Einstellung zu Governance
- Fragen zu Risiken von Social-Media-Algorithmen (HA und HB)
- Fragen zur Haftbarkeit und Aufsicht (HC, HD, HE und HF)
- Fragen zur Transparenz und Nachvollziehbarkeit (HG, HH und HI)

Die Erhebung wurde auf LimeSurvey erstellt und zwischen dem 26. Mai 2023 und dem 29. Mai 2023 durch fünf Probandinnen und Probanden getestet. Die Rückmeldungen, die sich insbesondere auf das Vokabular bezogen, wurden eingearbeitet. Die Umfrage wurde am 23. Juni 2023 publiziert. Eine vollständige Version des Fragebogens befindet sich im Anhang der Bachelor-Thesis. Die Datenerhebung wurde in zwei Phasen durchgeführt.

Phase	Zeitpunkt	Vorgehen
1	ab 23. Juni 2023 bis 3. Juli 2023	Rekrutierung im erweiterten Umfeld, Arbeitskolleginnen und Arbeitskollegen, Kommilitoninnen und Kommilitonen
2	ab 4. Juli 2023 bis 20. Juli 2023	Rekrutierung ausserhalb des Umfelds durch Flyer und bezahlte Werbekampagnen auf Social Media

Für Werbekampagnen wurden Instagram und LinkedIn ausgewählt. Aus der ersten Phase der Erhebung war bekannt, dass Instagram die meistgenutzte Plattform innerhalb der Zielgruppe ist. LinkedIn wurde mit dem Hintergedanken ausgewählt, dass eine wissenschaftliche Erhebung im Rahmen einer Bachelor-Thesis auf einer Karriereplattform das Interesse wecken könnte. Beide Kampagnen generierten jedoch nur geringe Klickzahlen. Insgesamt betrug das Werbebudget CHF 200.00.

Die Datenerhebung wurde am 20. Juli 2023 abgeschlossen. Insgesamt nahmen 189 Teilnehmerinnen und Teilnehmer an der Umfrage teil, wovon jedoch nur 125 den Fragebogen vollständig ausfüllten. Die unvollständigen Fragebogen wurden meist nach Angabe der demographischen Daten nicht weiter ausgefüllt. Aus diesem Grund wurden alle unvollständigen Fragebogen aus dem Datenset entfernt. Von den vollständigen 125 Fragebogen wurde ein Fragebogen manuell entfernt, weil die Teilnehmerin bei jeder Frage «weder noch» angekreuzt hat.

### 4.3 Struktur der Stichprobe

<b>Stichprobengrösse</b>	124		
<b>Alter</b>	25.1 ( $\pm$ 2.7)		
		<b>Anzahl</b>	<b>%</b>
<b>Geschlecht</b>			
Frauen		70	(56.5%)
Männer		47	(37.9%)
Divers		6	(4.8%)
Keine Angabe		1	(0.8%)
<b>Bildungsgrad</b>			
Obligatorische Schule		4	(3.2%)
Berufsabschluss		22	(17.7%)
Gymnasial-, Berufs-/ Fachmatura		58	(46.8%)
Höhere Fachschule		4	(3.2%)
Bachelorstudium		24	(19.4%)
Masterstudium		10	(8.1%)
Doktorat		0	(0%)
Andere		2	(1.6%)
<b>Region</b>			
Espace Mittelland		33	(26.6%)
Innerschweiz		20	(16.1%)
Zürich		34	(27.4%)
Nordwestschweiz		13	(10.5%)
Ostschweiz		18	(14.5%)
Ausland		4	(3.2%)
Genfersee Region		1	(0.8%)
Tessin		0	(0%)
Keine Angabe		1	(0.8%)
<b>Täglich genutzte Plattformen</b>			
Instagram		98	(79%)
Youtube		56	(45.2%)
TikTok		32	(25.8%)
Twitter		12	(9.7%)
Facebook		6	(4.8%)

Tabelle 2: Struktur der Stichprobe

Die Stichprobe zeigt diverse Auffälligkeiten. Etwa konnten deutlich mehr Frauen als Männer erreicht werden. Der höhere Anteil wird zum einen damit erklärt, dass über den Studiengang «Information Science» an der Fachhochschule Graubünden mehr Studentinnen als Studenten erreicht wurden. Ebenfalls konnten über die Social-Media-Kampagnen fast ausschliesslich Frauen zur Teilnahme motiviert werden. Weiterhin gilt, dass die Stichprobe mit 27.5% Anteil von Personen mit Hochschulstudium einen tieferen Bildungsgrad aufweist als die schweizerische Bevölkerung mit einem Anteil von 42.5% (Bundesamt für Statistik, 2023), was dem Fokus auf junge Personen, die zum Teil noch in Ausbildung sind, geschuldet ist. Die Verteilung nach schweizerischen Regionen kann als ausgewogen betrachtet werden. Der grösste Anteil der Stichprobe nutzt täglich Instagram (79%). Dies ist nicht überraschend, da frühere Studien bereits gezeigt haben, dass Instagram in dieser demographischen Gruppe populär ist, während Facebook tendenziell ältere Nutzende und TikTok eher jüngere anspricht (Taylor & Choi, 2022, S. 9–10). Youtube wird von der Hälfte der Stichprobe regelmässig genutzt, während Twitter in der Schweiz weiterhin nur eine Plattform am äusseren Rande des Marktes ist, was sich in der Stichprobe widerspiegelt. Die Plattformnutzung kann für die Gruppe der 18- bis 35-Jährigen als repräsentativ beschrieben werden, jedoch ist bei der Interpretation der Daten stets zu berücksichtigen, dass die meisten Antworten sich somit implizit auf Instagram von Meta sowie Youtube von Google beziehen.

Statistische Werte		Bestimmung Stichprobengrösse	
Konfidenzniveau (KIV)	90%	KIV <sub>90</sub> =	$\frac{z^2 \times \hat{p} (1 - \hat{p})}{e^2}$
Ablehnungswert (z)	1.65		
Fehlermarge (e)	7.1%		
Populationsproportion ( $\hat{p}$ )	65%	KIV <sub>90</sub> =	$\frac{1.65^2 \times 0.65(1 - 0.65)}{0.071^2} = 123$

Tabelle 3: Statistische Werte

Die Stichprobe kann, ausgehend von einem Konfidenzniveau von 90%, einer Fehlermarge von 7.1% und einer Populationsproportion von 65%, als statistisch relevant betrachtet werden. Die Populationsproportion wurde basierend auf empirischen Erkenntnissen definiert. Unterschiedliche Studien haben die Zustimmung zu Regulation von KI, ML und Social Media erhoben. Insgesamt zeigte sich in allen Studien ein klarer Trend zur Zustimmung, jedoch mit einer Spannweite an Resultaten von 60% bis 90% Zustimmung (Chung & Wihbey, 2022, S. 7–10; Kelley et al., 2021, S. 628; B. Zhang & Dafoe, 2020, S. 187). Basierend auf diesen Studien wurde für die Berechnung der Stichprobengrösse ein tiefer, aber robuster  $\hat{p}$ -Wert von 65% gewählt. Bei der Auswertung der Likert-

Skala wurden die Daten von «Stimme eher zu» und «Stimme zu» aggregiert, wie auch die Daten von «Stimme eher nicht zu» und «Stimme überhaupt nicht zu».



## 5 Ergebnisse

Nachfolgend wird der Zustimmungswert nach Likert-Skala je Umfrageelement angezeigt, jeweils mit den Antwortmöglichkeiten «Stimme eher zu» und «Stimme zu» zu einem Wert aggregiert.

### 5.1 Einstellung zu Aufsicht, Regulation und den algorithmischen Systemen

		Zustimmung	
		Anzahl	%
<b>Social-Media-Plattformen...</b>			
werden im Falle von Schäden genügend zur Rechenschaft gezogen.	(A)	12	(9.7%)
sollen reguliert und beaufsichtigt werden.	(G)	103	(83.7%)
<b>Social-Media-Algorithmen...</b>			
sind nach meinen Bedürfnissen einstellbar.	(D)	39	(31.5%)
sind vertrauenswürdig und agieren in meinem Interesse.	(V)	29	(23.4%)
sind transparent in Bezug auf ihre Funktionsweise.	(T)	10	(8.1%)
behandeln alle Nutzer*innen gleich.	(F)	12	(9.7%)
empfehlen mir Inhalte, die mir entsprechen.	(E)	82	(66.7%)

Tabelle 4: Ergebnisse zur Einstellung zu Aufsicht, Regulation und Algorithmen

Im ersten Abschnitt des Fragebogens wurde die generelle Einstellung gegenüber Governance von Social-Media-Plattformen ermittelt (G) und das Ausmass an Vertrauen (V), dass Nutzende derzeit in die Algorithmen haben. Die überwiegende Mehrheit (83.7%) gab an, dass Social-Media-Plattformen beaufsichtigt und reguliert werden sollten. Ebenfalls wurde ermittelt, dass nur (23.4%) der Nutzenden den algorithmischen Systemen vertrauen.

Daneben wurde evaluiert, ob die FATED-Faktoren derzeit von algorithmischen Systemen erfüllt werden. Fast alle Faktoren erhielten in diesem Zusammenhang tiefe Zustimmungswerte: Gerechtigkeit (9.7%), Haftbarkeit (9.7%), Transparenz (8.1%), Erklärbarkeit (66.7%) und Definierbarkeit (31.5%). Der einzige erhöhte Wert der Erklärbarkeit kann mit der Formulierung der Frage erklärt werden. Im Fragebogen war nicht explizit ausgeführt, dass Nutzende den Empfehlungsvorgang verstehen müssen, sondern lediglich, dass die Empfehlungen zum eigenen Nutzungsverhalten passen und somit nachvollziehbar sind.

Der Wert kann auch als eine generelle Zufriedenheit mit der Empfehlungsqualität interpretiert werden.

## 5.2 Einstellung zu Risiken auf sozialen Medien

	Zustimmung	
	Anzahl	%
<b>H<sub>A</sub>: Social-Media-Plattformen sollten Massnahmen zum Schutz von Nutzer*innen auf Grundlage von...</b>		
klaren ethischen Vorgaben im Gesetz treffen.	105	(84.7%)
Analyse und Bewertung möglicher Risiken durch die Plattformbetreiber selbst treffen.	56	(45.2%)
<b>H<sub>B</sub>: Social-Media-Plattformen sollten verpflichtet sein, ...</b>		
die Entstehung schwerer Schäden an Menschen (Depression, Trauma) durch ihre Dienste zu verhindern.	97	(78.2%)
die Zunahme von geschlechterspezifischer Gewalt zu verhindern.	115	(92.7%)
die Verbreitung illegaler Inhalte über ihre Dienste zu verhindern.	108	(87.1%)
Risiken für gesellschaftliche Debatten, Wahlprozesse und die öffentliche Sicherheit zu verhindern.	73	(58.9%)
persönliche Diskriminierungen und diskriminierende Inhalte zu verhindern.	112	(90.3%)
<b>schwere Schäden zu mindern.</b>		<b>81.5%</b>
Verbreitung von Desinformation («Fake News») zu verhindern.	111	(89.5%)
gesamtesellschaftliche Polarisierung (Verstärkung von Meinungsunterschieden) zu verhindern.	59	(47.6%)
politische Radikalisierung von Einzelnen oder Personengruppen zu verhindern.	89	(71.8%)
übermässigen Online-Konsum (Internetsucht) zu verhindern.	72	(58.1%)
moderate nachteilige Auswirkungen (Schlafstörungen, «FOMO») zu verhindern.	49	(39.5%)
<b>alle Schäden zu mindern.</b>		<b>61.3%</b>

Tabelle 5: Ergebnisse zur Einstellung zu Risiken auf sozialen Medien

Im zweiten Abschnitt der Umfrage wurde Governance nach Risikobewertung und Risikominderung (Art. 34, Art. 35) der Governance nach klaren ethischen Richtlinien gegenübergestellt. Wobei die zweite Variante deutlich höher bewertet wurde (84.7%)

gegenüber der ersten (45.2%). Anschliessend wurden fünf potenzielle Schäden von Social-Media-Algorithmen aus der Literatur fünf Schäden gegenübergestellt, die gemäss DSA zwingend zu mindern sind. Die Ergebnisse zeigten eine teilweise Übereinstimmung mit dem Gesetz, wobei die im DSA verankerte Schadenminderungspflicht, von Risiken für gesellschaftliche Debatten, Wahlprozesse und die öffentliche Sicherheit, vergleichsweise tief bewertet wurde (58.9%). Zwei der theoretischen, nicht im DSA enthaltenen Minderungspflichten, erhielten hohe Zustimmungswerte, darunter die Verhinderung von politischer Radikalisierung von Einzelnen oder Personengruppen (71.8%) sowie die Verhinderung von Verbreitung von Desinformation (89.5%).

### 5.3 Einstellung zu Haftbarkeit und Aufsichtsinstanz

	Zustimmung	
	Anzahl	%
<b>H<sub>C</sub>: Social-Media-Plattformen sollten im Schadenfall...</b>		
in Form der verantwortlichen Programm-Entwickler*innen strafrechtlich haftbar sein.	34	(27.4%)
in Form der Personen im Management strafrechtlich haftbar sein.	91	(73.4%)
als Unternehmung wirtschaftsrechtlich haftbar sein.	111	(89.5%)
<b>H<sub>D</sub>: Social-Media-Plattformen sollten ...</b>		
durch Plattform-Nutzer*innen demokratisch beaufsichtigt werden.	61	(49.2%)
durch staatliche Behörden beaufsichtigt werden.	70	(56.5%)
durch die Plattformbetreiber selbst anhand gesetzlicher Regeln beaufsichtigt werden.	77	(62.1%)
durch Plattformbetreiber auf freiwilliger Basis selbst beaufsichtigt werden.	31	(25%)
durch unabhängige, nichtstaatliche Prüforganisationen beaufsichtigt werden.	96	(77.4%)

Tabelle 6: Ergebnisse zur Einstellung zu Haftbarkeit und Aufsichtsinstanz (1 von 2)

Im dritten Abschnitt der Umfrage wurde ermittelt, welche Instanz in Fällen von Schäden haftbar gemacht werden sollte. Der DSA-Ansatz der wirtschaftsrechtlichen Haftung der Unternehmung wurde am höchsten bewertet (89.5%), gefolgt von strafrechtlicher Haftung für Personen im Management (73.4%) und vergleichsweise geringer Zustimmung für Haftbarkeit der Entwicklerinnen und Entwickler (27.4%).

Auch die Frage nach der Instanz, die mit der Aufsicht beauftragt werden sollte, wurde von den Umfrageteilnehmenden DSA-konform beantwortet. Die höchsten Zustimmungswerte erhielt die Aufsicht durch unabhängige, nichtstaatliche Prüforganisationen (77.4%), durch die Plattformbetreiber selbst anhand gesetzlicher Regeln (62.1%) sowie durch staatliche Behörden (56.5%). Weniger populär waren hingegen die Ansätze aus der wissenschaftlichen Literatur, etwa die demokratische Aufsicht durch Plattform-Nutzende (49.2%) oder der Vorschlag, dass sich Plattformbetreiber auf freiwilliger Basis selbst beaufsichtigen sollen (25%).

	Zustimmung	
	Anzahl	%
<b>H<sub>E</sub>: Social-Media-Plattformen sollten beaufsichtigt werden durch...</b>		
eine spezialisierte Behörde im Domizilland (Unternehmenssitz) des Social-Media-Dienstes.	46	(37.1%)
eine spezialisierte EU-Behörde im europäischen Binnenmarkt (inkl. Schweiz, Norwegen, Island, Liechtenstein)	93	(75%)
eine schweizerische Behörde, sofern Dienste hierzulande verfügbar sind.	86	(69.4%)
<b>H<sub>F</sub>: Ich vertraue der Aufsicht von...</b>		
chinesischen Diensten (z.B. TikTok) durch chinesische Behörden.	4	(3.2%)
amerikanischen Diensten (z. B. Facebook, Instagram, Twitter) durch amerikanische Behörden.	10	(8.1%)
amerikanischen und chinesischen Diensten durch EU-Behörden.	74	(59.7%)
amerikanischen und chinesischen Diensten durch schweizerische Behörden.	83	(66.9%)

Tabelle 7: Ergebnisse zur Einstellung zu Haftbarkeit und Aufsichtsinstanz (2 von 2)

Bei der Frage danach, in welchem Staat die Aufsichtstätigkeit angesiedelt werden sollte, schnitt die EU-Behörde, wie im DSA vorgesehen (75%) etwas besser als das hypothetische Szenario einer schweizerischen Behörde ab (69.4%). Das Szenario einer Aufsichtsbehörde im Domizilland des Social-Media-Dienstes schnitt am unzureichendsten ab (37.1%) was im Kontext der Resultate der darauffolgenden Frage interpretiert werden. Sie eruierte, welchen Staaten die Nutzenden grundsätzlich bei der Aufsicht von Social-Media-Plattformen vertrauen: Hohe Zustimmung erfuhr die Aufsicht von amerikanischen und chinesischen Diensten durch schweizerische Behörden (66.9%) und EU-Behörden (59.7%), wobei tiefe Zustimmung bei der Aufsicht von chinesischen Diensten (z. B. TikTok) durch chinesische Behörden (3.2%) und der Aufsicht von amerikanischen

Diensten (z. B. Facebook, Instagram oder Twitter) durch amerikanische Behörden (8.1%) messbar war.

## 5.4 Einstellung zu Transparenzpflichten von algorithmischen Systemen

	Zustimmung	
	Anzahl	%
<b>H<sub>G</sub>: Social-Media-Plattformen sollten die Funktionsweise der Social-Media-Algorithmen...</b>		
in den AGB offenlegen.	99	(80.5%)
in der App oder auf der Webseite offenlegen.	114	(91.9%)
<b>H<sub>H</sub>: Social-Media-Plattformen sollten die für Empfehlungen verwendeten Personendaten ...</b>		
in den AGB generalisiert beschreiben.	89	(72.4%)
bei jeder Empfehlung (vorgeschlagener Post, Tweet, Video) personalisiert anzeigen.	80	(64.5%)
<b>H<sub>I</sub>: Social-Media-Plattformen sollten mir die Möglichkeit bieten, ...</b>		
selbst zu entscheiden, welche meiner Personendaten für Empfehlungen verwendet werden.	111	(89.5%)
Empfehlungen ohne Verwendung von Personendaten über mich zu generieren.	96	(77.4%)

Tabelle 8: Ergebnisse zur Einstellung zu Transparenzpflichten

Im letzten Abschnitt der Umfrage wurde die Einstellung zu Transparenzpflichten ermittelt. Die im DSA definierte Offenlegung der Funktionsweise von Algorithmen in den AGB erhielt leicht geringere Zustimmungswerte (80.5%) als die Offenlegung direkt über das Interface (91.9%). Die Beschreibung der verwendeten Personendaten in den AGB erhielt leicht mehr Zustimmung (72.4%) als die XAI-Variante der personalisierten Beschreibung bei jeder Empfehlung oder bei jedem vorgeschlagenen Beitrag (64.5%). Bezüglich der Auswahl verwendeter Personendaten bei Empfehlungen wurden beide Ansätze positiv bewertet. Empfehlungen ohne Verwendung von Personendaten nach DSA erhielten eine leicht tiefere Zustimmung (77.4%) als die Variante, bei der Nutzende selbst entscheiden, welche Personendaten für Empfehlungen verwendet werden (89.5%).

## 5.5 Hypothesentest

In der nachfolgenden Tabelle repräsentieren Nullhypothesen den Ansatz nach DSA während die Alternativhypothesen andere Governance-Ansätze darstellen.

<b>F - Gerechtigkeit</b>	
H <sub>A</sub>	Nullhypothese H <sub>0A</sub> konnte nicht bestätigt werden. Es ist von Alternativhypothese H <sub>1A</sub> auszugehen, die bestätigt wurde.
H <sub>B</sub>	Nullhypothese H <sub>0B</sub> konnte bestätigt werden. Alternativhypothese H <sub>1B</sub> konnte ebenfalls bestätigt werden.
<b>A - Haftbarkeit</b>	
H <sub>C</sub>	Nullhypothese H <sub>0C</sub> konnte bestätigt werden. Alternativhypothese H <sub>1C</sub> konnte nicht bestätigt werden. Alternativhypothese H <sub>2C</sub> konnte bestätigt werden.
H <sub>D</sub>	Nullhypothese H <sub>0D</sub> konnte bestätigt werden. Alternativhypothesen H <sub>1D</sub> , H <sub>2D</sub> und H <sub>3D</sub> konnten nicht bestätigt werden.
H <sub>E</sub>	Nullhypothese H <sub>0E</sub> konnte nicht bestätigt werden. Alternativhypothesen H <sub>1E</sub> und H <sub>2E</sub> konnten beide bestätigt werden.
H <sub>F</sub>	Nullhypothese H <sub>0F</sub> konnte bestätigt werden. Alternativhypothesen H <sub>1F</sub> und H <sub>2F</sub> konnten nicht bestätigt werden. Alternativhypothese H <sub>3F</sub> konnte ebenfalls bestätigt werden.
<b>T - Transparenz</b>	
H <sub>G</sub>	Nullhypothese H <sub>0G</sub> konnte bestätigt werden. Alternativhypothese H <sub>1G</sub> konnte ebenfalls bestätigt werden.
<b>E - Erklärbarkeit</b>	
H <sub>H</sub>	Nullhypothese H <sub>0H</sub> konnte bestätigt werden. Alternativhypothese H <sub>1H</sub> konnte ebenfalls bestätigt werden.
<b>D - Definierbarkeit</b>	
H <sub>I</sub>	Nullhypothese H <sub>0I</sub> konnte bestätigt werden. Alternativhypothese H <sub>1I</sub> konnte ebenfalls bestätigt werden.

Tabelle 9: Tabelle der Hypothesentests

Die Hypothesentests der Teilhypothesen zeigten im Allgemeinen einen hohen Zustimmungswert zu den Governance-Ansätzen nach DSA. In fast allen Dimensionen nach FA-TED-Modell wurden DSA-Ansätze mehrheitlich positiv bewertet:

- In der Dimension Gerechtigkeit, bei der Teilhypothese A, wurde Governance nach Risikobewertung und Risikominderung gemäss DSA deutlich tiefer bewertet als die Governance nach klaren ethischen Richtlinien. Jedoch erhielt die Verhinderung schwerer Schäden nach DSA deutlich mehr Zustimmung als die Verhinderung weniger gravierender Schäden.

- In der Dimension Haftbarkeit wurde die wirtschaftsrechtliche Haftbarkeit der Unternehmung (C) sowie Governance mittels regulierter Selbstregulation in Kombination mit Kontrolle durch unabhängige Prüfgesellschaften am positivsten bewertet (D). Die Koordinationsstelle im Land der Dienstleister wurde schlecht bewertet (E), aber das Vertrauen in die Aufsicht von amerikanischen und chinesischen Diensten durch schweizerische oder EU-Behörden wurde als hoch eingestuft (F).
- In der Dimension Transparenz wurde die Offenlegung der Empfehlungsmechanismen in den AGB nach DSA schlechter bewertet als die Offenlegung der Empfehlungsmechanismen auf der Seite (G).
- In der Dimension Erklärbarkeit wurde die globale Erklärung des Systems nach DSA besser bewertet als lokale Erklärungen nach XAI-Literatur (H).
- In der Dimension Definierbarkeit wurde die dynamische Einstellbarkeit von RS besser bewertet als der DSA-Ansatz der Bereitstellung Profiling-freier RS. Jedoch hatten beide Varianten hohe Zustimmungswerte (I).

Aufgrund der hohen Zustimmungswerte zu den DSA-Ansätzen, die sich bei fast allen Fragen zeigten, konnte die Nullhypothese  $H_0$ , dass junge Erwachsene in der Deutschschweiz die Governance-Ansätze zu Social-Media-Algorithmen im Digital Services Act mehrheitlich ablehnen, nicht belegt werden. Eine Ablehnung der Governance-Ansätze nach DSA konnte von den wenigen erwähnten Punkten abgesehen nicht erkannt werden. In einigen Punkten wurden trotz hoher Zustimmung zu DSA-Ansätzen die alternativen Governance-Ansätze besser bewertet.

Auf den Folgeseiten finden sich zwei Gesamttabellen mit den Ergebnissen. Die erste Tabelle zeigt einen Abriss der Resultate mit Hypothesentests während die zweite Tabelle die Gesamtergebnisse beinhaltet.

Die Befragten...	Zustimmung	Hypothese
H <sub>0A</sub> befürworten Governance nach Risikobewertung und -minderung.	45.2% (38.1% - 52.3%)	Kein eindeutiges Ergebnis
H <sub>1A</sub> befürworten Governance nach klaren ethischen Richtlinien.	84.7% (77.6% - 91.8%)	Hypothese bestätigt
H <sub>0B</sub> befürworten Governance zur Verhinderung «schwerer Schäden».	81.45% (74.4% - 88.6%)	Hypothese bestätigt
H <sub>1B</sub> befürworten Governance zur Verhinderung «aller Schäden».	61.29% (54.2% - 68.4%)	Hypothese bestätigt
H <sub>0C</sub> befürworten Governance mit Haftbarkeit der Unternehmung.	89.5% (82.4% - 96.6%)	Hypothese bestätigt
H <sub>1C</sub> befürworten Governance mit Haftbarkeit des Entwicklers bzw. der Entwicklerin.	27.4% (20.3% - 34.5%)	Hypothese nicht bestätigt
H <sub>2C</sub> befürworten Governance mit Haftbarkeit des Managements.	73.4% (66.3% - 80.5%)	Hypothese bestätigt
H <sub>0D</sub> befürworten Governance mittels regulierter Selbstregulation mit Kontrolle unabhängige Prüfungsgesellschaften.	69.8% (62.7% - 76.9%)	Hypothese bestätigt
H <sub>1D</sub> befürworten Governance mittels Kontrolle durch staatliche Behörde.	56.5% (49.4% - 63.6%)	Kein eindeutiges Ergebnis
H <sub>2D</sub> befürworten Governance mittels unregulierter Selbstkontrolle der Unternehmung.	25% (17.9% - 32.1%)	Hypothese nicht bestätigt
H <sub>3D</sub> befürworten demokratische Governance durch Nutzerinnen und Nutzer selbst.	49.2% (42.1% - 56.3%)	Kein eindeutiges Ergebnis
H <sub>0E</sub> befürworten Governance durch Nationale Koordinationsstelle im Land der Dienstleister.	37.1% (30% - 44.2%)	Hypothese nicht bestätigt
H <sub>1E</sub> befürworten Governance durch eine zentrale Stelle der Europäischen Union.	75% (67.9% - 82.1%)	Hypothese bestätigt
H <sub>2E</sub> befürworten Governance durch schweizerische Aufsichtsbehörde.	69.4% (62.3% - 76.5%)	Hypothese bestätigt
H <sub>0F</sub> vertrauen in die Aufsicht von amerikanischen und chinesischen Diensten durch EU-Behörden.	59.7% (52.6% - 66.8%)	Hypothese bestätigt
H <sub>1F</sub> vertrauen in die Aufsicht von chinesischen Diensten (z. B. TikTok) durch chinesische Behörden	3.2% (0% - 10.3%)	Hypothese nicht bestätigt
H <sub>2F</sub> vertrauen in die Aufsicht von amerikanischen Diensten durch amerikanische Behörden.	8.1% (1% - 15.2%)	Hypothese nicht bestätigt
H <sub>3F</sub> vertrauen in die Aufsicht von amerikanischen und chinesischen Diensten durch schweizerische Behörden.	66.9% (59.8% - 74%)	Hypothese bestätigt
H <sub>0G</sub> befürworten die Offenlegung der Empfehlungsmechanismen in den AGB.	80.5% (73.4% - 87.6%)	Hypothese bestätigt
H <sub>1G</sub> befürworten die Offenlegung der Empfehlungsmechanismen auf der Seite.	91.9% (84.8% - 99%)	Hypothese bestätigt
H <sub>0H</sub> befürworten die Offenlegung der Empfehlungsparameter als globale Erklärung des Systems.	72.4% (65.3% - 79.5%)	Hypothese bestätigt
H <sub>1H</sub> befürworten die Offenlegung der Empfehlungsparameter als lokale Erklärung bei der Empfehlung selbst.	64.5% (57.4% - 71.6%)	Hypothese bestätigt
H <sub>0I</sub> befürworten die Bereitstellung Profiling-freier Recommender-Systeme.	77.4% (70.3% - 84.5%)	Hypothese bestätigt
H <sub>1I</sub> befürworten die dynamischer Einstellbarkeit von Recommender-Systemen.	89.5% (82.4% - 96.6%)	Hypothese bestätigt

Tabelle 10: Ergebnisse nach Hypothesen (ohne Grafik)



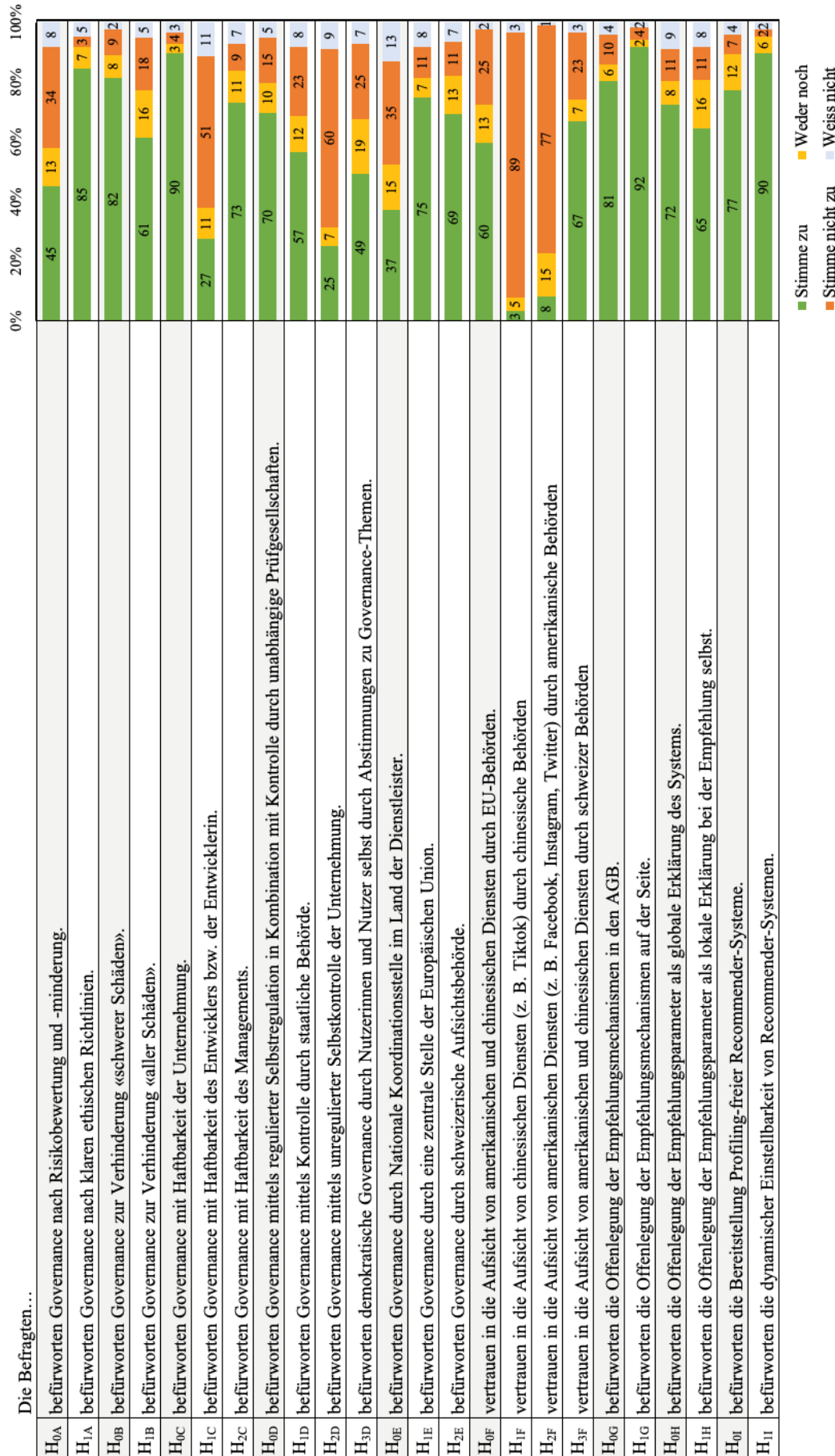


Tabelle 11: Ergebnisse nach Hypothesen (mit Grafik)

## 5.6 Teilmengenanalyse

### 5.6.1 Einstellung zu Governance

Für alle Teilnehmenden wurde ein Governance-Zustimmungswert berechnet. Der Wert gibt in Prozent an, wie hoch der Anteil der Governance-Fragen ist, die mit «Stimme zu» oder «Stimme völlig zu» beantwortet wurden und veranschaulicht, inwiefern Governance-Massnahmen DSA-unabhängig befürwortet werden. Anhand der Teilmengenanalyse lässt sich zudem eruieren, ob es durch Über- oder Unterrepräsentation gewisser Personengruppen zu einer möglichen Verzerrung gekommen ist.

Jedoch sind keine erheblichen Unterschiede zwischen den Teilmengen feststellbar. Der Zustimmungswert über die gesamte Stichprobe ergab 61%. Da Instagram in der Stichprobe die meistgenutzte Plattform ist, hat sich eine Auswertung nach Plattform erübrigt.

Geschlecht	Frauen	Männer	Divers
Anzahl	70	47	6
Zustimmung	61%	61%	63%

Bildungs-grad	Oblig. Schule	Berufs-ausbildung	Gymnasial-, Berufs- oder Fachmatura	Höhere Fachschule	Bachelor-Abschluss	Master-Abschluss
Anzahl	4	22	58	4	24	10
Zustimmung	58%	60%	60%	66%	62%	60%

Alter	18	19	20	21	22	23	24	25	26
Anzahl	3	2	6	20	12	8	9	6	15
Zustimmung	50%	64%	55%	57%	49%	54%	50%	57%	51%
Alter	27	28	29	30	31	32	33	34	35
Anzahl	9	7	11	8	2	0	1	0	5
Zustimmung	57%	61%	58%	51%	58%	-	55%	-	52%

Region	Espace Mittel-land	Inner-schweiz	Zürich	Nord-west-schweiz	Ost-schweiz	Genferse e-region	Ausland
Anzahl	33	21	34	13	18	1	4
Zustimmung	58%	52%	51%	56%	54%	65%	61%

Tabelle 12: Ergebnisse der Teilmengenanalyse zu allen Governance-Ansätzen

Es gab keine nennenswerten Unterschiede zwischen Angehörigen unterschiedlicher Geschlechter oder Bildungsschichten. Einzig die Inhaberinnen und Inhaber von HF-Abschlüssen hatten einen leicht erhöhten Wert, jedoch basiert der Wert auf nur vier Personen. Bei der Betrachtung nach Alter gab es neben zwei leicht erhöhten Werten bei 19- und 28-Jährigen und einem leicht tieferen Wert bei 22-Jährigen keine Auffälligkeiten. Eine Korrelation oder gar kausaler Zusammenhang ist wegen der geringen Teilmengen und der unregelmässigen Verteilung der Anomalien fraglich. Zwischen den Regionen zeigte sich ebenfalls ein gleichmässiges Bild. Der einzige erhöhte Wert repräsentiert die Genfersee-Region, die mit einer Person eine zu geringe Teilmenge darstellt, um Aussagen zu treffen.

### 5.6.2 Einstellung zum Digital Services Act

Es wurde zudem eine gesonderte Analyse der Teilmengen durchgeführt, die sich auf die Zustimmungswerten explizit zu DSA-Ansätzen bezog und eine Zustimmung von 62% ergab.

Geschlecht	Frauen		Männer		Divers	
Anzahl	70		47		6	
Zustimmung	60%		64%		67%	

Bildungs-grad	Oblig. Schule	Berufs-ausbildung	Gymnasial-, Berufs- oder Fachmatura	Höhere Fachschule	Bachelor-Abschluss	Master-Abschluss
Anzahl	4	22	58	4	24	10
Zustimmung	56%	59%	62%	66%	66%	59%

Alter	18	19	20	21	22	23	24	25	26
Anzahl	3	2	6	20	11	8	9	6	15
Zustimmung	42%	69%	65%	66%	56%	63%	56%	65%	57%

Alter	27	28	29	30	31	32	33	34	35
Anzahl	9	7	11	8	2	0	1	0	5
Zustimmung	68%	73%	65%	64%	63%	-	50%	-	55%

Region	Espace Mittel-land	Inner-schweiz	Zürich	Nord-west-schweiz	Ost-schweiz	Genfersee -region	Ausland
Anzahl	33	20	34	13	18	1	4
Zustimmung	69%	58%	56%	64%	60%	75%	66%

Tabelle 13: Ergebnisse der Teilmengenanalyse zu DSA-Ansätzen

Die Werte sind erhöht im Vergleich zur generellen Governance-Zustimmung, was sich damit erklären lässt, dass die DSA-Ansätze besser bewertet wurden. Die Schwankungen zwischen den Teilgruppen bewegen sich ebenfalls in einem geringen Bereich.

Es gab bei der Auswertung der DSA-Ansätze keine auffälligen Unterschiede zwischen Angehörigen unterschiedlicher Geschlechter. Die nichtbinären Teilnehmenden zeigen einen leicht erhöhten Wert, der jedoch auf nur sechs Teilnehmende zurückzuführen ist und somit nicht als belastbar betrachtet wird. Es gab keine augenscheinlichen Unterschiede zwischen Angehörigen unterschiedlicher Bildungsschichten. Leicht erhöhte Werte zeigen HF- und Bachelor-Absolventinnen und -Absolventen. Da sich der Trend jedoch bei den

Master-Absolventinnen und -Absolventen nicht fortsetzt, ist eine Korrelation nicht nahelegend. Es zeigen sich diverse Altersgruppen mit eher tiefen Werten, darunter 18-, 22-, 26-, 33- und 35-Jährige, während 19-, 21-, 27- und 28-Jährige erhöhte Werte haben. Da kein gradueller Anstieg ersichtlich ist, wird dies keine Korrelation anzeigen. Zwischen den Regionen zeigte sich ebenfalls ein gleichmässiges Bild. Der einzige erhöhte Wert repräsentiert die Genfersee-Region, die mit nur einer Person eine zu geringe Teilmenge darstellt, um Aussagen zu treffen.

Insgesamt scheinen Faktoren wie Geschlecht, Bildungsgrad, Wohnregion und Alter keinen Einfluss auf die Zustimmung zu Governance-Massnahmen sowie auf die Zustimmung zu DSA-Massnahmen zu haben, was durchaus als überraschendes Ergebnis der Umfrage betrachtet werden kann. Um eine zuverlässigere Aussage zu treffen, müsste jedoch eine höhere Stichprobenmenge analysiert werden.



## 6 Diskussion

### 6.1 Interpretation der Ergebnisse

Die Ergebnisse zeigen eine generelle Befürwortung der DSA-Ansätze. Ebenfalls kann daraus abgeleitet werden, dass mittels Umsetzung des DSA in allen Dimensionen nach AX-Modell eine Verbesserung bewirkt werden kann.

Bei der Teilhypothese A wurde Governance nach Risikobewertung und -minderung gemäss DSA deutlich kritischer bewertet als die Governance nach klaren ethischen Richtlinien. Daraus lässt sich ableiten, dass Nutzerinnen und Nutzer sich klar definierte Grenzen für akzeptable und inakzeptable Schadenspotenziale wünschen.

Bei der Teilhypothese B wurden sowohl Nullhypothese wie auch Alternativhypothese angenommen, jedoch mit markanten Zustimmungsunterschieden. Governance zur Verhinderung schwerer Schäden wurde zu 81.5% befürwortet, während Governance zur Verhinderung aller Schäden, inklusive subtiler Schäden, zu 61.3% befürwortet wurde. Der einzige Punkt innerhalb der Pflicht zur Verhinderung subtiler Schäden, der einen deutlich erhöhten Zustimmungswert genießt, ist die Verhinderung von «Fake News» (89.5%), was einen Zusammenhang zur Corona-Pandemie oder zur Kriegsdesinformation zur russischen Invasion der Ukraine im Februar 2022 haben könnte.

Bei der Teilhypothese C zeigte sich die höchste Zustimmung zur Haftbarkeit der Unternehmung, wie im DSA vorgesehen (89.5%), gefolgt von der Haftbarkeit der Management-Mitglieder (73.4%). Eine Haftbarkeit der Entwickelnden verlangten nur rund ein Viertel der Befragten (27.4%), was zeigt, dass die primäre Verantwortung der Systeme nicht den Entwickelnden zugeschrieben wird. Die hohe Zustimmung für die Haftbarkeit der Management-Mitglieder lässt vermuten, dass der DSA eine Chance verpasst hat, Entscheidungsträgerinnen und -träger in die Verantwortung zu nehmen. Der Zustimmungswert könnte damit in Zusammenhang stehen, dass Geschäftsführer amerikanischer Social-Media-Firmen wie Mark Zuckerberg (Meta) und Elon Musk (Twitter) medial exponiert sind und deshalb in den Augen der Öffentlichkeit persönlich zur Rechenschaft gezogen werden sollten.

Bei der Teilhypothese D befürworteten die Governance durch nationale Koordinationsstelle im Land der Dienstleister 37.1%, was eine teilweise Ablehnung des DSA-Ansatzes nationaler Koordinationsstellen vermuten lässt. Jedoch befürworteten 75% die Governance durch eine zentrale EU-Stelle, die im DSA mindestens teilweise vorgesehen ist. Governance durch eine schweizerische Aufsichtsbehörde wurde mit 69.4% positiv bewertet,

aber tiefer als Governance durch eine EU-Behörde. Warum ausgerechnet schweizerische Konsumentinnen und Konsumenten einer EU-Behörde mehr Vertrauen schenken als einer schweizerischen Behörde, geht aus den Resultaten nicht hervor und ist sicherlich einer vertieften Analyse würdig.

Bei der Teilhypothese E wurde die Koordinationsstelle im Land der Dienstleister schlecht bewertet. Die Aufgabenteilung zwischen der zentralen Stelle in der Europäischen Union, die nach aktuellem Wissensstand eine nachrangige Aufsichtstätigkeit ausführen wird und der Koordinationsstelle im Land, ist für Konsumentinnen und Konsumenten vermutlich nicht nachvollziehbar. Ebenfalls konnten in der Umfrage nicht alle Details des Aufsichtskonzepts aufgearbeitet werden, wie etwa, dass aussereuropäische Social-Media-Plattformen zwingend eine Rechtsvertretung im EU-Raum anmelden müssen. Da dieser Punkt nicht spezifiziert war, ist es möglich, dass die Befragten von Aufsicht durch US-Behörden im Falle von Facebook, Twitter und Instagram, oder von Aufsicht durch chinesische Behörden im Falle von TikTok ausgingen. Die Aufsicht durch die EU-Behörde wurde positiv bewertet, was ein generelles Vertrauen in Aufsichtstätigkeiten durch europäische Institutionen zu zeigen scheint.

Bei der Teilhypothese F erfährt die Aufsicht durch schweizerische Behörden leicht mehr Zustimmung (66.9%) als durch EU-Behörden (59.7%). Dies ist nicht als Kritik am DSA-Ansatz zu verstehen, sondern zeigt, dass schweizerische Konsumentinnen und Konsumenten sich wünschen, dass schweizerische Behörden in entsprechenden Aufsichtstätigkeiten involviert würden. Da die Werte jedoch nah beieinander liegen, ist ein abschließendes Fazit nicht möglich. Auffällig gering bewertet wurde die Aufsicht von chinesischen Diensten (z. B. TikTok) durch chinesische Behörden (3.2%) sowie die Aufsicht von amerikanischen Diensten (z. B. Facebook, Instagram, Twitter) durch amerikanische Behörden (8.1%). Dies könnte im Falle von China daran liegen, dass TikTok eine gewisse Staatsnähe zugeschrieben wird und der chinesische Staat entsprechend kein Interesse daran hat, Konsumentinnen und Konsumenten zu schützen, da er die Plattformen nutzt, um Informationen zu gewinnen. Im Falle der USA könnte der Wert an früheren Skandalen liegen, bei denen US-amerikanische Social-Media-Plattformen involviert waren, die bei Nutzerinnen und Nutzern womöglich das Gefühl von Untätigkeit seitens der US-Regierung erzeugt haben.

Bei der Teilhypothese G wurde sowohl DSA-Variante der Offenlegung der Empfehlungsmechanismen in den AGB (80.5%) wie auch auf der Seite (91.9%) deutlich angenommen. Interessant an der nicht geringfügigen Abweichung ist, dass die AGB von Online-Diensten immer auf der Seite der Betreiber aufzufinden sind. Das bedeutet, die



Konsumentinnen und Konsumenten wissen nicht, wo die AGB zu finden sind oder sie wünschen sich einen noch direkteren Zugriff als bis anhin.

Bei der Teilhypothese H befürworteten 72.4% die Offenlegung der Empfehlungsparameter als globale Erklärung des Systems nach DSA und 64.5% als lokale Erklärung bei der Empfehlung selbst. Beide Hypothesen konnten bestätigt werden und die Werte bewegen sich unweit voneinander. Dass der DSA-Ansatz leicht besser abschneidet als die «Best Practice» nach XAI ist überraschend, könnte aber daran liegen, dass sich Konsumentinnen und Konsumenten noch nicht vorstellen können, wie eine Parameteranzeige aussehen könnte oder befürchten, dass dadurch das Interface überladen wird. Wie eXplainableAI für Empfehlungssysteme oder im Social-Media-Kontext am besten integriert wird, damit Nutzende den Mehrwert erkennen, sollte deshalb in Zukunft vertieft analysiert werden.

Bei der Teilhypothese I wurde die dynamische Einstellbarkeit von Recommender-Systemen besser bewertet als der Profiling-freie RS nach DSA. Bei der Umsetzung einer schweizerischen Regulation sollte überprüft werden, ob Social-Media-Plattformen dynamisch einstellbare RS anbieten müssen. Zudem sollte vertieft evaluiert werden, welche Anforderungen Nutzende an die Parametrisierung ihrer Empfehlungssysteme stellen, sowie welches Ausmass an Daten- und Technikkompetenz nötig ist, damit sie die Parametrisierung selbstständig vornehmen können.

## 6.2 Kontextuelle Limitationen

Diese Bachelor-Thesis, die Fragen im Spannungsfeld zwischen aktuellem Policy-Making, sozialen Medien und maschinellem Lernen erforscht, war von Beginn bis Abschluss in ein hochdynamisches Diskursumfeld eingebettet. Die Vorrecherche zum Thema fand mehrheitlich im Oktober, November und Dezember des Jahres 2022 statt. Zu diesem Zeitpunkt war die öffentliche Debatte zu sozialen Medien von der Corona-Pandemie und der damit einhergegangenen Desinformation zur Verbreitung des Virus, zur Wirksamkeit von Impfungen und der Angemessenheit von Infektionsschutzmassnahmen geprägt. Somit war ursprünglich angedacht das Thema an Fragen rund um die Informationsqualität der algorithmisch empfohlenen Inhalte auszurichten, um der Nutzerperspektive gerecht zu werden. Im März 2023, nachdem das Thema gewählt wurde, wurde der KI-Bot «ChatGPT» als Freeware veröffentlicht. Daraufhin entbrannte eine intensive Debatte über verschiedene Verfahren des maschinellen Lernens. Folglich wurden auch Social-Media-Algorithmen in der öffentlichen Perspektive vermehrt als ML-Applikationen wahrgenommen. Entsprechend wurde die thematische Fragestellung rekaliibriert, um dem

diskursiven Trend sowie der aktuellen Nutzerperspektive gerecht zu werden. Abhängig davon, wie intensiv die Umfrage-Teilnehmenden Diskurse über künstliche Intelligenz, maschinelles Lernen und Social-Media-Regulation verfolgen, kann die Perspektive auf die Materie unterschiedlich sein. Die unterschiedlichen Perspektiven liessen sich in einer qualitativen Studie besser erheben.

Im Kontext der Schweiz war der Anfang des Jahres 2023 aus einer aufsichtsrechtlichen Perspektive ebenfalls turbulent. Die schweizerische Grossbank Credit Suisse, die in der jüngeren Vergangenheit im Mittelpunkt mehrerer Finanzskandale stand, geriet im Frühjahr 2023 in eine Liquiditätskrise. Am 19. März 2023 gab die Konkurrentin UBS bekannt, die Credit Suisse als Resultat einer staatlichen Einflussnahme, zu übernehmen. Es entbrannte eine öffentliche Debatte über die Schuld und Verantwortung der involvierten staatlichen Akteure, darunter das eidgenössische Finanzdepartement EDF, die schweizerische Nationalbank SNB und die eidgenössische Finanzmarktaufsicht FINMA. Aktuelle Meinungsumfragen zeigen, dass ein beträchtlicher Anteil der schweizerischen Bevölkerung als Folge dieser Bankenkrise eine negative Meinung gegenüber den involvierten Behörden vertritt. So ergab eine Umfrage nach dem Bankendebakel, dass nur 31% der Schweizerinnen und Schweizer die Aufsichtsbehörde FINMA als glaubwürdig wahrnehmen (gfs.bern, 2023). In welchem Ausmass sich dieses Misstrauen auf die Regulation von Social-Media-Diensten übertragen lässt, ist unklar, aber es kann nicht ausgeschlossen werden, dass die Brisanz des Themas die Teilnehmenden der Umfrage in eine kritischere Grundhaltung gegenüber Aufsichtsbehörden versetzt hat.

### 6.3 Methodische Limitationen

Einige methodische Aspekte bedürfen einer kritischen Würdigung. Erstens stützt sich die Literaturanalyse auf Quellen aus verschiedenen wissenschaftlichen Bereichen, von Psychologie und Betriebswirtschaft bis hin zu Informatik und Datenwissenschaft. Geschuldet der interdisziplinären Fragestellung der Bachelor-Thesis wurde eine globale Perspektive auf das Thema eingenommen, was unter anderem dazu führte, dass einige Detailfragen nicht näher beleuchtet werden konnten. Als weitere Limitation der Literaturanalyse ist der hohe Anteil amerikanischer Literatur in den Bereichen Social-Media-Forschung, Informatik, Datenwissenschaften und User eXperience anzumerken, was die Perspektive aufs Thema möglicherweise beeinflusst. Insgesamt ist der Anteil von europäischer und deutschsprachiger Literatur in der Analyse gering, abgesehen von den Quellen, die sich spezifisch mit der EU-Gesetzgebung beschäftigen. Aufgrund der Aktualität der politischen Vorgänge und ungeklärten Fragen zur DSA-Umsetzung mangelte es einigen

Teilhypothesen auch an Schärfe. Das Vorhandensein undefinierter Faktoren und Abhängigkeiten von der effektiven Umsetzung des DSA durch die Social-Media-Dienste erschwerte das Formulieren von verständlichen Fragen für die Datenerhebung.

Die Stichprobengrösse von 124 bedeutet eine limitierte statistische Belastbarkeit der Ergebnisse. Darüber hinaus sind gewisse Unausgewogenheit innerhalb der Stichprobe erwähnenswert. Die Stichprobe enthält eine Überrepräsentation von Frauen und Nutzenden von Instagram. Die Ergebnisse sind unter Berücksichtigung der gewählten statistischen Grenzwerte zwar robust, sofern von einer Grundmenge ausgegangen wird, in der beispielsweise Frauen überrepräsentiert sind und Instagram häufig genutzt wird.

Davon abgesehen gelten für die Erhebung die aus der empirischen Sozialforschung bekannten Limitationen von Online-Befragungen: Zum einen führen Selbstauskünfte immer zu Verzerrungen, weil Teilnehmende ihre Antworten an der sozialen Erwünschtheit ausrichten, also beispielsweise ihren Social-Media-Konsum geringer angeben als wahr. Zum anderen kann ein Non-Response-Bias auftreten, wenn eine bestimmte Personengruppe innerhalb der Grundgesamtheit nicht teilnimmt, wie in dieser Umfrage die Personen aus der französischsprachigen Schweiz. Das Risiko einer Rekrutierungsverzerrung hingegen kann als eher tief eingestuft werden, da ein breites Spektrum von Rekrutierungswegen gewählt wurde, was sich auch in der gleichmässigen geographischen Verteilung und der Altersverteilung innerhalb der Stichprobe zeigt.

Das Testen der Hypothesen anhand des Zustimmungswerts bei den Befragten hat ebenfalls Grenzen. Zum einen misst die Umfrage die persönlichen Einstellungen der Personen, basierend auf ihrem Vorwissen und ihrer Weltanschauung. Jedoch haben sich die wenigsten Nutzenden tiefergehend mit dem Thema beschäftigt. Die persönliche Meinung, die von Nutzerinnen und Nutzern vertreten wird, repräsentiert nicht zwingend die Lösung, die sich Fachleute, Expertinnen, Konsumentenschützer oder unabhängige Organisationen wünschen. Ebenfalls wurden die Hypothesen anhand einer einfachen Mehrheit des Zustimmungswerts getestet. Ob jedoch eine einfache Mehrheit genügt, um die Sinnhaftigkeit eines Governance-Ansatzes zu messen, ist zweifelhaft, insbesondere wenn sich das Resultat nahe an der 50-Prozent-Marke bewegt. Dies war in dieser Umfrage jedoch selten der Fall.

## 6.4 Implikationen

Für die Legislative in der Schweiz kann der Digital Services Act durchaus als Vorbild fungieren, da, so lassen die Umfrageergebnisse vermuten, junge Menschen in der Schweiz diese Governance-Strategien mittragen. Einige Punkte könnten jedoch für

schweizerische Konsumentinnen und Konsumenten optimiert werden. Insbesondere sollte eine dynamische Anpassbarkeit der Empfehlungssysteme, eine potenzielle persönliche Haftbarkeit des Managements, klare ethische Richtlinien, an die sich Plattformen halten müssen, sowie eine klare Pflicht zur Verhinderung von Verbreitung von Desinformation bedacht werden. Das Einbinden dieser Aspekte im Rahmen eines schweizerischen Gesetzesentwurfs könnte ergänzend zu den DSA-Ansätzen das Vertrauen der Nutzenden zusätzlich erhöhen. Ob und wie die Schweiz ein eigenes Gesetz für die Regulation von Social-Media-Plattformen und den eingesetzten Algorithmen durchsetzen kann, ist eine politische Frage und unterliegt den komplexen bilateralen politischen Prozessen zwischen der Schweiz und der EU.

Für künftige Forschung ergeben sich, wie in der Diskussion der Ergebnisse dargelegt spannende Fragen, die es näher zu beleuchten gilt. Mögliche Forschungsfragen für wissenschaftliche Projekte werden nachfolgend in Form einer Liste dargestellt und können künftigen Bachelor- oder Masterarbeiten, aber auch sonstigen Forschungsarbeiten als Inspiration dienen:

- Welche Darstellung nach explainableAI (XAI) ist für die Vermittlung von Empfehlungsparemtern bei Inhaltsempfehlungen für Nutzende am verständlichsten?
- Welche Anforderungen haben Nutzende an die Konfiguration der Empfehlungssysteme?
- Welche Daten-, Algorithmus- und Technikkompetenz ist notwendig, damit Nutzende ihre eigenen Empfehlungsalgorithmen selbstständig konfigurieren können?
- Warum scheinen Social-Media-Nutzende in der Schweiz den EU-Behörden mehr Vertrauen zu schenken, wenn es um Social-Media-Regulation geht?

## 7 Fazit

Die vorliegende Bachelor-Thesis beschäftigt sich mit der Fragestellung, ob die Governance-Strategie nach Digital Services Act für die schweizerische Legislative einen gangbaren Weg darstellt. Dazu wurden zunächst die Dimensionen Gerechtigkeit, Haftbarkeit, Transparenz, Erklärbarkeit und Definierbarkeit zu einem theoretischen Rahmen zusammengefasst. Anhand des Rahmens wurden die wichtigsten Gesetzesartikel im DSA identifiziert. Die Vorschriften aus dem Gesetzesartikel wurden alternativen Governance-Ansätzen aus der wissenschaftlichen Literatur gegenübergestellt. Dazu wurde eine Vielzahl an wissenschaftlicher Literatur aus unterschiedlichen Disziplinen ausgewertet, um den aktuellen Forschungs- und Wissensstand abzubilden und zentrale Konzepte zu identifizieren. Der Literaturteil der Arbeit bietet einen ausführlichen Überblick zur Systematik der Empfehlungssysteme, verschiedenen datenwissenschaftlichen Technologien im Bereich des maschinellen Lernens, Grundlagen von Governance und Policy, sowie eine kritische Auseinandersetzung mit sozialen Medien und den Risiken, die für Nutzende damit einhergehen können.

In der Operationalisierung wurde die Leithypothese, dass eine Mehrheit von jungen Erwachsenen in der Schweiz die Governance-Ansätze zu Social-Media-Algorithmen im Digital Services Act ablehnt, gefasst, um den DSA herauszufordern. Diese Leithypothese wurde in Teilhypothesen zerlegt, die einzelne Governance-Indikatoren darstellten und auf deren Grundlage eine Online-Umfrage konzipiert wurde, die im Juni und Juli 2023 stattfand. Trotz der Stichprobengrösse von 124 Personen kann diese Umfrage mit gewissen Einschränkungen als statistisch relevant betrachtet werden. Die Umfrage brachte zum einen hervor, dass die überwältigende Mehrheit der Befragten die Ansicht vertritt, dass algorithmische Systeme der Plattformen nicht vertrauenswürdig sind und Social-Media-Plattformen beaufsichtigt sowie reguliert werden sollten. Die Erfüllung der FATED-Faktoren wird aktuell als ungenügend bewertet. Daraus lässt sich die hohe Dringlichkeit von Governance zur Steigerung des Vertrauens von Nutzenden. Davon abgesehen zeigten die Ergebnisse eine generelle Befürwortung von Governance nach DSA-Ansätzen. Die Befürwortung war sogar in Punkten sichtbar, zu denen es in der wissenschaftlichen Literatur eine klare Präferenz für andere Ansätze gibt. Beispielsweise wurde die globale Erklärung des Systems nach DSA besser bewertet als lokale Erklärungen, die in der XAI-Literatur häufig vorgeschlagen werden. Abweichungen zu den Governance-Methoden nach DSA bestanden auch bei der Offenlegung der Empfehlungsmechanismen in den AGB nach DSA, die besser bewertet wurde als die Offenlegung der Empfehlungsmechanismen auf der Seite. Im Gegenteil wurde etwa die dynamische Einstellbarkeit von

Recommender-Systemen schlechter bewertet als der DSA-Ansatz der Profiling-freien Recommender-Systeme.

Aus den Ergebnissen lässt sich ableiten, dass in allen Dimensionen nach AX-Modell aus Sicht der Nutzenden mit dem DSA eine Verbesserung bewirkt werden kann. Die Teilmengenanalyse ergab zudem, dass Faktoren wie Geschlecht, Bildungsgrad, Wohnregion und Alter keinen Einfluss auf die Zustimmung zu DSA-Massnahmen oder Governance-Massnahmen insgesamt haben.

Aufgrund der hohen Zustimmungswerte zu den DSA-Ansätzen, die sich bei fast allen Fragen zeigten, konnte die Nullhypothese, dass junge Erwachsene in der Schweiz die Governance-Ansätze zu Social-Media-Algorithmen im Digital Services Act ablehnen, nicht belegt werden. Es ist also von der Gegenhypothese auszugehen. Der Digital Services Act scheint in den Augen von jungen Nutzerinnen und Nutzern durchaus eine attraktive und vertrauensstiftende Governance-Strategie darzustellen.

Die Dynamik in der Debatte rund um das brisante Forschungsthema erschwerte die inhaltliche Ausrichtung und verlangte auch laufende Anpassungen der Fragestellung, insbesondere aufgrund der Debatte die infolge des KI-Bots ChatGPT entfachte. Zudem war im Kontext der Schweiz der Anfang des Jahres 2023 aufsichtsrechtlich turbulent, da die Liquiditätskrise der Grossbank Credit Suisse und die Übernahme durch die Konkurrentin UBS nachweislich das Vertrauen in staatliche Aufsichtsapparate gesenkt hat.

Bei der Umsetzung eines schweizerischen Gesetzes nach Vorbild der EU-Regulation sollten die im DSA fehlenden Aspekte berücksichtigt werden, darunter: Die dynamische Anpassbarkeit der Empfehlungssysteme, eine persönliche Haftbarkeit des Managements, klare ethische Richtlinien, an die sich Plattformen halten müssen, sowie eine Pflicht zur Verhinderung von Desinformation. Künftige Forschung sollte sich unter anderem mit der Frage beschäftigen, welche XAI-Darstellung für die Erklärung von algorithmischen Ergebnissen für Nutzende am verständlichsten ist, sowie welche Anforderungen Nutzende an die Parametrisierung ihrer Empfehlungssysteme stellen. Auch sollte evaluiert werden, welche Daten-, Algorithmus- und Technikkompetenz notwendig ist, damit Nutzende ihre Empfehlungsalgorithmen selbstständig konfigurieren können. Diese Erkenntnisse können bei der Formulierung von künftigen Gesetzen und Richtlinien eine zentrale Rolle spielen, um die Bedürfnisse von Nutzenden gezielt anzusprechen.

Neben konkreten Vorschlägen für Governance-Strategien ist vertiefte Forschung in diesem Bereich nötig, um einen konstruktiven gesamtgesellschaftlichen Diskurs über die Verantwortung, Sicherheit und Konsumentenschutz im Zusammenhang mit Social-Media-Algorithmen zu führen. Diese Arbeit hat Lösungen identifiziert, diskutiert und evaluiert

und soll insbesondere zu weiterer Forschung in diesem Bereich anregen. Da sich die dynamische Landschaft der sozialen Medien und Technologien wie maschinelles Lernen stetig weiterentwickelt, sind viele Aspekte unserer digitalen Zukunft noch ungewiss. Diese Unsicherheit soll Akteurinnen und Akteure jedoch nicht beängstigen, sondern insbesondere Wissenschaftlerinnen und Forscher, politische Entscheidungsträger und Vertreterinnen aus Medien und Gesellschaft motivieren, sich für ein sicheres, gerechtes und transparentes digitales Ökosystem in Europa wie auch in anderen Regionen der Welt einzusetzen.





## 8 Literaturverzeichnis

- Albareda, L. (2008). Corporate responsibility, governance and accountability: From self-regulation to co-regulation. *Corporate Governance: The International Journal of Business in Society*, 8(4), 430–439. <https://doi.org/10.1108/14720700810899176>
- Alpaydin, E. (2010). *Introduction to machine learning* (2nd ed). MIT Press.
- Alvarado, O., Heuer, H., Vanden Abeele, V., Breiter, A., & Verbert, K. (2020). Middle-Aged Video Consumers' Beliefs About Algorithmic Recommendations on YouTube. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2), 1–24. <https://doi.org/10.1145/3415192>
- Alvarado, O., & Waern, A. (2018). Towards Algorithmic Experience: Initial Efforts for Social Media Contexts. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12. <https://doi.org/10.1145/3173574.3173860>
- Amtsblatt der EU (Hrsg.). (2022). Verordnung (EU) über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste). <https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32022R2065>
- Arogyaswamy, B. (2020). Big tech and societal sustainability: An ethical framework. *AI & SOCIETY*, 35(4), 829–840. <https://doi.org/10.1007/s00146-020-00956-6>
- Baccarella, C. V., Wagner, T. F., Kietzmann, J. H., & McCarthy, I. P. (2018). Social media? It's serious! Understanding the dark side of social media. *European Management Journal*, 36(4), 431–438. <https://doi.org/10.1016/j.emj.2018.07.002>
- Bandy, J., & Diakopoulos, N. (2021). More Accounts, Fewer Links: How Algorithmic Curation Impacts Media Exposure in Twitter Timelines. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–28. <https://doi.org/10.1145/3449152>
- Bastick, Z. (2021). Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation. *Computers in Human Behavior*, 116, 106633. <https://doi.org/10.1016/j.chb.2020.106633>
- Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI&SOCIETY*, 30(1), 89–116. <https://doi.org/10.1007/s00146-014-0549-4>
- Bobadilla, J., Alonso, S., & Hernando, A. (2020). Deep Learning Architecture for Collaborative Filtering Recommender Systems. *Applied Sciences*, 10(7), 2441.

<https://doi.org/10.3390/app10072441>

Bongard-Blanchy, K., Rossi, A., Rivas, S., Doublet, S., Koenig, V., & Lenzini, G. (2021). "I am Definitely Manipulated, Even When I am Aware of it. It's Ridiculous!"—Dark Patterns from the End-User Perspective. *Designing Interactive Systems Conference 2021*, 763–776. <https://doi.org/10.1145/3461778.3462086>

Bucher, T. (2017). The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30–44. <https://doi.org/10.1080/1369118X.2016.1154086>

Buchheim, J. (2022). Der Kommissionsentwurf eines Digital Services Act – Regelungsinhalte, Regelungsansatz, Leerstellen und Konfliktpotential. In I. Spiecker gen. Döhmann, M. Westland, & R. Campos (Hrsg.), *Demokratie und Öffentlichkeit im 21. Jahrhundert – zur Macht des Digitalen*. Nomos Verlagsgesellschaft mbH & Co. KG. <https://doi.org/10.5771/9783748932741>

Budzinski, O., & Mendelsohn, J. (2021). Regulating Big Tech: From Competition Policy to Sector Regulation? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3938167>

Bundesamt für Statistik. (2023, Juli 31). *Merkmale der Arbeitskräfte: Bildungsniveau*. <https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/erwerbstaetigkeit-arbeitszeit/merkmale-arbeitskraefte/bildungsniveau.html>

Carvalho, J. M., Arga E Lima, F., & Farinha, M. (2021). Introduction to the Digital Services Act, Content Moderation and Consumer Protection. *Revista de direito e tecnologia*, 3(1), 71–104.

Cauffman, C., & Goanta, C. (2021). A New Order: The Digital Services Act and Consumer Protection. *European Journal of Risk Regulation*, 12(4), 758–774. <https://doi.org/10.1017/err.2021.8>

Chomanski, B. (2021). The Missing Ingredient in the Case for Regulating Big Tech. *Minds and Machines*, 31(2), 257–275. <https://doi.org/10.1007/s11023-021-09562-x>

Chung, M., & Wihbey, J. (2022). Social media regulation, third-person effect, and public views: A comparative study of the United States, the United Kingdom, South Korea, and Mexico. *New Media & Society*, 146144482211229. <https://doi.org/10.1177/14614448221122996>

Cormen, T. H. (Hrsg.). (2009). *Introduction to algorithms* (3rd ed). MIT Press.

- Covington, P., Adams, J., & Sargin, E. (2016). Deep Neural Networks for YouTube Recommendations. *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198. <https://doi.org/10.1145/2959100.2959190>
- Dasgupta, S., Papadimitriou, C. H., & Vazirani, U. V. (2006). *Algorithms*. McGraw-Hill Higher Education.
- Di Noia, T., Tintarev, N., Fatourou, P., & Schedl, M. (2022). Recommender systems under European AI regulations. *Communications of the ACM*, 65(4), 69–73. <https://doi.org/10.1145/3512728>
- Eg, R., Demirkol Tønnesen, Ö., & Tennfjord, M. K. (2023). A scoping review of personalized user experiences on social media: The interplay between algorithms and human factors. *Computers in Human Behavior Reports*, 9, 100253. <https://doi.org/10.1016/j.chbr.2022.100253>
- Engelmann, S., Grossklags, J., & Herzog, L. (2020). Should users participate in governing social media? Philosophical and technical considerations of democratic social media. *First Monday*. <https://doi.org/10.5210/fm.v25i12.10525>
- Europäische Kommission. (2023, April 23). *A Europe fit for the digital age: New online rules for users*. [commission.europa.eu. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-users\\_en](https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment/europe-fit-digital-age-new-online-rules-users_en)
- Fan, J., & Zhang, A. X. (2020). Digital Juries: A Civics-Oriented Approach to Platform Governance. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3313831.3376293>
- Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3518482>
- Freeman, S., Gibbs, M., & Nansen, B. (2023). Personalised But Impersonal: Listeners' Experiences of Algorithmic Curation on Music Streaming Services. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–14. <https://doi.org/10.1145/3544548.3581492>
- Garrett, J. J. (2011). *The Elements of User Experience: User-Centered Design for the Web and Beyond*. Pearson Education.
- gfs.bern. (2023, Juli 27). *Übernahme der CS durch die UBS*. Cockpit gfs.

- <https://cockpit.gfsbern.ch/de/cockpit/234106-uebernahme-der-cs-durch-die-ubs/>
- Gill, A. (2008). Corporate Governance as Social Responsibility: A Research Agenda. *Berkeley Journal of International Law*, 26(2), 452–478.  
<https://doi.org/10.15779/Z38MS9P>
- Gonzales, A. L., & Hancock, J. T. (2011). Mirror, Mirror on my Facebook Wall: Effects of Exposure to Facebook on Self-Esteem. *Cyberpsychology, Behavior, and Social Networking*, 14(1–2), 79–83. <https://doi.org/10.1089/cyber.2009.0411>
- Google. (2023). *Information about Monthly Active Recipients under the Digital Services Act (EU)*. Google. [https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-24\\_2022-7-1\\_2022-12-31\\_en\\_v1.pdf](https://storage.googleapis.com/transparencyreport/report-downloads/pdf-report-24_2022-7-1_2022-12-31_en_v1.pdf)
- Gupta, U., Wu, C.-J., Wang, X., Naumov, M., Reagen, B., Brooks, D., Cottel, B., Hazelwood, K., Hempstead, M., Jia, B., Lee, H.-H. S., Malevich, A., Mudigere, D., Smelyanskiy, M., Xiong, L., & Zhang, X. (2020). The Architectural Implications of Facebook's DNN-Based Personalized Recommendation. *2020 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, 488–501.  
<https://doi.org/10.1109/HPCA47549.2020.00047>
- Hacker, P. (2021). Manipulation by algorithms. Exploring the triangle of unfair commercial practice, data protection, and privacy law. *European Law Journal*, eulj.12389.  
<https://doi.org/10.1111/eulj.12389>
- Haque, A. B., Islam, A. K. M. N., & Mikalef, P. (2023). Explainable Artificial Intelligence (XAI) from a user perspective: A synthesis of prior literature and problematizing avenues for future research. *Technological Forecasting and Social Change*, 186, 122120. <https://doi.org/10.1016/j.techfore.2022.122120>
- Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products – Facets of user experience. *Interacting with Computers*, 22(5), 353–362.  
<https://doi.org/10.1016/j.intcom.2010.04.002>
- Hassenzahl, M., & Tractinsky, N. (2006). User experience—A research agenda. *Behaviour & Information Technology*, 25(2), 91–97.  
<https://doi.org/10.1080/01449290500330331>
- Hoffmann-Riem, W. (2020). Artificial Intelligence as a Challenge for Law and Regulation. In T. Wischmeyer & T. Rademacher (Hrsg.), *Regulating Artificial Intelligence* (S. 1–29). Springer International Publishing. [https://doi.org/10.1007/978-3-030-32361-5\\_1](https://doi.org/10.1007/978-3-030-32361-5_1)

- Jung, A., & Nardelli, P. H. J. (2020). An Information-Theoretic Approach to Personalized Explainable Machine Learning. *IEEE Signal Processing Letters*, 27, 825–829. <https://doi.org/10.1109/LSP.2020.2993176>
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in Social Media Research: Past, Present and Future. *Information Systems Frontiers*, 20(3), 531–558. <https://doi.org/10.1007/s10796-017-9810-y>
- Kelley, P. G., Yang, Y., Heldreth, C., Moessner, C., Sedley, A., Kramm, A., Newman, D. T., & Woodruff, A. (2021). Exciting, Useful, Worrying, Futuristic: Public Perception of Artificial Intelligence in 8 Countries. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 627–637. <https://doi.org/10.1145/3461702.3462605>
- Klumbyte, G., Lücking, P., & Draude, C. (2020). Reframing AX with Critical Design: The Potentials and Limits of Algorithmic Experience as a Critical Design Concept. *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 1–12. <https://doi.org/10.1145/3419249.3420120>
- Koza, J. R., Bennett, F. H., Andre, D., & Keane, M. A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In J. S. Gero & F. Sudweeks (Hrsg.), *Artificial Intelligence in Design '96* (S. 151–170). Springer Netherlands. [https://doi.org/10.1007/978-94-009-0279-4\\_9](https://doi.org/10.1007/978-94-009-0279-4_9)
- Krönke, C. (2020). Artificial Intelligence and Social Media. In T. Wischmeyer & T. Rademacher (Hrsg.), *Regulating Artificial Intelligence* (S. 145–173). Springer International Publishing. [https://doi.org/10.1007/978-3-030-32361-5\\_7](https://doi.org/10.1007/978-3-030-32361-5_7)
- Latzer, M., & Festic, N. (2019). A guideline for understanding and measuring algorithmic governance in everyday life. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1415>
- Leerssen, P. (2023). An end to shadow banning? Transparency rights in the Digital Services Act between content moderation and curation. *Computer Law & Security Review*, 48, 105790. <https://doi.org/10.1016/j.clsr.2023.105790>
- Liao, Q. V., Zhang, Y., Luss, R., Doshi-Velez, F., & Dhurandhar, A. (2022). Connecting Algorithmic Research and Usage Contexts: A Perspective of Contextualized Evaluation for Explainable AI. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 10(1), 147–159. <https://doi.org/10.1609/hcomp.v10i1.21995>

- Lodder, A. R., & Carvalho, J. M. (2022). Online Platforms: Towards an Information Tsunami with New Requirements on Moderation, Ranking, and Traceability. *European Business Law Review*, 33(Issue 4), 537–556.  
<https://doi.org/10.54648/EULR2022025>
- Lopes, P., Silva, E., Braga, C., Oliveira, T., & Rosado, L. (2022). XAI Systems Evaluation: A Review of Human and Computer-Centred Methods. *Applied Sciences*, 12(19), 9423. <https://doi.org/10.3390/app12199423>
- Mac, R., Warzel, C., & Kantrowitz, A. (2018, März 30). Growth At Any Cost: Top Facebook Executive Defended Data Collection In 2016 Memo—And Warned That Facebook Could Get People Killed. *Buzzfeed News*. <https://www.buzzfeednews.com/article/ryanmac/growth-at-any-cost-top-facebook-executive-defended-data>
- McKay, S., & Tenove, C. (2021). Disinformation as a Threat to Deliberative Democracy. *Political Research Quarterly*, 74(3), 703–717.  
<https://doi.org/10.1177/1065912920938143>
- Medvedev, I., Gordon, T., & Wu, H. (2019). Powered by AI\_ Instagram’s Explore recommender system.pdf. *Facebook AI Blog*. <https://ai.facebook.com/blog/powered-by-ai-instagrams-explore-recommender-system/>
- Medzini, R. (2022). Enhanced self-regulation: The case of Facebook’s content governance. *New Media & Society*, 24(10), 2227–2251.  
<https://doi.org/10.1177/1461444821989352>
- Medzini, R., & Levi-Faur, D. (2023). Self-Governance via Intermediaries: Credibility in Three Different Modes of Governance. *Journal of Comparative Policy Analysis: Research and Practice*, 1–23. <https://doi.org/10.1080/13876988.2022.2155516>
- Meta (Hrsg.). (2023). Digital Services Act—Information on Average Monthly Active Recipients in the European Union. <https://transparency.fb.com/sr/dsa-report-feb2023/>
- Papadakis, H., Papagrigoriou, A., Panagiotakis, C., Kosmas, E., & Fragopoulou, P. (2022). Collaborative filtering recommender systems taxonomy. *Knowledge and Information Systems*, 64(1), 35–74. <https://doi.org/10.1007/s10115-021-01628-7>
- Pérez, B., & Fernández-Aller, C. (2022). Facebook and Artificial Intelligence: A Review of Good Practices.
- Popat, A., & Tarrant, C. (2023). Exploring adolescents’ perspectives on social media and mental health and well-being – A qualitative literature review. *Clinical Child Psychology and Psychiatry*, 28(1), 323–337.

<https://doi.org/10.1177/13591045221092884>

- Prakash Yadav, G., & Rai, J. (2017). The Generation Z and their Social Media Usage: A Review and a Research Outline. *Global Journal of Enterprise Information System*, 9(2), 110. <https://doi.org/10.18311/gjeis/2017/15748>
- Pu, P., Chen, L., & Hu, R. (2011). A user-centric evaluation framework for recommender systems. *Proceedings of the Fifth ACM Conference on Recommender Systems*, 157–164. <https://doi.org/10.1145/2043932.2043962>
- Rader, E. (2017). Examining user surprise as a symptom of algorithmic filtering. *International Journal of Human-Computer Studies*, 98, 72–88. <https://doi.org/10.1016/j.ijhcs.2016.10.005>
- Rader, E., & Gray, R. (2015). Understanding User Beliefs About Algorithmic Curation in the Facebook News Feed. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173–182. <https://doi.org/10.1145/2702123.2702174>
- Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Hrsg.). (2011). *Recommender Systems Handbook*. Springer US. <https://doi.org/10.1007/978-0-387-85820-3>
- Robles Carrillo, M. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy*, 44(6), 101937. <https://doi.org/10.1016/j.telpol.2020.101937>
- Roski, J., Maier, E. J., Vigilante, K., Kane, E. A., & Matheny, M. E. (2021). Enhancing trust in AI through industry self-governance. *Journal of the American Medical Informatics Association*, 28(7), 1582–1590. <https://doi.org/10.1093/jamia/ocab065>
- Rossiter, N., & Zehle, S. (2015). The Aesthetics of Algorithmic Experience. In R. Martin (Hrsg.), *The Routledge Companion to Art and Politics* (1. Aufl., S. 214–221). Routledge. <https://doi.org/10.4324/9781315736693-26>
- Salgado-Criado, J., & Fernandez-Aller, C. (2021). A Wide Human-Rights Approach to Artificial Intelligence Regulation in Europe. *IEEE Technology and Society Magazine*, 40(2), 55–65. <https://doi.org/10.1109/MTS.2021.3056284>
- Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3), 210–229. <https://doi.org/10.1147/rd.33.0210>
- Schwemer, S. F. (2021). Recommender Systems in the EU: From Responsibility to Regulation. *Morals & Machines*, 1(2), 60–69. <https://doi.org/10.5771/2747-5174->

2021-2-60

Senffleben, M. (2021). Trademark Law, AI-Driven Behavioural Advertising and the Digital Services Act – Towards Source and Parameter Transparency for Consumers, Brand Owners and Competitors. *SSRN Electronic Journal*.

<https://doi.org/10.2139/ssrn.3947739>

Shin, D. (2020a). How do users interact with algorithm recommender systems? The interaction of users, algorithms, and performance. *Computers in Human Behavior*, 109, 106344. <https://doi.org/10.1016/j.chb.2020.106344>

Shin, D. (2020b). User Perceptions of Algorithmic Decisions in the Personalized AI System: Perceptual Evaluation of Fairness, Accountability, Transparency, and Explainability. *Journal of Broadcasting & Electronic Media*, 64(4), 541–565.

<https://doi.org/10.1080/08838151.2020.1843357>

Shin, D., Lim, J. S., Ahmad, N., & Ibahrine, M. (2022). Understanding user sensemaking in fairness and transparency in algorithms: Algorithmic sensemaking in over-the-top platform. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01525-9>

Shokeen, J., & Rana, C. (2020). Social recommender systems: Techniques, domains, metrics, datasets and future scope. *Journal of Intelligent Information Systems*, 54(3), 633–667. <https://doi.org/10.1007/s10844-019-00578-5>

Sigfrids, A., Leikas, J., Salo-Pöntinen, H., & Koskimies, E. (2023). Human-centricity in AI governance: A systemic approach. *Frontiers in Artificial Intelligence*, 6, 976887. <https://doi.org/10.3389/frai.2023.976887>

Smith, B. (2021, Dezember 5). How TikTok Reads Your Mind. *The New York Times*. <https://www.nytimes.com/2021/12/05/business/media/tiktok-algorithm.html>

Steurer, R. (2013). Disentangling governance: A synoptic view of regulation by government, business and civil society. *Policy Sciences*, 46(4), 387–410.

<https://doi.org/10.1007/s11077-013-9177-y>

Su, X., & Khoshgoftaar, T. M. (2009). A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 2009, 1–19. <https://doi.org/10.1155/2009/421425>

Sundar, S. S. (2020). Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74–88. <https://doi.org/10.1093/jcmc/zmz026>

Taylor, S. H., & Choi, M. (2022). An Initial Conceptualization of Algorithm



- Responsiveness: Comparing Perceptions of Algorithms Across Social Media Platforms. *Social Media + Society*, 8(4), 205630512211443. <https://doi.org/10.1177/20563051221144322>
- TikTok. (2023). *Investing in our 150m strong community in Europe*. <https://news-room.tiktok.com/en-eu/investing-for-our-150-m-strong-community-in-europe>
- Turillazzi, A., Taddeo, M., Floridi, L., & Casolari, F. (2023). The digital services act: An analysis of its ethical, legal, and social implications. *Law, Innovation and Technology*, 15(1), 83–106. <https://doi.org/10.1080/17579961.2023.2184136>
- Twitter. (2023a, März 31). Twitter's Recommendation Algorithm. *Twitter Blog*. [https://blog.twitter.com/engineering/en\\_us/topics/open-source/2023/twitter-recommendation-algorithm](https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm)
- Twitter. (2023b, April 24). *Average Monthly Active Recipients of Service (AMARS) in the EU*. Twitter. <https://transparency.twitter.com/en/reports/amars-in-the-eu.html>
- Vogel, E. A., Rose, J. P., Roberts, L. R., & Eckles, K. (2014). Social comparison, social media, and self-esteem. *Psychology of Popular Media Culture*, 3(4), 206–222. <https://doi.org/10.1037/ppm0000047>
- Wang, G., Guo, Y., Zhang, W., Xie, S., & Chen, Q. (2023). What type of algorithm is perceived as fairer and more acceptable? A comparative analysis of rule-driven versus data-driven algorithmic decision-making in public affairs. *Government Information Quarterly*, 40(2), 101803. <https://doi.org/10.1016/j.giq.2023.101803>
- Wolfers, L. N., & Utz, S. (2022). Social media use, stress, and coping. *Current Opinion in Psychology*, 45, 101305. <https://doi.org/10.1016/j.copsy.2022.101305>
- Zhang, B., & Dafoe, A. (2020). U.S. Public Opinion on the Governance of Artificial Intelligence. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 187–193. <https://doi.org/10.1145/3375627.3375827>
- Zhang, Y., Liao, Q. V., & Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295–305. <https://doi.org/10.1145/3351095.3372852>
- Zódi, Z. (2022). Characteristics of the European Platform Regulation: Platform Law and User Protection. *Public Governance, Administration and Finances Law Review*, 7(1), 91–108. <https://doi.org/10.53116/pgaf.2022.1.7>



## 9 Anhang

### 9.1 Anhang 1: Fragebogen

# Aufsicht und Regulation von Social-Media-Algorithmen

Meinungsumfrage unter Social-Media-Nutzer\*innen

Vielen Dank, dass Sie sich die Zeit nehmen, bei dieser Umfrage mitzumachen.

Die Meinungsumfrage ist anonym, richtet sich an Personen im Alter **ab 18 Jahren bis und mit 35 Jahren, die in der Schweiz leben oder in der Schweiz arbeiten (Grenzgänger\*innen)** und **mindestens eine der folgenden Plattformen** benutzen:

- Facebook,
- Instagram,
- TikTok,
- Twitter und/oder
- Youtube.

Diese Umfrage wird im Rahmen einer Bachelor-Thesis über verschiedene Ansätze der Regulation und Aufsicht (Governance) von Social-Media-Algorithmen und -Plattformen durchgeführt. Das Ziel der Erhebung ist, festzustellen, welche Governance-Strategien aus Sicht von Konsument\*innen positiv wahrgenommen werden.

Die Umfrage besteht aus **16 Fragen** und die Teilnahme dauert ungefähr **10 bis 15 Minuten**.

Die Umfrage endet am 20.07.2023 um 23:59.

*Kontakt:*  
*R. Jevdenic*  
*Bachelor-Student Information Science*  
*Fachhochschule Graubünden*

## Demographische Daten

Bitte geben Sie Ihr Alter an.

*\*(Zielgruppe der Umfrage sind 18- bis und mit 35-jährige)*

🔗 Dies ist eine Frage-Hilfetext.

🔗 Bitte überprüfen Sie das Format Ihrer Antwort.

Bitte geben Sie Ihr Geschlecht an.

🔗 Bitte wählen Sie eine der folgenden Antworten:

- Mann
- Frau
- Divers

Bitte geben Sie Ihren höchsten Bildungsabschluss an.

🔗 Bitte wählen Sie eine der folgenden Antworten:

- Obligatorischer Schulabschluss
- Berufsabschluss (Lehre, EFZ)
- Gymnasial-, Berufs- oder Fachmatura
- Höhere Fachschule, Meisterdiplom
- Bachelor-Abschluss
- Master-Abschluss
- Doktorat
- Andere

Diese Umfrage richtet sich an Personen, die in der Schweiz leben, ihren Lebensmittelpunkt in der Schweiz haben oder eine enge Bindung zur Schweiz haben (z.B. Grenzgänger\*innen).

Bitte geben Sie die Region an, in der Sie wohnhaft sind.

🔗 Bitte wählen Sie eine der folgenden Antworten:

- Genferseeregion (GE, VD, VS)
- Espace Mittelland (BE, FR, JU, NE, SO)
- Nordwestschweiz (AG, BL, BS)
- Zürich
- Ostschweiz (AR, AI, GL, GR, SG, SH, TG)
- Innerschweiz (LU, NW, OW, SZ, UR, ZG)
- Tessin
- Ausland (Grenzgänger\*in oder im Auslandsaufenthalt)

### Einleitung

Im Einleitungsteil dieser Umfrage wird Ihre generelle Einstellung gegenüber Social-Media-Plattformen, den eingesetzten Algorithmen und Governance (Regulation und Aufsicht) erhoben.

Social-Media-Algorithmen sind Systeme, die Inhalte wie Posts, Tweets, Bilder oder Videos für Sie automatisch **selektieren, priorisieren und empfehlen**. Dies betrifft **personalisierte Feeds** (Instagram Discovery-Page, TikTok For-You-Page, Twitter Feed, Facebook Timeline) wie auch die Youtube-Seitenleiste mit Videorempfehlungen.

Wie häufig\* verwenden Sie schätzungsweise diese Social-Media-Plattformen?

\*ausgehend von einer Interaktionsdauer von mind. 10min

	nie	selten (1-3 mal monatlich)	ab und zu (1-3 mal wöchentlich)	häufig (1-3 mal täglich)	sehr häufig (mehr als 3 mal täglich)
Facebook	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Instagram	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
TikTok	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Twitter	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Youtube	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wie stehen Sie derzeit zur Aufsicht und Haftbarkeit von Social-Media-Plattformen?

Social-Media-Plattformen ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... werden im Falle von Schäden an Menschen (z.B. durch Desinformation) oder der Gesellschaft (z.B. Wahlmanipulation) genügend zur Rechenschaft gezogen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... sollen reguliert und beaufsichtigt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Social-Media-Algorithmen **selektieren, priorisieren und empfehlen** auf Social-Media-Plattformen Inhalte wie Posts, Tweets, Bilder oder Videos. Wie stehen Sie derzeit zu diesen Algorithmen?

Diese Algorithmen ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... sind nach meinen Bedürfnissen einstellbar.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... sind vertrauenswürdig und agieren in meinem Interesse.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... sind transparent in Bezug auf ihre Funktionsweise.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... behandeln alle Nutzer*innen gleich (Reichweite, Empfehlungsqualität).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... empfehlen mir Inhalte, die mir entsprechen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Risiken von Social-Media-Algorithmen

In diesem Abschnitt wird Ihre Meinung zum Umgang mit Risiken, die durch Social-Media-Algorithmen entstehen können, erhoben.  
 Dabei geht es um Algorithmen, die Inhalte wie Posts, Tweets, Bilder oder Videos für Sie automatisiert **selektieren, priorisieren und empfehlen**. Dies betrifft **personalisierte Feeds** (Instagram Discovery-Page, TikTok For-You-Page, Twitter Feed, Facebook Timeline) wie auch die YouTube-Seitenleiste mit Videoempfehlungen.  
 Schäden sind **Nachteile, Beeinträchtigungen oder Verluste**, die durch die Nutzung von Social-Media-Plattformen auftreten können. Hierbei kann es sich persönliche Schäden handeln, die die Lebensqualität von Menschen beeinträchtigen, oder auch Schäden an der Gesellschaft, wenn sie den sozialen Zusammenhalt oder demokratische Prozesse negativ beeinflussen.

Auf welcher **Grundlage** sollten **Massnahmen zum Schutz** von Nutzer\*innen getroffen werden?  
 Social-Media-Plattformen sollten Massnahmen zum Schutz von Nutzer\*innen auf Grundlage von ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... klaren ethischen Vorgaben im Gesetz treffen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... Analyse und Bewertung möglicher Risiken durch die Plattformbetreiber selbst treffen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Welche Arten von Schäden und Risiken sollten Social-Media-Plattformen verhindern müssen?  
 Social-Media-Plattformen sollten verpflichtet sein, ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... die Entstehung schwerer Schäden (z. B. Depression, Trauma) an Menschen durch ihre Dienste zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... Verbreitung von Desinformation ("Fake News") zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... die Zunahme von geschlechterspezifischer Gewalt (z.B. durch Gewaltverherrlichung online) zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... gesamtgesellschaftliche Polarisierung (Verstärkung von Meinungsunterschieden) zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... die Verbreitung illegaler Inhalte über ihre Dienste zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... politische Radikalisierung von Einzelnen oder Personengruppen zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... Risiken für gesellschaftliche Debatten, Wahlprozesse und die öffentliche Sicherheit zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... die übermässigen Online-Konsum (Internetsucht) zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... persönliche Diskriminierungen und diskriminierende Inhalte zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... moderate nachteilige Auswirkungen (Schlafstörungen, "FOMO") zu verhindern.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

### Haftbarkeit und Aufsicht

In diesem Abschnitt wird Ihre Meinung zur Haftbarkeit und Rechenschaftspflicht von Social-Media-Plattformen erhoben.

**Haftbarkeit** beschreibt, welche Person oder Instanz rechtlich belangt werden kann. **Aufsicht** beschreibt, dass die Tätigkeiten der Dienste überwacht werden.

Schäden sind **Nachteile, Beeinträchtigungen oder Verluste**, die durch die Nutzung von Social-Media-Plattformen auftreten können. Hierbei kann es sich persönliche Schäden handeln, die die Lebensqualität von Menschen beeinträchtigen, oder auch Schäden an der Gesellschaft, wenn sie den sozialen Zusammenhalt oder demokratische Prozesse negativ beeinflussen.

Welche **Person oder Instanz** sollte im Falle von Schäden haftbar sein?

Social-Media-Plattformen sollten im Schadenfall...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
...in Form der verantwortlichen Programm-Entwickler*innen strafrechtlich haftbar sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...in Form der Personen im Management (Entscheidungsträger*innen) strafrechtlich haftbar sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
...als Unternehmung wirtschaftsrechtlich haftbar sein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Durch welche **Prüf- und Kontrollinstanz** sollten Social-Media-Plattformen beaufsichtigt werden?

Social-Media-Plattformen sollten ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... durch Plattform-Nutzer*innen demokratisch beaufsichtigt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... durch staatliche Behörden beaufsichtigt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... durch die Plattformbetreiber selbst anhand gesetzlicher Regeln beaufsichtigt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... durch Plattformbetreiber auf freiwilliger Basis selbst beaufsichtigt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... durch unabhängige, nicht-staatliche Prüforganisationen beaufsichtigt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

In welchem Land resp. Länderbund sollten **staatliche Kontroll- und Prüfinstanzen** angesiedelt sein?

Social-Media-Plattformen sollten beaufsichtigt werden durch ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... eine spezialisierte Behörde im Domizilland (Unternehmenssitz) des Social-Media-Dienstes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... eine spezialisierte EU-Behörde im europäischen Binnenmarkt (inkl. Schweiz, Norwegen, Island, Liechtenstein).	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... eine schweizerische Behörde, sofern Dienste hierzulande verfügbar sind.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wie hoch ist Ihr Vertrauen in diese staatlichen Instanzen bezüglich der Aufsicht der angegebenen Dienste?

Ich vertraue der Aufsicht von...

	Überhaupt nicht	Eher nicht	Weder noch	Eher schon	Voll und Ganz	Weiss nicht
... chinesischen Diensten (z.B. Tiktok) durch chinesische Behörden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... amerikanischen Diensten (z. B. Facebook, Instagram, Twitter) durch amerikanische Behörden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... amerikanischen und chinesischen Dienste durch EU-Behörden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... amerikanischen und chinesischen Dienste durch schweizer Behörden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



## Transparenz und Nachvollziehbarkeit

In diesem Abschnitt wird Ihre Meinung zu Transparenz und Nachvollziehbarkeit der Social-Media-Algorithmen erhoben. Dabei geht es um Algorithmen, die Inhalte wie Posts, Tweets, Bilder oder Videos für Sie automatisiert selektieren, priorisieren und empfehlen.

Dies betrifft **personalisierte Feeds** (Instagram Discovery-Page, TikTok For-You-Page, Twitter Feed, Facebook Timeline) wie auch die Youtube-Seitenleiste mit Videoempfehlungen.

Diese Empfehlungen werden auf Grundlage von **Personendaten** (z. B. Alter, Geschlecht, Wohnort, persönliche Interessen, politische Einstellung, sexuelle Orientierung) über Sie und andere Nutzer\*innen getroffen.

Wo sollte die **Funktionsweise von Social-Media-Algorithmen** offengelegt werden?

Social-Media-Plattformen sollten die Funktionsweise der Social-Media-Algorithmen...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... in den AGB offenlegen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... in der App oder auf der Webseite offenlegen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Sollten die zur Empfehlung verwendeten Daten generalisiert (z.B. in den AGBs) oder fallbezogen (z.B. bei jeder Empfehlung) angezeigt werden?

Social-Media-Plattformen sollten die Personendaten, die für die Generierung von Empfehlungen verwendet werden, ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... in den AGB generalisiert beschreiben.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... bei jeder Empfehlung (vorgeschlagener Post, Tweet, Video) personalisiert anzeigen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wünschen Sie sich eine Möglichkeit, **Social-Media-Algorithmen nach Ihren Bedürfnissen** anzupassen?

Social-Media-Plattformen sollten mir die Möglichkeit bieten ...

	Stimme überhaupt nicht zu	Stimme nicht zu	weder noch	Stimme zu	Stimme völlig zu	Weiss nicht
... selbst zu entscheiden, welche Personendaten über mich zur Erstellung von Empfehlungen (vorgeschlagener Post, Tweet, Video) verwendet werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
... Empfehlungen (vorgeschlagener Post, Tweet, Video) ohne Verwendung von Personendaten über mich zu generieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## 9.2 Anhang 2: Gesamtergebnisse Umfrage

<b>Einstellung Aufsicht, Regulation und den algorithmischen Systemen</b>	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Einstellung zu Aufsicht und Regulation von Social-Media-Plattformen								
Social-Media-Plattformen werden im Falle von Schäden genügend zur Rechenschaft gezogen	12	(9.7%)	9	(7.3%)	89	(71.8%)	14	(11.3%)
Social-Media-Plattformen sollen reguliert und beaufsichtigt werden.	103	(83.7%)	13	(10.6%)	4	(3.3%)	3	(2.4%)
<b>Einstellung zu algorithmischen Systemen</b>								
Social-Media-Algorithmen sind nach meinen Bedürfnissen einstellbar.	39	(31.5%)	18	(14.5%)	61	(49.2%)	6	(4.8%)
Social-Media-Algorithmen sind vertrauenswürdig und agieren in meinem Interesse.	29	(23.4%)	27	(21.8%)	64	(51.6%)	4	(3.2%)
Social-Media-Algorithmen sind transparent in Bezug auf ihre Funktionsweise.	10	(8.1%)	9	(7.3%)	101	(81.5%)	4	(3.2%)
Social-Media-Algorithmen behandeln alle Nutzer*innen gleich.	12	(9.7%)	16	(12.9%)	69	(55.6%)	27	(21.8%)
Social-Media-Algorithmen empfehlen mir Inhalte, die mir entsprechen.	82	(66.7%)	20	(16.3%)			1	(0.8%)
<b>Einstellung zu Risiken auf sozialen Medien</b>								
Social-Media-Plattformen sollten Massnahmen zum Schutz von Nutzer*innen auf Grundlage von klaren ethischen Vorgaben im Gesetz treffen.	105	(84.7%)	9	(7.3%)	4	(3.2%)	6	(4.8%)
Analyse und Bewertung möglicher Risiken durch die Plattformbetreiber selbst treffen.	56	(45.2%)	16	(12.9%)	42	(33.9%)	10	(8.1%)
Social-Media-Plattformen sollten verpflichtet sein, ...								
die Entstehung schwerer Schäden an Menschen durch ihre Dienste zu verhindern.	97	(78.2%)	6	(4.8%)	16	(12.9%)	5	(4%)
Verbreitung von Desinformation ("Fake News") zu verhindern.	111	(89.5%)	5	(4%)	7	(5.6%)	1	(0.8%)
die Zunahme von geschlechterspezifischer Gewalt zu verhindern.	115	(92.7%)	3	(2.4%)	5	(4%)	1	(0.8%)
gesamtgemeinschaftliche Polarisierung (Verstärkung von Meinungsunterschieden) zu verhindern.	59	(47.6%)	24	(19.4%)	32	(25.8%)	9	(7.3%)
die Verbreitung illegaler Inhalte über ihre Dienste zu verhindern.	108	(87.1%)	10	(8.1%)	6	(4.8%)	0	(0%)
politische Radikalisierung von Einzelnen oder Personengruppen zu verhindern.	89	(71.8%)	14	(11.3%)	14	(11.3%)	7	(5.6%)
Risiken für gesellschaftliche Debatten, Wahlprozesse und die öffentliche Sicherheit zu verhindern.	73	(58.9%)	25	(20.2%)	19	(15.3%)	7	(5.6%)
die übermässigen Online-Konsum (Internetsucht) zu verhindern.	72	(58.1%)	23	(18.5%)	24	(19.4%)	5	(4%)
persönliche Diskriminierungen und diskriminierende Inhalte zu verhindern.	112	(90.3%)	4	(3.2%)	7	(5.6%)	1	(0.8%)
moderate nachteilige Auswirkungen (Schlafstörungen, "FOMO") zu verhindern.	49	(39.5%)	33	(26.6%)			9	(7.3%)

	Stimme zu		Weder noch		Stimme nicht zu		Weiss nicht	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
<b><u>Einstellung zu Haftbarkeit und Aufsicht</u></b>								
Social-Media-Plattformen sollten im Schadenfall ...								
in Form der verantwortlichen Programm-Entwickler*innen strafrechtlich haftbar sein.	34	(27.4%)	13	(10.5%)	63	(50.8%)	14	(11.3%)
in Form der Personen im Management (Entscheidungsträger*innen) strafrechtlich haftbar sein.	91	(73.4%)	13	(10.5%)	11	(8.9%)	9	(7.3%)
als Unternehmung wirtschaftsrechtlich haftbar sein.	111	(89.5%)	4	(3.2%)	5	(4%)	4	(3.2%)
Social-Media-Plattformen sollten ...								
durch Plattform-Nutzer*innen demokratisch beaufsichtigt werden.	61	(49.2%)	23	(18.5%)	31	(25%)	9	(7.3%)
durch staatliche Behörden beaufsichtigt werden.	70	(56.5%)	15	(12.1%)	29	(23.4%)	10	(8.1%)
durch die Plattformbetreiber selbst anhand gesetzlicher Regeln beaufsichtigt werden.	77	(62.1%)	18	(14.5%)	22	(17.7%)	7	(5.6%)
durch Plattformbetreiber auf freiwilliger Basis selbst beaufsichtigt werden.	31	(25%)	8	(6.5%)	74	(59.7%)	11	(8.9%)
durch unabhängige, nichtstaatliche Prüforganisationen beaufsichtigt werden.	96	(77.4%)	7	(5.6%)	15	(12.1%)	6	(4.8%)
Social-Media-Plattformen sollten beaufsichtigt werden durch ...								
eine spezialisierte Behörde im Domizilland (Unternehmenssitz) des Social-Media-Dienstes.	46	(37.1%)	19	(15.3%)	43	(34.7%)	16	(12.9%)
eine spezialisierte EU-Behörde im europäischen Binnenmarkt (inkl. Schweiz, Norwegen, Island, L.	93	(75%)	8	(6.5%)	13	(10.5%)	10	(8.1%)
eine schweizerische Behörde, sofern Dienste hierzulande verfügbar sind.	86	(69.4%)	16	(12.9%)	14	(11.3%)	8	(6.5%)
Ich vertraue der Aufsicht von...								
chinesischen Diensten (z.B. Tiktok) durch chinesische Behörden	4	(3.2%)	6	(4.8%)	110	(88.7%)	4	(3.2%)
amerikanischen Diensten (z. B. Facebook, Instagram, Twitter) durch amerikanische Behörden	10	(8.1%)	18	(14.5%)	95	(76.6%)	1	(0.8%)
amerikanischen und chinesischen Dienste durch EU-Behörden	74	(59.7%)	16	(12.9%)	31	(25%)	3	(2.4%)
amerikanischen und chinesischen Dienste durch schweizer Behörden	83	(66.9%)	9	(7.3%)	28	(22.6%)	4	(3.2%)

<b>Einstellung zu Transparenzpflichten von algorithmischen Systemen</b>	<b>Stimme zu</b>		<b>Weder noch</b>		<b>Stimme nicht zu</b>		<b>Weiss nicht</b>	
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
Social-Media-Plattformen sollten die Funktionsweise der Social-Media-Algorithmen... in den AGB offenlegen.	99	(80.5%)	7	(5.7%)	12	(9.8%)	5	(4.1%)
in der App oder auf der Webseite offenlegen.	114	(91.9%)	3	(2.4%)	5	(4%)	2	(1.6%)
Social-Media-Plattformen sollten die für Empfehlungen verwendeten Personendaten ... in den AGB generalisiert beschreiben.	89	(72.4%)	10	(8.1%)	13	(10.6%)	11	(8.9%)
bei jeder Empfehlung (vorgeschlagener Post, Tweet, Video) personalisiert anzeigen.	80	(64.5%)	20	(16.1%)	14	(11.3%)	10	(8.1%)
Social-Media-Plattformen sollten mir die Möglichkeit bieten ... selbst zu entscheiden, welche meiner Personendaten für Empfehlungen verwendet werden	111	(89.5%)	7	(5.6%)	3	(2.4%)	3	(2.4%)
Empfehlungen ohne Verwendung von Personendaten über mich zu generieren.	96	(77.4%)	15	(12.1%)	8	(6.5%)	5	(4%)

---

## Bisher erschienene Schriften

Ergebnisse von Forschungsprojekten erscheinen jeweils in Form von Arbeitsberichten in Reihen.  
Sonstige Publikationen erscheinen in Form von alleinstehenden Schriften.

Derzeit gibt es in den Churer Schriften zur Informationswissenschaft folgende Reihen:  
Reihe Berufsmarktforschung

### Weitere Publikationen

Churer Schriften zur Informationswissenschaft – Schrift 161

Herausgegeben von Wolfgang Semar

Céline Graf

«Ghostbusters Münstergasse»

Vermittlung von regionalen Onlineressourcen und Recherchekompetenzen mit einem digitalen

Educational Escape Room an der Bibliothek Münstergasse der Universitätsbibliothek Bern

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 162

Herausgegeben von Wolfgang Semar

Mahmoud Hemila

Qualitätsanalyse von inhaltsbasierten Empfehlungssystemen für Journals

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 163

Herausgegeben von Wolfgang Semar

Nicolas Brauchli

Inwiefern unterscheiden sich die Online-Plattformen der Legacy-Medien von den Digital Born

Plattformen in der Deutschschweizer Medienlandschaft?

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 164

Herausgegeben von Wolfgang Semar

Flurin Böni

Das verborgene Gold am Ende des Rainbow-Washing

Eine Analyse der Vereinbarkeit sozialen Engagements mit unternehmerischen Zielen

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 165

Herausgegeben von Wolfgang Semar

Alina Viert

Herausforderungen in der Aufbewahrung von Videospielen und ihrer Peripherie

Fragen und Antworten insbesondere zur Peripherie und zur Emulation als Lösungsansatz

Chur 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 166

Herausgegeben von Wolfgang Semar

Susanne Knöpfel

Wissenslandkarten als Grundlage für Visualisierungen im Wissensmanagement

Chur, 2023

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 167

Herausgegeben von Wolfgang Semar

Lorena Staiger

Deep Web und Bibliotheken: Stand der Dinge

Chur, 2023

ISSN 1660-945X

---

Churer Schriften zur Informationswissenschaft – Schrift 168  
Herausgegeben von Wolfgang Semar  
Karin Mattmann  
Positive Darstellungen archivarischer Tätigkeiten in Fiktion  
Wie das Abbild von fiktionalem Archivpersonal in der Öffentlichkeit positiv und realistisch dargestellt werden kann  
Chur, 2023  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 169  
Herausgegeben von Wolfgang Semar  
Stefan Banzer  
Codemigration mit ChatGPT  
Evaluation von ChatGPT als Tool zur teilautomatisierten Codeübersetzung von COBOL Code zu Python Code  
Chur, 2023  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 170  
Herausgegeben von Wolfgang Semar  
Marion Spitz  
Digitale Nudges zwischen Moral und Manipulation  
Eine quantitative Inhaltsanalyse zu den Auswirkungen ethischer Aspekte auf die erforschte Wirksamkeit von digitalen Nudges  
Chur, 2024  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 171  
Herausgegeben von Wolfgang Semar  
Joy Walser  
Erschliessungsmöglichkeiten einer Sammlung mit Records in Contexts  
Entwicklung und Anwendung eines konzeptionellen Modells für die Sammlung  
«Pfarrer F. Tschugmell, Siegel- und Stempelsammlung»  
Chur, 2024  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 172  
Herausgegeben von Wolfgang Semar  
Alessio Monte  
Potenzialanalyse zur Anwendung von KI-basierten  
Chur, 2024  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 173  
Herausgegeben von Wolfgang Semar  
Lisa Köllner  
Der Familienbezug und seine Bedeutung für die Nutzung von Firmenarchiven durch Familienunternehmen am Beispiel aktuell tätiger Unternehmen  
Chur, 2024  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 174  
Herausgegeben von Wolfgang Semar  
Silvia Rutz  
Psychologische Sicherheit in virtuellen agilen Teams  
Eine explanative Analyse der Einflussfaktoren auf die psychologische Sicherheit in virtuellen agilen Software-Teams  
Chur, 2024  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 175  
Herausgegeben von Wolfgang Semar  
Jérôme Gander  
Information Governance und öffentliche Verwaltung  
Definitionen, Nutzen und die Rolle der Verwaltungsarchive.  
Chur, 2024  
ISSN 1660-945X

---

## Über die Informationswissenschaft der Fachhochschule Graubünden

Die Informationswissenschaft ist in der Schweiz noch ein relativ junger Lehr- und Forschungsbereich. International weist diese Disziplin aber vor allem im anglo-amerikanischen Bereich eine jahrzehntelange Tradition auf. Die klassischen Bezeichnungen dort sind Information Science, Library Science oder Information Studies. Die Grundfragestellung der Informationswissenschaft liegt in der Betrachtung der Rolle und des Umgangs mit Information in allen ihren Ausprägungen und Medien sowohl in Wirtschaft und Gesellschaft. Die Informationswissenschaft wird in Chur integriert betrachtet.

Diese Sicht umfasst nicht nur die Teildisziplinen Bibliothekswissenschaft, Archivwissenschaft und Dokumentationswissenschaft. Auch neue Entwicklungen im Bereich Medienwirtschaft, Informations- und Wissensmanagement und Big Data werden gezielt aufgegriffen und im Lehr- und Forschungsprogramm berücksichtigt.

Der Studiengang Informationswissenschaft wird seit 1998 als Vollzeitstudiengang in Chur angeboten und seit 2002 als Teilzeit-Studiengang in Zürich. Seit 2010 rundet der Master of Science in Business Administration das Lehrangebot ab.

Der Arbeitsbereich Informationswissenschaft vereinigt Cluster von Forschungs-, Entwicklungs- und Dienstleistungspotenzialen in unterschiedlichen Kompetenzzentren:

- Information Management & Competitive Intelligence
- Collaborative Knowledge Management
- Information and Data Management
- Records Management
- Library Consulting
- Information Laboratory
- Digital Education

Diese Kompetenzzentren werden im Swiss Institute for Information Science (SII) zusammengefasst.

---

## Impressum

### Impressum

FHGR - Fachhochschule  
Graubünden  
Information Science  
Pulvermühlestrasse 57  
CH-7000 Chur

[www.informationsscience.ch](http://www.informationsscience.ch)

[www.fhgr.ch](http://www.fhgr.ch)

**ISSN 1660-945X**

### Institutsleitung

Prof. Dr. Ingo Barkow

Telefon: +41 81 286 24 61

Email: [ingo.barkow@fhgr.ch](mailto:ingo.barkow@fhgr.ch)

### Sekretariat

Telefon: +41 81 286 24 24

Fax: +41 81 286 24 00

Email: [clarita.decurtins@fhgr.ch](mailto:clarita.decurtins@fhgr.ch)