



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Digitale Souveränität und Künstliche Intelligenz: Eine
genderbasierte Analyse von Herausforderungen

verfasst von | submitted by

Linda Norma Brocke BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Arts (MA)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 824

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Politikwissenschaft

Betreut von | Supervisor:

Mag. Dr. Marion Löffler Privatdoz.

Inhaltsverzeichnis

Einleitung.....	4
1 Theorie und Begriffsklärung.....	6
1.1 Digitale Souveränität.....	6
1.2 Plattformmacht.....	9
1.3 Algorithmische Verzerrung (Bias) & Hegemoniale Männlichkeit	11
1.3.1 Algorithmische Verzerrungen	11
1.3.2 Hegemoniale Männlichkeit	13
2 Forschungsstand	15
2.1 Machtstrukturen und demokratische Herausforderungen bei Meta	16
2.2 Geschlechterungleichheit und algorithmischer Bias	18
2.3 EU-Regulierungen für KI & Plattformen.....	20
2.3.1 Regulierungsbedarf im digitalen Raum.....	20
2.3.2 Datenschutz-Grundverordnung (DSGVO).....	22
2.3.3 Der Digital Markets Act (DMA)	23
2.3.4 Der Digital Services Act (DSA).....	24
2.3.5 Der EU AI-Act	25
2.3.6 Ergänzende Regulierungen.....	27
2.4 Zwischenfazit	28
3 Methodologie	30
3.1 Theoretische Fundierung.....	30
3.1.1 Methodische Anwendung.....	31
3.1.2 Datenbasis und Materialauswahl	32
3.1.3 Grenzen und Erkenntnispotenzial.....	33
3.2 Critical Frame Analysis.....	33
3.2.1 Methodologische Grundlagen	33
3.2.2 Anwendung in dieser Arbeit und Materialauswahl	34
3.2.3 Erkenntnispotenzial der CFA	36
3.3 Datenbasis & Abgrenzung	36
4 Strukturelle Machtanalyse von Meta nach Foucault	36
4.1 Plattformmacht und algorithmische Inhaltsmoderation bei Meta	37
4.2 Algorithmische Inhaltsmoderation als Dispositiv struktureller Macht	40
4.3 Normalisierung: Technische Reproduktion gesellschaftlicher Normen	41
4.4 Gouvernamentalität: Steuerung durch Plattformlogiken	42
4.5 Digitale Souveränität und die Grenzen regulatorischer Kontrolle	44

4.6 Zwischenfazit.....	46
5 Critical Frame Analyse.....	47
5.1 Neutralitäts- und Objektivitätsframe	48
5.1.1 Blogpost: Objektivität durch Forschung und Technik.....	49
5.1.2 Senatsanhörung: Technik als Überwachungsinstanz	50
5.1.3 Instagram-Statement: Metas Strategiewechsel	51
5.1.4 EU- und Geschlechterperspektive zum Frame Neutralität & Objektivität	53
5.2 Effizienz- und Sicherheitsframe	54
5.2.1 Blogpost: Technologische Machbarkeit und moralischer Fortschritt	55
5.2.2 Senatsanhörung: Konsolidierung des Frames Effizienz & Sicherheit.....	56
5.2.3 Instagram-Statement: Neujustierung von Effizienz & Sicherheit	57
5.2.4 EU- und Geschlechterperspektive zum Frame Effizienz & Sicherheit	58
5.3 Fortschritts- und Innovationsframe	59
5.3.1 Blogpost: Technologische Fortschritte in Metas Kommunikationsstrategie ..	60
5.3.2 Senatsanhörung: Fortschrittsrhetorik Zuckerbergs	61
5.3.3 Instagram-Statement: Neuausrichtung des Frames.....	62
5.3.4 EU- und Geschlechterperspektive zum Frame Fortschritt & Innovation	64
5.4 Vermeidung von Verantwortung (Responsibility Shifting)	65
5.4.1 Blogpost: Verantwortungsdiffusion durch Open Source und Kollaboration..	66
5.4.2 Senatsanhörung: Ablehnung redaktioneller Verantwortung.....	66
5.4.3 Instagram-Statement: Delegation an Community-basierte Systeme	67
5.4.4 EU- und Geschlechterperspektive zum Frame Responsibility-Shifting	69
5.5 Zwischenfazit.....	70
6 Diskussion der Ergebnisse	71
6.1 Zusammenführung der Erkenntnisse	71
6.2 Digitale Souveränität und die Rolle der EU	72
6.3 Reproduktion von Geschlechterungleichheit durch algorithmische Systeme	74
6.4 Grenzen und Implikationen der Analyse	77
7 Beantwortung der Forschungsfrage	79
8 Fazit	82
Literaturverzeichnis.....	84
Anhang A	93
Anhang B.....	95
Anhang C	102

Abstract (Deutsch)

Diese Masterarbeit untersucht, wie KI-gestützte Moderationssysteme von Meta die digitale Souveränität der Europäischen Union und geschlechtsspezifische Ungleichheiten beeinflussen. Theoretisch basiert sie auf Foucaults Konzept struktureller Macht und Raewyn Connells Theorie hegemonialer Männlichkeit. Methodisch kombiniert sie eine Critical Frame Analysis öffentlicher Meta-Kommunikation mit einer Analyse der Plattforminfrastruktur. Die Ergebnisse zeigen, dass Meta durch proprietäre technische Systeme, algorithmische Sichtbarkeitssteuerung und eigene Community-Standards erhebliche strukturelle Macht ausübt, die durch Frames wie *Neutralität & Effizienz* oder *Fortschritt & Innovation* legitimiert wird. Dies erschwert die wirksame Umsetzung von EU-Regulierungen wie DSA, DMA und AI-Act und wird durch transatlantische Dynamiken verstärkt. Zudem können Moderationsalgorithmen marginalisierte Perspektiven benachteiligen, wodurch digitale Souveränität ohne Geschlechtergerechtigkeit unvollständig bleibt. Die Arbeit kommt zu dem Schluss, dass nachhaltige Regulierung nur gelingt, wenn strukturelle Plattformmacht und ihre diskursive Absicherung zugleich adressiert werden – durch technische Transparenz, konsequente Durchsetzung und kritische Beobachtung gesellschaftlicher Effekte.

Abstract (English)

This master's thesis examines how Meta's AI-supported moderation systems influence the digital sovereignty of the European Union and gender inequalities. Theoretically, it is based on Foucault's concept of structural power and Raewyn Connell's theory of hegemonic masculinity. Methodologically, it combines a critical frame analysis of public Meta communication with an analysis of the platform infrastructure. The results show that Meta exercises considerable structural power through proprietary technical systems, algorithmic visibility control, and its own community standards, which are legitimized by frames such as *neutrality & efficiency* or *progress & innovation*. This hinders the effective implementation of EU regulations such as the DSA, DMA, and AI Act and is reinforced by transatlantic dynamics. In addition, moderation algorithms can disadvantage marginalized perspectives, leaving digital sovereignty incomplete without gender equality. The paper concludes that sustainable regulation can only succeed if structural platform power and its discursive safeguards are addressed simultaneously through technical transparency, consistent enforcement, and critical observation of societal effects.

Einleitung

In den vergangenen Jahren hat sich die Nutzung und Weiterentwicklung von KI-Systemen rasant entwickelt und ist mittlerweile in zahlreichen gesellschaftlichen und wirtschaftlichen Bereichen fest verankert. In Branchen wie Medizin, Finanzwesen oder Verkehr üben Algorithmen und automatisierte Entscheidungssysteme zunehmend Einfluss auf zentrale Entscheidungsprozesse aus (Wiegerling et al., 2018: 4). Generative KI-Tools wie ChatGPT, das seit 2022/2023 einen anhaltenden Hype erlebt, verdeutlichen die zunehmende Verbreitung und Macht dieser Technologien.

Unter anderem durch die Nutzung solcher KI-Tools hinterlassen Nutzer:innen fortlaufend Daten auf den Servern großer Technologieunternehmen. Diese Daten dienen nicht nur der Verbesserung der Algorithmen, sondern bilden die Grundlage eines datenbasierten Geschäftsmodells (Kaufmann, 2015: 1). Künstliche Intelligenz (KI) umfasst in diesem Zusammenhang Systeme, die in der Lage sind, ihre Umgebung zu analysieren und auf Basis dieser Analyse selbstständig Handlungen auszuführen, um bestimmte Ziele zu erreichen. Dabei kann es sich sowohl um rein softwarebasierte Anwendungen wie Sprachassistenten oder Bilderkennungssoftware als auch um physisch eingebettete Systeme wie autonome Fahrzeuge oder Drohnen handeln (European Commission, 2018: 1). Das hebt KI von konventionellen Analyseprogrammen ab, wobei die Grenzen dabei oftmals nicht eindeutig sind. Unternehmen wie Google, Microsoft, Meta oder Amazon, aber auch chinesische Konzerne wie Alibaba oder Baidu, sammeln systematisch Daten, kategorisieren diese und verkaufen sie weiter, um gezielte Marketingstrategien und Produktempfehlungen zu ermöglichen. Der Einsatz von KI hat diese datenbasierten Prozesse nicht nur effizienter und automatisierter gestaltet, sondern ermöglicht auch eine deutlich präzisere und individuellere Ausrichtung von Marketingstrategien und Produktempfehlungen. Daten sind damit zu einer zentralen Ressource geworden, die nicht nur wirtschaftlichen, sondern auch politischen Wert besitzt und in der digitalen Wirtschaft als Ware, Kapital oder Vermögenswert betrachtet wird (Sadowski, 2019; Birch et al., 2021).

Diese Form der Datenkontrolle hat weitreichende Konsequenzen für die digitale Souveränität von Staaten. Christian Katzenbach und Lena Ulbricht (2019) diskutieren in diesem Zusammenhang, wie digitale Technologien neue Formen der Kontrolle und politischen Gestaltung ermöglichen und dabei Fragen von Macht, Interessen und Regulierung ins Zentrum rücken. Die Fähigkeit, digitale Infrastrukturen und Datenströme eigenständig zu gestalten und zu regulieren, wird dabei zunehmend als zentrale Voraussetzung für

politische Handlungsfähigkeit verstanden (Katzenbach & Ulbricht, 2019). In der Europäischen Union hat dies zu intensiven Debatten über den Einfluss großer Technologieunternehmen und über die Notwendigkeit regulatorischer Maßnahmen geführt. Als politische Reaktionen auf diese Problemlagen wurden unter anderem die EU-KI-Verordnung und der Digital Services Act auf den Weg gebracht.

Besonders problematisch ist, dass KI-Systeme bestehende Geschlechterungleichheiten reproduzieren oder sogar verstärken können – etwa durch diskriminierende Trainingsdaten, stereotype Mustererkennung oder unausgewogene technische Entwicklungsteams (European Institute for Gender Equality, 2022: 12). Insofern stellt sich die Frage, inwiefern Staaten in der Lage sind, faire und inklusive digitale Infrastrukturen zu gestalten und entsprechende Regulierungen zu entwickeln. Algorithmen sind nicht neutral, sondern reproduzieren häufig bestehende Stereotypen, die in den Trainingsdaten enthalten sind (Feast, 2019: Abs. 7). Ein prägnantes Beispiel hierfür ist der Einsatz von KI-gestützter Bewerbungssoftware. So wurde in mehreren Fällen dokumentiert, dass Bewerbungsalgorithmen Frauen* systematisch benachteiligten, indem bestimmte Berufe oder Qualifikationen aufgrund historisch männlich geprägter Datensätze priorisiert wurden. Dies verdeutlicht, wie automatisierte Entscheidungsprozesse bestehende Stereotype und Ungleichheiten fortschreiben können. Ein weiteres Beispiel zeigt sich auf Social Media Plattformen, wo Algorithmen Inhalte, die traditionelle Schönheitsideale betonen, bevorzugt anzeigen. Dies kann zu einer verstärkten Objektifizierung von Frauen* und der Reproduktion geschlechtsspezifischer Rollenbilder führen. Der Einsatz von KI kann somit bestehende strukturelle Ungleichheiten verstärken, während der Einfluss der betroffenen Bevölkerungsgruppen auf die Technologiegestaltung begrenzt bleibt (Beer, 2019: 127). Diese Arbeit untersucht, wie algorithmische Entscheidungen von US-Tech-Konzernen digitale Souveränität und Geschlechtergerechtigkeit in der EU erschweren oder deren Realisierung blockieren. Als Beispiel für US-Tech Konzerne wird die Plattform Meta betrachtet, insbesondere der Einsatz von KI zur Inhaltsmoderation und deren algorithmische Bias. Die Untersuchung analysiert Material aus dem Zeitraum von 2018 bis 2025. Zum einen um aktuelle Entwicklungen und EU-Regulierungen (z. B. Digital Services Act, EU-KI-Verordnung) zu berücksichtigen und zum anderen, um die seit 2018 verstärkt politische und regulatorischen Maßnahmen zur Kontrolle von KI-Technologien mitebeziehen zu können. Gleichzeitig markieren die Jahre 2018 und 2025 signifikante Entwicklungen: 2018 trat die DSGVO in Kraft, was die Debatte um digitale Souveränität intensivierte,

und bis 2025 soll die Umsetzung der EU-KI-Verordnung abgeschlossen sein, wodurch sich wesentliche Veränderungen in der Regulierung und Nutzung von KI abzeichnen.

Des Weiteren soll Mark Zuckerberg als Beispiel eines Technologieführers, dessen unternehmerische Entscheidungen erheblichen Einfluss auf digitale Souveränität und Geschlechtergerechtigkeit in der EU haben, dienen. Zuckerberg, als Gründer und CEO von Meta (ehemals Facebook), kontrolliert eine der größten sozialen Plattformen weltweit und sein Handeln hat durch den Umgang seines Unternehmens mit Daten, Algorithmen und Inhaltsmoderation direkte Auswirkungen auf globale digitale Machtverhältnisse. Seine Strategien, insbesondere in Bezug auf Datenschutz, algorithmische Transparenz und politische Einflussnahme, sind zentrale Faktoren, die untersucht werden, um die Verbindung zwischen digitaler Souveränität und Geschlechterungleichheit zu analysieren. Dabei ist hervorzuheben, dass Zuckerbergs weitreichender Einfluss auf digitale Infrastrukturen und gesellschaftliche Diskurse nicht demokratisch legitimiert ist, sondern auf seiner unternehmerischen Macht basiert. Dennoch beeinflussen seine Entscheidungen die digitale Souveränität von Staaten weltweit. Die besondere Rolle der EU, die in der Vergangenheit immer wieder versucht hat, mit verschiedenen Instrumenten – etwa durch strengere gesetzliche Rahmenbedingungen – die Macht von Technologiekonzernen gezielt einzuschränken, soll in diesem Zusammenhang genauer untersucht werden.

1 Theorie und Begriffsklärung

Die folgenden theoretischen Grundlagen dienen dazu, die zentralen Begriffe und Konzepte dieser Arbeit in ihren politischen, technologischen und gesellschaftlichen Kontext einzuordnen. Dabei wird insbesondere herausgearbeitet, wie digitale Souveränität durch globale Plattformkonzerne herausgefordert werden, algorithmische Systeme gesellschaftliche Ungleichheiten verstärken und welche strukturelle Macht diesen Prozessen zugrunde liegt.

1.1 Digitale Souveränität

Für die theoretische Fundierung des Begriffs *digitale Souveränität* wird zunächst ein klassisches Verständnis von Souveränität herangezogen, wie es Georg Glasze et al. (2022) in Anlehnung an Gerhard Kiersch (1977) formulieren. Souveränität bedeutet dabei nach außen die rechtliche Gleichrangigkeit souveräner Staaten im Völkerrecht – einschließlich der Freiheit, Bündnisse einzugehen oder außenpolitisch sowie militärisch zu handeln. Nach innen verweist sie auf die politische und rechtliche Hoheitsgewalt über ein

Territorium und dessen Bevölkerung, die notfalls auch mit Gewalt durchgesetzt werden kann (vgl. Glasze et al., 2022: 9).

Auf dieser Grundlage hat sich das Konzept der digitalen Souveränität als Reaktion auf gesellschaftliche Transformationen durch Digitalisierung und Globalisierung herausgebildet. Um seine Entwicklung angemessen einzuordnen, ist ein Blick auf die politischen und historischen Kontexte unerlässlich.

Bereits in den 1990er- und frühen 2000er-Jahren tauchten in Politik und Medien Begriffe wie „Ende des Nationalstaats“ oder „Postsouveränität“ auf, die im Kontext von Globalisierung und Digitalisierung diskutiert wurden (ebd.: 11). Zahlreiche Autor:innen stellten die Verbindung von staatlicher Herrschaft und territorialer Infrastruktur infrage. Sie diagnostizierten eine Auflösung der sogenannten „infrastructural power“ nationalstaatlicher Strukturen (ebd.). In den 2010er-Jahren verlagerte sich der Diskurs zunehmend hin zur Forderung nach digitaler Souveränität, die in den 2020er-Jahren weiter an Bedeutung gewann. Schlagworte wie „Rückkehr des Nationalstaats“ oder „Ende der Globalisierung“ prägten nun zunehmend die politische Landschaft (ebd.: 12).

Auch auf Ebene der EU ist der Begriff „digitale Souveränität“ mittlerweile zentraler Bestandteil politischer Agenden. Die Europäische Kommission (n.d.-a) betont in offiziellen Erklärungen, Europa müsse seine digitale Souveränität stärken, eigene Standards setzen und sich technologisch unabhängiger machen – mit besonderem Fokus auf Daten, Technologie und Infrastruktur.

Dabei ist zu beachten, dass die Auslegung digitaler Souveränität stark vom jeweiligen politischen und geografischen Kontext abhängt. Im europäischen Diskurs ist der Begriff oft mit demokratisch legitimierten Selbstbestimmungsansprüchen verbunden, während er in autoritäreren Kontexten – etwa in Russland oder China – zur Legitimation nationaler Kontrolle über digitale Räume herangezogen wird (ebd.: 17).

Der Begriff der digitalen Souveränität ist demnach weder eindeutig definiert noch statisch. Er ist vielmehr kontextabhängig und eng verknüpft mit politischen Interessen und Machtverhältnissen (ebd.: 11). Für die vorliegende Arbeit wird er daher im Sinne des europäischen Verständnisses verwendet – als demokratisch legitimer Anspruch auf Kontrolle über digitale Infrastruktur und Datenflüsse.

Dieses kontextgebundene Begriffsverständnis kann durch die Perspektive von Lokke Moerel und Paul Timmers (2021) ergänzt werden, die digitale Souveränität als moderne Form strategischer Autonomie verstehen, welche im digitalen Raum spezifische Anforderungen an staatliche Handlungsfähigkeit mit sich bringt (Moerel & Timmers, 2021: 8). Während

Souveränität klassisch mit territorialer Kontrolle, rechtlicher Zuständigkeit sowie interner und externer Legitimität verbunden ist, bezieht sich digitale Souveränität auf die Fähigkeit eines Staates, in zentralen digitalen Bereichen – etwa hinsichtlich digitaler Infrastrukturen, Datenflüsse und technologischer Systeme – unabhängig entscheiden und handeln zu können. Sie umfasst sowohl die Kontrolle über digitale Systeme selbst als auch über die in diesen erzeugten und verarbeiteten Daten. In diesem Zusammenhang wird auch der Begriff der Datensouveränität verwendet, der die Hoheit über Speicherung, Verarbeitung und Zugriffsrechte auf Daten einschließt. Damit bildet digitale Souveränität einen wesentlichen Baustein langfristiger staatlicher Handlungsfähigkeit im digitalen Zeitalter (ebd, 2021: 8). Dazu zählen unter anderem Datenhoheit, regulatorische Unabhängigkeit und technologische Eigenständigkeit.

Diese Sichtweise wird von Luciano Floridi (2020: 370) noch deutlich weiter gefasst. Er versteht digitale Souveränität als den Kampf um die Kontrolle über sämtliche Ebenen des Digitalen: Daten, Software, Standards und Protokolle, Prozesse, Hardware, digitale Dienste und schließlich Infrastrukturen – kurz: um die Kontrolle des Digitalen insgesamt. In Floridis Worten: „In short, for the control of the digital“ (Floridi, 2020: 371). Dieses umfassende Verständnis zeigt, dass digitale Souveränität nicht allein mit staatlicher Unabhängigkeit gleichzusetzen ist, sondern auch mit der Fähigkeit, technologische, normative und infrastrukturelle Grundlagen des Digitalen selbst gestalten zu können.

In diesem Zusammenhang erscheint die Rolle globaler Technologieunternehmen besonders problematisch: Sie kontrollieren zentrale digitale Infrastrukturen, ohne demokratisch legitimiert zu sein – ein Zustand, der fundamentale Souveränitätsfragen aufwirft. Globale Technologiekonzerne wie Meta prägen durch ihre marktbeherrschende Stellung und technologische Überlegenheit nationale sowie europäische digitale Infrastrukturen nachhaltig. Dies hat tiefgreifende Auswirkungen auf die Bevölkerung, da die Dominanz weniger Plattformen die digitale Souveränität der EU-Mitgliedstaaten gefährdet – nicht zuletzt, weil diese Unternehmen oft außerhalb des Zugriffs nationalstaatlicher Regulierung agieren. Gleichzeitig werden im weiteren Verlauf die spezifischen Herausforderungen europäischer Digitalpolitik herausgearbeitet.

Shoshana Zuboff (2018) zeigt in ihrem Werk *Das Zeitalter des Überwachungs kapitalismus* eindrucklich, wie US-amerikanische Plattformunternehmen durch systematische Datensammlung und -verwertung asymmetrische Machtverhältnisse gegenüber Staaten und Individuen etablierten. Daraus resultiert eine strukturelle Abhängigkeit Europas von den digitalen Infrastrukturen dieser Unternehmen. Dies ermöglicht Konzernen wie Meta,

Google oder Amazon nicht nur ökonomische, sondern auch politische Einflussnahme. Zuboff (2018: 37) warnt, dass dadurch private Unternehmen und der Markt innerhalb dieses Systems, sowohl die totale Herrschaft über die menschlichen, gesellschaftlichen und politischen Sphären beanspruchen, als auch die kollektive menschliche Souveränität untergraben. Dementsprechend stelle dies auch eine Bedrohung für liberale Demokratien des Westens dar (ebd.: 37). Sie unterstreicht insbesondere, dass die regulatorischen Handlungsspielräume und die technologische Eigenständigkeit von Staaten massiv eingeschränkt sind. Die Logiken des Überwachungskapitalismus festigen asymmetrische Machtverhältnisse durch Datenextraktion und algorithmisch gesteuerte Informationsverarbeitung (ebd.: 552-554).

Um jedoch zu verstehen, warum digitale Souveränität in der europäischen Debatte zunehmend an Bedeutung gewinnt, reicht ein rein staatenzentrierter Blick nicht aus. Vielmehr muss die Rolle derjenigen Akteure in den Blick genommen werden, die digitale Infrastrukturen nicht nur bereitstellen, sondern zunehmend kontrollieren: globale Plattformunternehmen. Ihre ökonomische und infrastrukturelle Macht stellt eine zentrale Herausforderung für staatliche Gestaltungsansprüche dar – insbesondere dann, wenn diese Macht die Regeln des digitalen Raums faktisch mitbestimmt. Das folgende Kapitel zeigt deshalb auf, wie sich die Plattformmacht von Konzernen wie Meta konstituiert, welche Mechanismen sie stärkt und inwiefern diese Macht die politischen und wirtschaftlichen Ordnungen Europas herausfordert.

1.2 Plattformmacht

Im Kontext digitaler Souveränität ist das Verständnis von Plattformmacht von zentraler Bedeutung, da die Kontrolle über digitale Plattformen maßgeblich die Handlungsfähigkeit von Staaten und Gesellschaften im digitalen Raum beeinflusst. Um die Auswirkungen dieser Macht zu erfassen, ist es notwendig, die Unternehmensstrukturen globaler Technologiekonzerne im Rahmen des aktuellen Wirtschaftssystems und ihrer politischen Bedeutung zu beleuchten. Dabei stellen sich zentrale Fragen: Wie entsteht die wirtschaftliche Macht dieser Konzerne? Welches Ausmaß erreicht sie? Und inwiefern verändert die Konzentration dieser Macht den traditionellen Umgang mit Wettbewerbsrecht?

Mit diesen Fragen beschäftigt sich Philipp Staab ausführlich in seinem Werk *Digitaler Kapitalismus. Markt und Herrschaft in der Ökonomie der Unknappheit* (2019). Er bezeichnet die durch Digitalisierung und Globalisierung bedingte Transformation des Wirtschaftssystems als »digitalen Kapitalismus«. Diese neue ökonomische Ordnung zeichnet

sich durch neu konfigurierte Machtverhältnisse aus, die durch die technologischen Infrastrukturen global agierender Plattformen ermöglicht werden (Staab, 2019: 8).

Zentral für Staabs Analyse ist der Begriff des „proprietären Marktes“: Plattformen wie Meta, Google oder Amazon bilden nicht nur Märkte ab – sie konstruieren sie aktiv mit und setzen deren Regeln. Das bedeutet, sie kontrollieren Zugänge, bestimmen die Rahmenbedingungen und beeinflussen maßgeblich den Wettbewerb (ebd.: 134).

Auch Nick Srnicek (2018) analysiert die strukturelle Macht digitaler Plattformen und beschreibt sie als eine neue dominierende Unternehmensform im digitalen Kapitalismus. In *Plattform-Kapitalismus* unterscheidet er verschiedene Plattfortmtypen – etwa Werbeplattformen (z. B. Google, Facebook) oder Cloud-Plattformen (z. B. Amazon Web Services) – und zeigt, dass ihr Geschäftsmodell systematisch auf der Sammlung, Analyse und Monetarisierung von Nutzerdaten basiert (Srnicek, 2018: 53-54.). Dabei profitieren sie von Netzwerkeffekten, Skalierbarkeit und Tendenzen zur Monopolisierung, die zu einer enormen Konzentration ökonomischer Macht führen (ebd.: 98-99).

Obwohl Nutzer:innen theoretisch frei entscheiden können, welche Dienste sie in Anspruch nehmen oder wie ihre Daten verarbeitet werden, ist diese Wahl in der Praxis stark eingeschränkt. Die Marktdominanz weniger großer Anbieter führt dazu, dass Alternativen kaum wettbewerbsfähig sind – ihre Nutzung wird schnell zur sozialen Hürde oder gar zur Ausgrenzung (Staab 2019: 150, 136). Wer etwa keine Dienste der marktführenden Anbieter nutzt, kann kaum online Reisen buchen, kommunizieren, einkaufen oder auf öffentliche Dienstleistungen zugreifen.

Obwohl es Fortschritte im Verbraucherschutz gibt, bestimmen weiterhin wenige große Konzerne die Spielregeln. Die wirtschaftlich mächtigsten Plattformunternehmen – insbesondere die sogenannten Big Five (Google, Apple, Facebook/Meta, Amazon und Microsoft) – agieren nicht innerhalb einer staatlich regulierten Marktwirtschaft im Sinne liberal-kapitalistischer Prinzipien. Vielmehr verfügen sie über eine Form infrastruktureller Kontrolle, die klassische Marktmechanismen unterläuft und ihnen ermöglicht, zentrale Bedingungen wirtschaftlicher und gesellschaftlicher Teilhabe eigenständig zu gestalten (ebd.: 222).

Zusammenfassend lässt sich feststellen, dass digitale Plattformen durch ihre Kontrolle über Datenflüsse, Infrastruktur und Interaktionslogiken eine erhebliche strukturelle Macht ausüben. Sie dominieren nicht nur wirtschaftliche Märkte, sondern auch die Bedingungen öffentlicher Kommunikation. Ihre Fähigkeit, Marktmechanismen

algorithmisch zu gestalten und das Verhalten der Nutzer:innen zu steuern, beschreibt Staab als neue Form digitaler Herrschaft (ebd.: 144–147).

1.3 Algorithmische Verzerrung (Bias) & Hegemoniale Männlichkeit

Mit der neuen Form digitaler Herrschaft erweitert sich die Entscheidungsmacht digitaler Plattformen über öffentliche Kommunikationsräume – insbesondere durch den Einsatz KI-gestützter Moderationssysteme. Diese Systeme gelten oft als neutral, reproduzieren jedoch bestehende Geschlechterungleichheiten (Noble, 2018; Crawford, 2021). Besonders problematisch ist dies im Bereich der automatisierten Inhaltsmoderation, bei der Beiträge von Frauen*, marginalisierten Gruppen oder feministischen Bewegungen überproportional häufig gelöscht oder herabgestuft werden (Buolamwini & Gebru, 2018).

Diese Arbeit überträgt das Konzept der hegemonialen Männlichkeit (Connell, 2015) auf die technologische Vorherrschaft von Unternehmen wie Meta, um strukturelle Machtverhältnisse im digitalen Raum sichtbar zu machen. Raewyn Connell versteht hegemoniale Männlichkeit nicht nur als individuelles Verhalten, sondern als gesellschaftlich dominante Norm, die andere Geschlechtsidentitäten und Männlichkeitsformen systematisch unterordnet (ebd.: 131–132). In digitaltechnologischen Kontexten zeigt sich diese Hegemonie beispielsweise in der strukturellen Bevorzugung männlich konnotierter Werte wie Objektivität, Effizienz und Kontrolle – etwa in der Entwicklung und Anwendung algorithmischer Systeme. So hilft Connells Ansatz zu erklären, wie Plattformrichtlinien und KI-gestützte Moderation nicht geschlechtsneutral sind, sondern geschlechtsspezifische Hierarchien – etwa durch Verzerrungen in Datensätzen, Entwickler:innenkulturen oder Inhaltsbewertungen – aktiv stabilisieren und reproduzieren.

1.3.1 Algorithmische Verzerrungen

Die Forschung von Safiya Noble (2018) legt vielfältigen Diskriminierungsmechanismen offen, die durch Algorithmen entstehen können. In *Algorithms of Oppression: How Search Engines Reinforce Racism* zeigt Noble, wie insbesondere Google-Suchalgorithmen sexistische und rassistische Stereotype verstärken. Sie weist nach, dass Suchergebnisse systematisch Vorurteile reproduzieren, weil die Algorithmen von Menschen entwickelt wurden, die in einer von strukturellem Rassismus und Sexismus geprägten Gesellschaft sozialisiert wurden (Noble, 2018).

Ein markantes Beispiel hierfür war der erste Treffer bei der Suchanfrage „black girls“: Die Top-Ergebnisse verwiesen auf pornografische Inhalte, was die rassistisch-sexistische Funktionsweise der algorithmischen Logik eindrücklich illustriert (Noble, 2018: 3).

Noble kritisiert die weit verbreitete Annahme, Technologien seien objektiv oder neutral, und argumentiert, dass Algorithmen ebenso wie ihre Entwickler:innen Vorurteile mittragen und verstärken.

Auch David Beer (2019) zeigte in seiner Forschung, dass algorithmische Systeme nicht nur bestehende diskriminierende Muster reproduzieren, sondern diese sogar verstärken können. Zudem führen sie zu neuen Formen von Kontrolle und Wahrnehmung, wodurch sie aktiv an der Konstruktion gesellschaftlicher Wirklichkeit mitwirken. So können bestimmte demografische Gruppen unterschiedlich kategorisiert und im Vergleich zu anderen systematisch benachteiligt werden. Beer (2019: 127) verdeutlicht darüber hinaus, wie Unternehmen durch intransparente algorithmische Prozesse gezielt Einfluss auf die Sichtbarkeit von Inhalten nehmen – ein Mechanismus, der bestehende soziale Hierarchien stabilisiert. Dabei betont er, dass es sich hierbei nicht lediglich um ein technisches, sondern um ein tiefgreifend gesellschaftlich verankertes Machtproblem handelt.

Eine ebenfalls besonders relevante empirische Studie stammt von Joy Buolamwini und Timnit Gebru (2018). Sie untersuchen in ihrem Artikel „*Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*“ die Genauigkeit von kommerziellen Geschlechtsidentifikationssystemen (z.B. von IBM, Microsoft und Face++) in Bezug auf intersektionale Unterschiede. Dabei fokussieren sie sich auf die Diskriminierung aufgrund von Geschlecht und Hautfarbe. Bei der Analyse der Fehlerraten der KI-Systeme, wurden insgesamt höhere Fehlerquoten bei Frauen festgestellt. Am höchsten war die Fehlerquote bei Schwarzen Frauen (bis zu 34,7%) (Buolamwini & Gebru, 2018: 11). Gleichzeitig wiesen hellhäutige Männer eine nur sehr geringe Fehlerquote auf (unter 1%) (ebd.). Dadurch konnte verdeutlicht werden, dass intersektionale Diskriminierungen durch Algorithmen, Ungleichheiten erheblich verstärken. Das impliziert die erhebliche Verzerrung kommerzieller KI-Systeme und die dadurch verursachte Benachteiligung von marginalisierten Gruppen. Daher weisen die Autor:innen auf die Notwendigkeit einer diverseren Datengrundlage, transparente Algorithmen und intersektionale Fairness in der KI-Entwicklung hin (ebd.: 12). Sie fordern zudem Regulierungsmaßnahmen und ethische Richtlinien, um diskriminierenden Technologien entgegenzuwirken.

Die genannte Literatur zeigt, dass algorithmische Verzerrungen keine zufälligen technischen Fehler sind, sondern Ausdruck und Verstärker bestehender gesellschaftlicher Machtverhältnisse. Algorithmen reproduzieren soziale Ungleichheiten nicht nur, sie können diese – wie die Ergebnisse von Buolamwini und Gebru (2018) sowie Noble (2018) zeigen – sogar systematisch verschärfen, insbesondere entlang intersektionaler

Kategorien wie Geschlecht und Race. Es wurde verdeutlicht, dass algorithmische Moderationssysteme auf Plattformen wie denen von Meta nicht als neutrale Werkzeuge betrachtet werden dürfen, sondern als machtvolle Akteure, die aktiv in soziale Ordnungsprozesse eingreifen.

Aus dieser Perspektive wird verständlich, warum die Analyse algorithmischer Verzerrungen im Kontext der digitalen Souveränität und geschlechtsspezifischen Diskriminierung relevant ist. Digitale Plattformen beeinflussen nicht nur, welche Inhalte sichtbar sind, sondern auch, welche Gruppen gesellschaftlich repräsentiert oder marginalisiert werden. Insofern bilden algorithmische Verzerrungen eine Verbindung zwischen technischer Infrastruktur und gesellschaftlicher Macht.

1.3.2 Hegemoniale Männlichkeit

Die Analyse algorithmischer Moderationssysteme darf nicht bei Fragen technischer Effizienz oder regulatorischer Steuerbarkeit stehen bleiben. Digitale Technologien sind in gesellschaftliche Machtverhältnisse eingebettet und reproduzieren diese. Um zu verstehen, wie geschlechtsspezifische Ungleichheiten durch KI-gestützte Moderation stabilisiert werden, bedarf es theoretischer Konzepte, die Macht nicht nur als formale Struktur, sondern als kulturell verankerte und performativ erzeugte Hierarchie begreifen.

Eine Grundlegende Rolle spielt dabei das Patriarchat. Das Patriarchat beschreibt ein System männlicher Dominanz, das sich sowohl in politischer und wirtschaftlicher Macht äußert als auch in familiären und persönlichen Beziehungen fortsetzt. Es beruht auf der strukturellen Bevorzugung von Männern – etwa durch privilegierten Zugang zu Ressourcen, Entscheidungsgewalt und symbolischer Autorität – und stellt somit ein umfassendes Geflecht geschlechterbasierter Herrschaft dar (Connell, 2015: 90). Aufbauend auf dieser Struktur macht das Konzept der hegemonialen Männlichkeit sichtbar, wie sich männliche Dominanz kulturell verankert und fortschreibt.

Das Konzept der hegemonialen Männlichkeit nach Connell (2015), wurde ursprünglich nicht im Kontext digitaler Technologien entwickelt, bietet jedoch einen entscheidenden Zugang zur Analyse, da dadurch sichtbar gemacht werden kann, wie sich geschlechterbezogene Machtasymmetrien in technokratischen Diskursen und digitalen Infrastrukturen einschreiben. Für die vorliegende Arbeit wird es eingesetzt, um die Legitimation algorithmischer Moderation bei Meta als Teil eines maskulin codierten Machtregimes zu deuten. Es ermöglicht, die Reproduktion von Geschlechterverhältnissen innerhalb technologischer Infrastrukturen als Ausdruck sozialer Normierungen und Hierarchisierungen zu

analysieren. Im Unterschied zu frühen Positionen der feministischen Technologiekritik – etwa bei Sandra Harding (1986) oder Judy Wajcman (1991) – die primär auf die Unsichtbarkeit weiblicher Perspektiven, epistemische Ausschlüsse und den Mythos technologischer Objektivität fokussierten, lenkt Connell (2015: 106, 284) den Blick auf die aktive Konstruktion von Überlegenheit. Diese Perspektive soll in dieser Arbeit damit verknüpft werden, wie sie sich in der Sprache und Logik der Tech-Branche manifestiert.

Gerade im Diskurs global agierender Plattformunternehmen wie Meta lassen sich diskursive Strategien identifizieren, die zur Legitimation technologischer Macht beitragen. Connells Theorie hegemonialer Männlichkeit bietet dabei einen analytischen Zugang, um diese Legitimationsfiguren als Bestandteil eines strukturell verankerten Geschlechterverhältnisses zu deuten. Connells Theorie erlaubt es, diese Machtlogik als Teil eines strukturellen Geschlechterverhältnisses zu analysieren, das weit über Fragen von Repräsentation oder „gender bias“ hinausgeht.

Connell differenziert in ihrem Konzept zwischen hegemonialer Männlichkeit, die eine dominante und privilegierte Position in den Geschlechterverhältnissen einnimmt, und nicht-hegemonialen Männlichkeiten (Connell, 2015: 129–135). Letztere können entweder untergeordnet sein, die hegemoniale Männlichkeit stützen oder sie herausfordern.

Hervorzuheben ist dabei, dass „Männer“ und „Männlichkeit“ keine homogenen Kategorien darstellen. Die in Relation zueinanderstehenden unterschiedlichen Männlichkeiten werden als soziale Konstruktion angesehen und stehen in Abgrenzung zu Weiblichkeiten. Sowohl Weiblichkeiten als auch Männlichkeiten werden als eine dynamische Konstellation von Praktiken betrachtet, die sich historisch wandeln (Connell, 2015: 130). Innerhalb der Männlichkeiten stellt die Hegemoniale Männlichkeit das hetero- und homosoziale Zentrum der Macht dar (ebd.: 130).

Connell (2015: 130) beschreibt hegemoniale Männlichkeit wie folgt:

„Hegemoniale Männlichkeit kann man als jene Konfiguration geschlechtsbezogener Praxis definieren, welche die momentan akzeptierte Antwort auf das Legitimitätsproblem des Patriarchats verkörpert und die Dominanz der Männer sowie die Unterordnung der Frauen gewährleistet (oder gewährleisten soll).“

Alle anderen Männlichkeiten und Weiblichkeiten müssen sich in Relation zur hegemonialen Männlichkeit setzen.

Connell (2015: 131) unterscheidet zwischen verschiedenen nicht-hegemonialen Männlichkeiten. Erstens, der untergeordneten Männlichkeit, die in einer patriarchalen

Gesellschaft als illegitim angesehen wird. Es wird eine Strategie der Feminisierung und Vermischung mit weiblichen Praktiken angewendet, um eine Abwertung oder Ausgrenzung zu rechtfertigen.

Zweitens, der komplizenhaften Männlichkeit. Diese stützt und akzeptiert die hegemoniale Männlichkeit, weicht aber selbst in Teilen von ihr ab. Es werden beispielsweise im Berufs- oder Privatleben Kompromisse mit Frauen* eingegangen. Dennoch werden die Vorteile, die Männer in einer patriarchalen Gesellschaft durch die hegemoniale Männlichkeit haben, für sich genutzt (ebd.: 133).

Als dritte Form führt Connell (ebd.: 134) die marginalisierte Männlichkeit an. Diese hat grundsätzlich keine gesellschaftliche Autorität oder Macht, oft aufgrund ökonomischer und/oder ethnischer Hierarchisierung. Mit Blick auf die hegemoniale Männlichkeit steht sie am Rande der Gesellschaft und wird oft mit „Hypermaskulinität“ (z. B. hohe Gewaltbereitschaft) assoziiert.

Im digitalen Raum reproduzieren Plattformen wie Meta diese geschlechtlichen Machtverhältnisse, indem sie algorithmische Systeme einsetzen, die auf verzerrten Normen und Datensätzen basieren (Beck et al., 2019: 8). Diese Systeme operieren scheinbar neutral, sind jedoch strukturell geprägt von einem männlich konnotierten Ideal rationaler Kontrolle (Noble 2018: 1). In dieser Form fungieren Plattformen nicht nur als technische Infrastrukturen, sondern haben das Potenzial auch als kulturelle Verstärker hegemonialer Männlichkeitsmuster zu dienen. Dies soll in der Analyse sichtbar gemacht werden.

Vor diesem Hintergrund widmet sich das folgende Kapitel dem aktuellen Forschungsstand. Es stellt die zentralen wissenschaftlichen Debatten zu den Herausforderungen von Macht, Kontrolle und Verantwortung im Kontext algorithmischer Inhaltsmoderation dar. Im Fokus steht vor allem die Rolle von Plattformen wie Meta in der Konstruktion gesellschaftlicher Normen sowie auf die daraus resultierenden Implikationen für demokratische Teilhabe und Gleichstellung.

2 Forschungsstand

Die wachsende Bedeutung digitaler Plattformen wie Meta stellt Wissenschaft und Gesellschaft vor neue Herausforderungen im Hinblick auf Macht, Kontrolle und Verantwortung im digitalen Raum. Insbesondere die algorithmische Inhaltsmoderation rückt dabei ins Zentrum aktueller Debatten, da sie maßgeblich beeinflusst, welche Informationen Nutzer:innen wahrnehmen und wie gesellschaftliche Normen, Werte und Identitäten geformt werden.

Im Spannungsfeld zwischen privatwirtschaftlichen Interessen und gesellschaftlicher Verantwortung entstehen neuartige Machtverhältnisse, die sich weder durch klassische Marktmechanismen noch durch staatliche Regulierung vollständig erfassen lassen. Die Forschung richtet ihren Fokus daher zunehmend auf die strukturellen Bedingungen digitaler Macht – etwa auf intransparente Unternehmenspraktiken, algorithmische Verzerrungen (Bias) und die politische Instrumentalisierung von Moderationsrichtlinien.

Die Erforschung algorithmischer Inhaltsmoderation ist ein interdisziplinäres Feld, das zentrale Fragen zu Machtasymmetrien, digitaler Souveränität, algorithmischem Bias und demokratischer Kontrolle aufwirft. Am Beispiel von Meta wird deutlich, wie Plattformunternehmen zwischen ökonomischer Selbstregulierung und politischer Verantwortung agieren. Die im Folgenden dargestellten Studien und theoretischen Ansätze legen das Fundament für die anschließende Analyse und verdeutlichen, wie technologische Machtdynamiken und geschlechtsspezifische Diskriminierung im Kontext digitaler Plattformen untersucht werden.

2.1 Machtstrukturen und demokratische Herausforderungen bei Meta

An der Schnittstelle von privatwirtschaftlichen Interessen und gesellschaftlicher Verantwortung entstehen neuartige Machtverhältnisse, die sich weder durch klassische Marktmechanismen noch durch staatliche Regulierung vollständig erfassen lassen. Während die Europäische Union mit regulatorischen Maßnahmen wie dem Digital Services Act (DSA) und der KI-Verordnung auf mehr Transparenz und Rechenschaftspflicht abzielt, betont Meta seine Autonomie, etwa durch eigene Richtlinien zur Moderation von Inhalten. Die Plattform versteht ihre Eingriffe in digitale Diskurse als Beitrag zum Schutz der Meinungsfreiheit – zugleich aber auch als wirtschaftliche Notwendigkeit zur Gewährleistung von „Engagement“ und Plattformstabilität (Meta, n.d., siehe Anhang A).

Die Forschung hebt hervor, dass Plattformunternehmen wie Meta durch ihre zentrale Rolle in der digitalen Ökonomie neue Machtasymmetrien und Abhängigkeiten schaffen. So analysieren José van Dijck, Thomas Poell und Martijn de Waal (2018), wie Plattformen als infrastrukturelle Akteure grundlegende gesellschaftliche und politische Prozesse prägen und dabei als Gatekeeper über Zugänge zu Öffentlichkeit und Information entscheiden (van Dijck et al., 2018: 8, 13, 24). Christopher Grieser (2021: 1) argumentiert, dass Plattformen als infrastrukturzentrierte Märkte fungieren, in denen die Autonomie der Nutzer:innen durch mehrfache Abhängigkeiten von den Plattformbetreibern

eingeschränkt wird. Diese Struktur begünstigt eine Konzentration von Entscheidungsmacht und erschwert demokratische Kontrolle.

Robert Gorwa und Timothy Garton Ash (2020) kritisieren die mangelnde Transparenz in der Plattform-Governance und sehen darin eine strukturelle Schwächung demokratischer Prinzipien. Sie fordern, dass die Entscheidungsfindung über Inhalte und Moderationspraktiken nicht allein privatwirtschaftlichen Interessen überlassen bleiben darf, sondern stärker politisch und zivilgesellschaftlich kontrolliert werden muss (Gorwa & Garton Ash, 2020: 286-287).

Kate Klonick (2018) bezeichnet soziale Plattformen wie Meta sogar als „neue Regierungen“ der digitalen Öffentlichkeit, da sie maßgeblich darüber entscheiden, welche Inhalte Sichtbarkeit erhalten, und damit öffentliche Diskurse prägen (Klonick, 2018: 1607–1610). Die Debatte um demokratische Kontrolle, Transparenz und Rechenschaftspflicht bleibt daher zentral für das Verständnis der Rolle von Plattformunternehmen im digitalen Zeitalter.

Die zuvor beschriebenen allgemeinen Machtasymmetrien und Governance-Probleme digitaler Plattformen lassen sich am Beispiel von Meta besonders deutlich nachvollziehen. Im Folgenden werden die spezifischen Unternehmensstrukturen und die daraus resultierenden Herausforderungen für Transparenz und demokratische Kontrolle bei Meta näher betrachtet.

Eine zentrale Problematik der Forschung liegt in der strukturellen Verfasstheit von Meta selbst. Das Ranking Digital Rights (2022) dokumentiert erhebliche Defizite in der Transparenz und Menschenrechtsverpflichtung des Unternehmens. Meta erfüllte lediglich 46 % der Kriterien für digitale Verantwortung – insbesondere im Bereich der Inhaltsmoderation und Datennutzung. Die Plattform ist durch eine undurchsichtige Aktienstruktur geprägt, in der CEO Mark Zuckerberg allein 57 % der Stimmrechte kontrolliert – trotz massiver Zustimmung anderer Aktionäre zur Reform der Governance (Ranking Digital Rights, 2022). Diese strukturelle Machtkonzentration steht im klaren Gegensatz zu demokratischen Prinzipien von Teilhabe und Kontrolle.

Zudem lassen sich Defizite in der Umsetzung menschenrechtsbasierter Prüfverfahren feststellen. Die von Meta veröffentlichten Transparenzberichte erfassen nur einen Teil der Plattformaktivitäten (Facebook und Instagram) und weisen große Lücken bei Werbedaten und Moderationsentscheidungen auf. Gerade hochrangige politische Akteure erhalten teils Sonderbehandlungen – ein Phänomen, das durch die Enthüllungen von Whistleblowerin Frances Haugen belegt wurde (Ranking Digital Rights, 2022).

Im Frühjahr 2025 berichtete auch die ehemalige Facebook-Managerin Sarah Wynn-Williams als Whistleblowerin vor dem US-Senat über interne Abläufe und Entscheidungsstrukturen bei Meta. Ihren Aussagen zufolge kooperierte das Unternehmen in bestimmten Fällen mit staatlichen Behörden, unter anderem auch mit der chinesischen Regierung, und nahm Einfluss auf die Sichtbarkeit kritischer Inhalte. Darüber hinaus schilderte Wynn-Williams, dass Meta gezielt Werbeanzeigen auf Jugendliche ausrichtete, basierend auf deren emotionalem Zustand, und dass das Unternehmen rechtliche Schritte gegen die Veröffentlichung kritischer Berichte und Bücher einleitete (Jamali, 2025; Perez, 2025). Diese Berichte verweisen auf strukturelle Herausforderungen im Hinblick auf Transparenz, externe Kontrolle und die Rolle von Whistleblowern bei der Offenlegung unternehmensinterner Praktiken.

2.2 Geschlechterungleichheit und algorithmischer Bias

Die zuvor beschriebenen Defizite hinsichtlich Transparenz und demokratischer Kontrolle manifestieren sich nicht nur auf institutioneller Ebene, sondern haben auch unmittelbare Auswirkungen auf die Gestaltung und Wirkung algorithmischer Systeme. Im Folgenden wird daher dargestellt, wie sich strukturelle Ungleichheiten und mangelnde Diversität in der Tech-Industrie auf die Entstehung von algorithmischem Bias und geschlechterspezifischer Diskriminierung auswirken.

Eine internationale Analyse der UNESCO (2022) verdeutlicht, dass algorithmische Systeme weltweit dazu tendieren, bestehende gesellschaftliche Ungleichheiten zu verstärken, sofern Diversität und Inklusion in der Entwicklung und Implementierung nicht gezielt gefördert werden. Die Studie hebt hervor, dass insbesondere Frauen sowie durch historische Ungerechtigkeiten benachteiligte Gruppen von algorithmischem Bias betroffen sind, und fordert daher eine stärkere Berücksichtigung vielfältiger Perspektiven in allen Phasen der KI-Entwicklung (UNESCO, 2022: 14). Dies unterstreicht, dass geschlechtsspezifische Diskriminierung durch Algorithmen nicht lediglich ein unternehmensspezifisches, sondern ein globales Phänomen darstellt, das umfassende politische und gesellschaftliche Maßnahmen erfordert.

Ein weiterer zentraler Forschungsaspekt betrifft die Zusammensetzung der Entwickler- und Review-Teams bei Meta. In der globalen Tech-Industrie ist der Anteil weiblicher Fachkräfte weiterhin niedrig: In den USA liegt er bei lediglich 26% (AnitaB.org, 2021), in der EU bei 22% (McKinsey & Company, 2021). Bei Meta selbst beträgt der Frauenanteil insgesamt 37,1%, im Tech-Bereich sogar nur 25,8% (Meta, 2022a, siehe Anhang A).

Diese Unterrepräsentation spiegelt sich auch in der Gestaltung und Implementierung algorithmischer Systeme wider, da Perspektiven marginalisierter Gruppen weniger Berücksichtigung finden, was sich wiederum in den Ergebnissen der Moderation niederschlagen kann.

Da algorithmische Entscheidungen häufig auf maschinellem Lernen basieren, sind sie unmittelbar von menschlichen Bewertungen und Vorgaben abhängig, die oftmals nicht die gesellschaftliche Vielfalt widerspiegeln. Auf diese Weise werden soziale Ordnungen – beispielsweise patriarchale, rassistische oder kapitalistische Strukturen – im digitalen Raum reproduziert. Auch wenn Unternehmen wie Meta Diversitätsziele formulieren, zeigt sich, dass ein grundlegender Umbau interner Machtverhältnisse notwendig wäre, um strukturelle Ungleichheiten zu verringern.

In diesem Zusammenhang stellen sogenannte algorithmische Bias eine besondere Herausforderung dar, da sie auf eine systematische Verzerrung algorithmischer Entscheidungen hinweisen. Empirische Studien belegen, dass selbst bei vermeintlich neutralen Eingabeparametern bestimmte Gruppen – etwa Frauen oder People of Color – unterschiedlich behandelt werden. Muhammad et al. (2019: 23) zeigen beispielsweise, dass Frauen durch Finanzierungs- und Vorhersageeffekte bei der Anzeigenschaltung seltener gut bezahlte Stellenanzeigen angezeigt bekommen. Auch Algorithm Watch (Kayser-Bril, 2021) bestätigt diese Beobachtungen und stellt fest, dass Facebook-Algorithmen Bewerbungen von Frauen systematisch schlechter bewerten als die von Männern.

Diese Befunde stehen in engem Zusammenhang mit der Datengrundlage, auf der die Systeme trainiert werden. Magdalena Freudenschuss und Sophie Mützel et al. (2014: 10; 2018: 2) betonen, dass Daten und Algorithmen keine objektive Realität abbilden, sondern sozial konstruiert und durch historische Machtverhältnisse geprägt sind. Dies zeigt sich auch in Experimenten mit Chatbots, die bei unregulierter Datenaufnahme sexistische und rassistische Aussagen übernehmen (Beck et al., 2019: 6). Die Forschung fordert daher verstärkte politische Kontrolle über algorithmische Systeme sowie umfassende Transparenzmechanismen (Muhammad et al., 2019; Algorithm Watch, 2021).

Die inhaltliche Steuerung der Plattform durch Mark Zuckerberg ist dabei nicht nur ein wirtschaftlicher, sondern auch ein politischer Faktor. In der aktuellen US-Legislaturperiode positionierte sich Meta offener gegenüber konservativen Strömungen und kündigte an, Faktenprüfungsprogramme durch Community-Notizen zu ersetzen (Zuckerberg, 2025: Abs. 8, siehe Anhang B). Parallel dazu wurden bei Meta und anderen Tech-Unternehmen Diversity-Initiativen zurückgefahren. Natalie Sherman (2025) berichtet, dass

Mitarbeitende und Branchenbeobachter:innen diese Entwicklung kritisch sehen und befürchten, dass der Abbau von Diversity-Programmen negative Auswirkungen auf die Vielfalt und Inklusion in den Unternehmen haben könnte. Vor diesem Hintergrund wird deutlich, dass männlich dominierte Strukturen und fehlende Diversität in der Tech-Branche – wie auch Caroline Criado Perez (2019) hervorhebt – maßgeblich dazu beitragen, dass Diskriminierung und Benachteiligung von Frauen* fortbestehen.

Der in diesem Kapitel dargestellte Forschungsstand verdeutlicht, dass algorithmische Inhaltsmoderation weit mehr ist als ein technischer Prozess: Sie ist eng mit gesellschaftlichen Machtverhältnissen verwoben und beeinflusst öffentliche Diskurse, politische Entscheidungsprozesse sowie individuelle Handlungsspielräume. Insbesondere die Zusammensetzung von Entwickler- und Entscheidungsteams spielt dabei eine zentrale Rolle, da die systematische Unterrepräsentation bestimmter Perspektiven – etwa von Frauen oder anderen marginalisierten Gruppen – dazu führen kann, dass gesellschaftliche Vorurteile und bestehende Machtstrukturen in digitalen Produkten fortgeschrieben werden. Daraus ergibt sich die Notwendigkeit, Diversität nicht nur als moralisches Ziel, sondern als grundlegende Voraussetzung für die Entwicklung fairer und inklusiver Technologien zu begreifen. Wie Unternehmen wie Meta strukturelle Veränderungen in ihren Teams und Prozessen wirksam gestalten oder durch geeignete Regulierungen dazu angehalten werden können, stellt eine der zentralen Herausforderungen für die Zukunft dar.

2.3 EU-Regulierungen für KI & Plattformen

Die zuvor skizzierten Dynamiken zeigen, dass digitale Plattformen nicht nur technische, sondern auch gesellschaftliche Wirkungen entfalten. Daraus ergibt sich ein wachsender Bedarf nach politischer Steuerung. Im Anschluss an die Darstellung des Regulierungsbedarfs werden die wichtigsten Regulierungsansätze der EU näher beleuchtet.

2.3.1 Regulierungsbedarf im digitalen Raum

Digitale Plattformen übernehmen zunehmend zentrale Funktionen in Gesellschaft, Wirtschaft und Kommunikation. Gleichzeitig ist ihre Rolle nicht umfassend gesetzlich geregelt. Insbesondere im Hinblick auf Transparenz, Rechenschaft und den Umgang mit Nutzer:innen entstehen dadurch Spannungsfelder, die politische Aufmerksamkeit erfordern. Die Auswirkungen algorithmischer Verzerrungen sind besonders gravierend, da Plattformen wie Meta den Zugang von umfangreichen Datenströmen bestimmen und eigenständig Regeln für die Inhaltsmoderation festlegen können (Gillespie, 2018: 121). Anders als bei klassischen redaktionellen Entscheidungen erfolgt diese Regulierung zunehmend

automatisiert, algorithmisch und in Echtzeit, bei gleichzeitig fehlender Transparenz. Dadurch können sich Verzerrungen unbemerkt und in großem Maßstab auswirken, etwa durch systematische Benachteiligung bestimmter Gruppen oder durch gezielte Beeinflussung öffentlicher Meinungen. Der Fall Cambridge Analytica hat gezeigt, wie algorithmisch gesteuerte Datenverarbeitung in sozialen Netzwerken demokratische Prozesse untergraben kann. Dabei wurden ohne Wissen der Nutzer:innen Millionen Facebook-Profile ausgewertet, um gezielt personalisierte politische Werbung zu schalten und so versucht Wahlentscheidungen zu beeinflussen.

Die Europäische Union verfolgt vor diesem Hintergrund eine ambitionierte Digitalpolitik, insbesondere bei der Regulierung digitaler Plattformen. Ihre Digitalstrategie und Rechtsinstrumente zielen darauf ab, die Machtasymmetrien im digitalen Raum einzudämmen und die digitale Souveränität Europas zu stärken – etwa durch Maßnahmen wie den Digital Markets Act (DMA). Neben dem DMA sind der Digital Services Act (DSA) und die KI-Verordnung zentrale Bausteine, um algorithmische Transparenz, Rechenschaftspflicht und menschenrechtsorientierte Moderationspraktiken durchzusetzen. Das Leitprinzip der digitalen Souveränität verfolgt einen eigenständigen europäischen Weg, der sich zwischen marktbasierten US-amerikanischen und staatsgesteuerten chinesischen Modellen positioniert, wobei Initiativen wie GAIA-X technologische Unabhängigkeit und faire Wettbewerbsbedingungen fördern (Köhler & Schmid, 2022; Pohle et al., 2023: 5–7, 14). Zugleich wird digitale Souveränität als ein mehrdimensionaler Prozess verstanden: Durch die NIS-Richtlinie, das Cybersicherheitsgesetz, DSA, DMA sowie den transatlantischen Handels- und Technologierat (TTC) stärkt die EU ihren Einfluss auf globale Standards. Dabei bestehen weiterhin Herausforderungen – etwa hinsichtlich der Definition von Gatekeepern und Datenschutzproblemen – sodass eine enge Koordination insbesondere mit den USA empfohlen wird, um Europas digitale Souveränität zu sichern und weltweit normative Standards zu setzen (Bendiek & Stürzer, 2022: 1–8). In den vergangenen Jahren wurden wiederholt maßgebliche Sanktionen gegen Meta verhängt. So musste das Unternehmen 2023 eine Rekordstrafe von 1,2 Milliarden Euro zahlen, nachdem es personenbezogene Daten europäischer Nutzer:innen unzulässig in die USA übermittelt hatte – ein Verstoß gegen die Datenschutz-Grundverordnung (DSGVO). 2024 folgte eine weitere Geldstrafe in Höhe von 797,7 Millionen Euro wegen Missbrauchs marktbeherrschender Stellung. Für 2025 werden laut Medienberichten erneut Sanktionen erwartet, da Meta gegen zentrale Bestimmungen des DMA verstoßen haben soll. Die Strafen könnten bis zu 10 % des weltweiten Jahresumsatzes betragen.

In ihrer Digitalstrategie betont die Europäische Kommission (2022) die Bedeutung technologischer Eigenständigkeit und die Notwendigkeit, europäische Werte durch den Aufbau eigener digitaler Infrastrukturen und Standards zu sichern. Die Strategie zielt explizit darauf ab, Europas Handlungsfähigkeit im digitalen Wandel zu gewährleisten:

„Europas digitale Souveränität erfordert digitale Infrastrukturen, Produkte und Dienste, die geeignet sind, die Sicherheit, die strategischen Vermögenswerte und die Interessen Europas im Einklang mit den europäischen Werten und Anforderungen zu wahren.“ (Europäische Kommission 2022: 5)

Die bisherigen Maßnahmen zeigen, dass die EU regulatorisch auf Machtungleichgewichte im digitalen Raum reagiert. Wie die genannten Beispiele zeigen, wird gleichzeitig deutlich, dass rechtliche Schritte allein nicht ausreichen, um strukturelle Abhängigkeiten und intransparente Praktiken dauerhaft zu begrenzen. Der wachsende Einfluss weniger Technologiekonzerne unterstreicht die Notwendigkeit, regulatorische Instrumente weiterzuentwickeln und an die Dynamiken digitaler Infrastrukturen anzupassen.

2.3.2 Datenschutz-Grundverordnung (DSGVO)

Die Datenschutz-Grundverordnung (DSGVO), die seit Mai 2018 in der Europäischen Union gilt, bildet einen der zentralen Pfeiler des europäischen Datenschutzrechts. Sie wurde eingeführt, um die Persönlichkeitsrechte von Bürger:innen im digitalen Raum zu stärken und europaweit zu vereinheitlichen (Wolford, n. d.). Die Verordnung zählt zu den strengsten Datenschutzgesetzen weltweit und gilt nicht nur für Organisationen mit Sitz in der EU, sondern für alle, die personenbezogene Daten von EU-Bürger:innen verarbeiten – unabhängig vom Standort des Unternehmens.

Dies betrifft insbesondere globale Technologieunternehmen wie Meta, deren Geschäftsmodelle maßgeblich auf der Erhebung, Analyse und Weiterverwertung von Nutzerdaten beruhen. Verstöße gegen die DSGVO können mit Geldstrafen von bis zu 4 % des weltweiten Jahresumsatzes geahndet werden (Art. 83 DSGVO). In der Vergangenheit wurden bereits Rekordsummen gegen Meta verhängt – unter anderem wegen unzulässiger Datenübermittlungen in die USA (siehe Abschnitt 2.3.1).

Die DSGVO basiert auf sieben zentralen Grundsätzen:

1. **Rechtmäßigkeit, Fairness und Transparenz** – Die Datenverarbeitung muss auf einer gültigen Rechtsgrundlage beruhen und für Betroffene nachvollziehbar sein.
2. **Zweckbindung** – Daten dürfen nur für festgelegte, eindeutige und legitime Zwecke verarbeitet werden.

3. **Datenminimierung** – Nur so viele Daten wie nötig dürfen erhoben werden.
4. **Richtigkeit** – Es muss gewährleistet sein, dass personenbezogene Daten korrekt und aktuell sind.
5. **Speicherbegrenzung** – Daten dürfen nur so lange gespeichert werden, wie es für die Zwecke notwendig ist.
6. **Integrität und Vertraulichkeit** – Sicherheit der Daten durch technische und organisatorische Maßnahmen.
7. **Rechenschaftspflicht** – Verantwortliche müssen die Einhaltung aller Grundsätze nachweisen können (Wolford, n.d.)

Zur Durchsetzung der Verordnung wurde der Europäische Datenschutzausschuss (EDSA) eingerichtet. Er koordiniert die Zusammenarbeit der nationalen Datenschutzbehörden im Europäischen Wirtschaftsraum und stellt die einheitliche Anwendung der DSGVO sicher (Europäischer Datenschutzausschuss, n.d.).

Die Datenschutz-Grundverordnung (DSGVO) der Europäischen Union gilt international als ein besonders weitreichendes und einflussreiches Regelwerk für den Datenschutz. In der wissenschaftlichen Diskussion wird ihre normative Bedeutung vielfach hervorgehoben. Colin J. Bennett und Charles D. Raab (2020) zeigen jedoch, dass die konkrete Auswahl und Wirksamkeit von Datenschutzinstrumenten innerhalb der EU stark von institutionellen, politischen und kulturellen Unterschieden zwischen den Mitgliedstaaten geprägt ist (vgl. Bennett & Raab, 2020: 448, 462). Diese Unterschiede beeinflussen maßgeblich, wie Datenschutz in der Praxis umgesetzt und durchgesetzt wird.

Trotz dieser Herausforderungen gilt die DSGVO als zentrales Instrument zur Stärkung digitaler Souveränität der EU – sie verschiebt die Kontrolle über personenbezogene Daten zumindest teilweise von Unternehmen zurück in die Hände der Nutzer:innen.

2.3.3 Der Digital Markets Act (DMA)

Der Digital Markets Act (DMA) stellt einen zentralen Baustein der EU-Digitalstrategie dar und adressiert die wachsende Marktkonzentration und Machtstellung großer Plattformunternehmen. Ziel des DMA ist es, faire Wettbewerbsbedingungen im digitalen Binnenmarkt zu schaffen und die strukturelle Dominanz sogenannter „Gatekeeper-Plattformen“ zu regulieren (Europäische Kommission, n.d.-b).

Gatekeeper sind Plattformen, die als Vermittler zwischen Nutzer:innen und Anbietenden auftreten und dabei den Zugang zu bestimmten digitalen Märkten kontrollieren. Zu diesen gehören etwa Google, Amazon oder Meta. Der DMA definiert klare Kriterien, um solche

Gatekeeper zu identifizieren, und verpflichtet sie zu bestimmten Verhaltensnormen. Dazu zählen etwa:

- das Verbot, eigene Dienste systematisch zu bevorzugen,
- die Pflicht zur Interoperabilität mit anderen Diensten,
- sowie Maßnahmen zur Verhinderung von Datenmonopolen (Europäisches Parlament und Rat der Europäischen Union, 2022a, siehe Anhang A).

Ein zentrales Merkmal des DMA ist seine präventive Ausrichtung. Statt lediglich auf bereits entstandene Wettbewerbsverzerrungen zu reagieren, soll der DMA marktbeherrschenden Unternehmen frühzeitig Grenzen setzen. Rupprecht Podszun (2023: 2) weist in seiner Analyse des Gesetzes darauf hin, dass klassische Wettbewerbsinstrumente im digitalen Raum vielfach nicht ausreichen, um den neuen Machtverhältnissen gerecht zu werden. Stattdessen bedarf es gezielter Plattformregulierung, um Marktversagen proaktiv zu verhindern.

Podszun (2023) betont zudem die Notwendigkeit, die Flexibilität des DMA zu gewährleisten. Angesichts sich schnell wandelnder digitaler Märkte muss der Rechtsrahmen anpassungsfähig bleiben, um seine Wirksamkeit nicht zu verlieren. Gleichzeitig stellt die Umsetzung der Vorgaben eine Herausforderung dar: Nationale Wettbewerbsbehörden verfügen oft nicht über ausreichende Ressourcen oder technische Expertise, um die komplexen Geschäftsmodelle der Plattformkonzerne effektiv zu kontrollieren. Hinzu kommt die politische Debatte über eine Balance zwischen Regulierung und Innovationsförderung.

Der DMA zeigt exemplarisch, wie sich der Umgang mit globalen Plattformkonzernen in Europa verändert hat: Weg vom rückblickenden Wettbewerbsrecht hin zu zukunftsorientierter Plattformpolitik (Podszun 2023: 11). Er steht damit sinnbildlich für die strategische Neuausrichtung europäischer Digitalpolitik.

2.3.4 Der Digital Services Act (DSA)

Der Digital Services Act (DSA) ist ein weiteres zentrales Regulierungsvorhaben der Europäischen Union und ergänzt den Digital Markets Act in seiner Funktion, den digitalen Raum gerechter und sicherer zu gestalten. Während der DMA vor allem die strukturelle Marktmacht adressiert, zielt der DSA primär auf die Regulierung von Online-Inhalten und den Schutz der Nutzer:innenrechte ab (Europäische Kommission, n.d.-c, siehe Anhang A).

Der DSA verfolgt das Ziel, einheitliche Vorschriften für alle digitalen Dienste, insbesondere für sehr große Online-Plattformen (VLOPs), zu schaffen. Zu seinen zentralen Elementen gehören:

- die Verpflichtung zur Entfernung illegaler Inhalte,
- die Transparenz von Empfehlungsalgorithmen,
- die Regulierung personalisierter Werbung,
- sowie die Stärkung der Rechte von Nutzer:innen bei Inhaltsentscheidungen (Europäische Kommission, n.d.-c, siehe Anhang A).

Besonders betont wird dabei die Bedeutung der Nutzungssicherheit und des Schutzes der Grundrechte im digitalen Raum. So sollen Nutzer:innen unter anderem verständlich über die Funktionsweise von algorithmischer Inhaltsmoderation informiert werden und effektive Beschwerdemechanismen nutzen können.

Wissenschaftlich wird der DSA vielfach als ein Paradigmenwechsel hin zu einer stärker demokratisch kontrollierten Plattformordnung eingeordnet. Ben Wagner et al. (2022) heben hervor, dass der DSA nicht nur eine Reaktion staatlicher Institutionen gegenüber Tech-Konzernen darstellt, sondern auch auf eine strukturelle Neugestaltung der Verantwortlichkeiten im digitalen Raum zielt. Sie bewerten das Gesetzespaket als Versuch, Plattformen stärker in die Pflicht zu nehmen, ohne deren Innovationspotenzial zu ersticken.

Die praktische Durchsetzung zentraler Aspekte des DSA bleibt herausfordernd. Kritik besteht insbesondere darin, dass die technische Komplexität und globale Reichweite von Plattformen wie Meta nationale Regulierungsbehörden oft überfordern. Zudem zeigt sich, dass viele Maßnahmen stark von der Kooperationsbereitschaft der Plattformen selbst abhängig sind (vgl. Hulkó et al., 2025: 8).

Insgesamt stellt der DSA einen wichtigen Schritt zur Stärkung der digitalen Souveränität und der Grundrechte im digitalen Raum dar. Er reagiert auf die zunehmende gesellschaftliche Bedeutung algorithmischer Entscheidungssysteme und institutioneller Verantwortungslosigkeit, die insbesondere dort problematisch wird, wo Moderation intransparent, automatisiert und ohne klare Rechenschaftspflicht erfolgt.

2.3.5 Der EU AI-Act

Mit dem EU AI-Act (Verordnung über Künstliche Intelligenz) legt die Europäische Union den weltweit ersten umfassenden Rechtsrahmen zur Regulierung von Künstlicher Intelligenz vor. Der Gesetzesentwurf wurde 2021 von der Europäischen Kommission vorgelegt

und basiert auf einem risikobasierten Ansatz, der KI-Systeme je nach ihrem Gefährdungspotenzial in verschiedene Risikoklassen einteilt (Europäisches Parlament, 2023).

Im Zentrum der Verordnung steht das Ziel, KI-Systeme sicher, transparent, diskriminierungsfrei und menschenzentriert zu gestalten. Die Anwendungen wurden von der Europäischen Kommission (2021: 15) in folgende Risikostufen unterteilt:

- Unannehmbares Risiko: etwa soziale Bewertungssysteme (Social Scoring) durch staatliche Akteure oder KI-Systeme zur bewussten Manipulation des Verhaltens, die Menschen Schaden zufügen könnten.
- Hochrisikosysteme: z. B. biometrische Überwachung, automatisierte Bewerbungsprozesse, Scoring bei Kreditvergabe. Diese unterliegen besonders strengen Anforderungen hinsichtlich Transparenz, Datenqualität, menschlicher Aufsicht und Dokumentationspflichten.
- Geringes Risiko: z. B. Chatbots, die als solche gekennzeichnet werden müssen.
- Minimales Risiko: z. B. KI-basierte Spamfilter oder Spiele-KI, für die keine besonderen Auflagen bestehen (Europäische Kommission, 2021).

Besonders relevant im Kontext dieser Arbeit ist, dass der AI-Act gezielt auf die transparente Gestaltung algorithmischer Entscheidungssysteme abzielt – etwa in den Bereichen automatisierter Inhaltsmoderation, algorithmischer Empfehlungssysteme oder personalisierter Werbung. Unternehmen wie Meta müssen künftig offenlegen, auf welchen Daten ihre Systeme trainiert wurden, welche Logiken diesen zugrunde liegen und wie potenziell diskriminierende Wirkungen minimiert werden sollen.

Zugleich betont die EU in ihrer politischen Kommunikation, dass technologische Innovation durch Regulierung nicht behindert, sondern verantwortungsvoll gestaltet werden soll. Die EU verfolgt mit dem AI-Act das Ziel, einen globalen Standard für menschenzentrierte KI zu etablieren, der Grundrechte achtet und wirtschaftliche Wettbewerbsfähigkeit nicht ausschließt (Europäisches Parlament, 2023).

Kritische Stimmen bemängeln jedoch, dass viele technische Details offen oder interpretationsbedürftig bleiben und die Durchsetzung regulatorischer Standards stark von der Ausstattung nationaler Behörden abhängt. Zudem wird diskutiert, inwiefern Unternehmen wie Meta die Anforderungen durch vage Compliance-Strategien oder juristische Spielräume umgehen könnten (Goanta et al., 2022).

Dennoch stellt der AI-Act einen zentralen Meilenstein europäischer Digitalpolitik dar. Er verknüpft normative Zielsetzungen (Grundrechtsschutz, Diskriminierungsfreiheit,

Transparenz) mit konkreten Anforderungen an technische Systeme und markiert somit den Versuch, eine ethisch fundierte Gestaltung algorithmischer Macht zu institutionalisieren.

2.3.6 Ergänzende Regulierungen

Neben den zentralen Regelwerken wie dem DMA, DSA, AI-Act und der DSGVO existieren weitere regulatorische Instrumente auf EU-Ebene, die der zunehmenden Macht digitaler Plattformen und Technologieunternehmen entgegenwirken sollen.

Ein wesentlicher Bestandteil bleibt dabei das traditionelle EU-Kartell- und Wettbewerbsrecht, das den freien Wettbewerb sichern soll. Dieses umfasst insbesondere das Verbot illegaler Absprachen und Informationsaustausche zwischen Unternehmen, die dazu dienen könnten, Unsicherheiten auf dem Markt zu minimieren – etwa im Hinblick auf Preise, Kapazitäten oder Geschäftsstrategien (Europäische Union, 2025). Zusätzlich regelt es, welche Praktiken als Missbrauch einer marktbeherrschenden Stellung gelten, darunter etwa die Ausbeutung von Kund:innen oder die Verdrängung von Wettbewerbern.

Eine wichtige Ergänzung zur DSGVO stellt die Verordnung über Privatsphäre und elektronische Kommunikation (auch bekannt als *ePrivacy-Verordnung*) dar. Sie modernisiert bestehende Datenschutzregelungen und weitet den Schutzbereich auf neue Kommunikationsformen aus. Neben dem Schutz von Inhalten umfasst sie auch Metadaten elektronischer Kommunikation. Nutzer:innenrechte werden durch die verpflichtende Einwilligung zur Datenverarbeitung gestärkt, und Anbieter müssen technische Maßnahmen zur Sicherstellung von Sicherheit und Vertraulichkeit ergreifen (Europäisches Parlament und Rat der Europäischen Union, 2022b).

Ein weiteres zentrales Instrument ist die sogenannte NIS-2-Richtlinie (Richtlinie (EU) 2022/2555), die auf ein hohes gemeinsames Cybersicherheitsniveau in der EU abzielt. Sie verpflichtet Mitgliedstaaten zur Entwicklung nationaler Cyberstrategien und betroffene Einrichtungen zur Meldung sicherheitsrelevanter Vorfälle. Zudem fördert sie den Informationsaustausch zwischen relevanten Akteuren und erweitert die Handlungsbefugnisse nationaler Behörden im Bereich der Cybersicherheit (Europäisches Parlament und Rat der Europäischen Union, 2022c).

Auch die Initiative Gaia-X ist ein zentraler Baustein der europäischen Digitalstrategie. Ziel ist es, eine föderierte und vertrauenswürdige Dateninfrastruktur zu etablieren, welche EU-Standards bei Datenschutz und Sicherheit einhält. Im Gegensatz zu kommerziellen US-Anbietern soll Gaia-X ein offenes Ökosystem bieten, in dem verschiedene Cloud-

Anbieter interoperabel agieren können. Auf diese Weise soll Europas digitale Souveränität langfristig gestärkt werden (Europäische Organisation für Gaia-X, 2023).

Insgesamt ergänzen diese Regulierungen die großen Gesetzeswerke der EU und verdeutlichen, dass der Schutz der digitalen Selbstbestimmung und der freien Märkte auf einer Vielzahl von rechtlichen Maßnahmen beruht. Sie zeigen zugleich den Anspruch der EU, digitale Räume umfassend, nachhaltig und wertebasiert zu gestalten.

Das erste Kapitel hat die theoretischen Grundlagen zur Analyse digitaler Plattformen gelegt. Im Zentrum stand, wie sich Machtverhältnisse im digitalen Raum – insbesondere im Hinblick auf digitale Souveränität und geschlechtsspezifische Ungleichheiten – konstituieren und stabilisieren. Dabei wurde herausgearbeitet, dass Unternehmen wie Meta nicht nur ökonomische, sondern auch politische und kulturelle Steuerungsmacht ausüben, etwa durch algorithmische Moderationssysteme, die strukturelle Verzerrungen reproduzieren können. Plattformen übernehmen zunehmend zentrale Infrastrukturfunktionen, ohne dabei demokratischer Kontrolle zu unterliegen. Gleichzeitig wurde deutlich, dass die EU mit Instrumenten wie dem DMA und der DSGVO versucht, diesen Machtasymmetrien regulatorisch zu begegnen.

2.4 Zwischenfazit

Der Forschungsstand verdeutlicht, dass digitale Plattformen wie Meta zu den zentralen Machtakteuren im digitalen Raum zählen. Ihre algorithmische Inhaltsmoderation beeinflusst maßgeblich, welche Informationen zugänglich sind, wie gesellschaftliche Normen entstehen und wie demokratische Prozesse gestaltet werden. Die Forschung zeigt, dass die Konzentration von Macht bei privaten Plattformunternehmen durch intransparente Strukturen, fehlende demokratische Kontrolle und algorithmische Verzerrungen weiter verstärkt wird. Insbesondere die Zusammensetzung von Entwickler- und Moderationsteams sowie strukturelle Ungleichheiten in der Tech-Industrie tragen dazu bei, dass gesellschaftliche Vorurteile und Diskriminierung – vor allem gegenüber Frauen und marginalisierten Gruppen – in digitalen Systemen fortbestehen.

Europäische Regulierungsinitiativen wie der Digital Services Act und die KI-Verordnung setzen wichtige Impulse für mehr Transparenz, Rechenschaftspflicht und den Schutz demokratischer Werte im digitalen Raum. Dennoch bleibt die Machtverteilung weiterhin zugunsten global agierender Unternehmen verschoben, was die Effektivität staatlicher Regulierung einschränkt.

Zukünftige Forschung und politische Maßnahmen sollten daher verstärkt darauf abzielen, technologische Innovation, demokratische Kontrolle und gesellschaftliche Teilhabe im digitalen Zeitalter miteinander zu verbinden. Besonders herausfordernd bleibt die Frage, wie Diversität und Inklusion auf allen Ebenen der Plattformgestaltung und -regulierung nachhaltig verankert werden können. Hier besteht weiterhin großer Handlungsbedarf für Wissenschaft, Politik und Zivilgesellschaft, um strukturelle Ungleichheiten abzubauen und digitale Räume gerechter zu gestalten.

Forschungsfrage

Damit lauten die für diese Arbeit zentralen Forschungsfragen:

Inwiefern erschweren KI-gestützte Moderationssysteme von Meta die Umsetzung von digitaler Souveränität und Geschlechtergleichheit innerhalb der EU?

Unterfragen:

1. Tragen algorithmische Bias in Metas Inhaltsmoderation zur Verstärkung von Geschlechterungleichheiten bei?
2. Inwiefern untergräbt Metas Kontrolle über algorithmische Entscheidungen die digitale Souveränität Europas?
3. Welche regulatorischen Maßnahmen der EU adressieren diese Herausforderungen, und wie wirksam sind sie?

In dieser Arbeit sind insbesondere die Konzepte der digitalen Souveränität und der algorithmischen Verzerrung, insbesondere in Bezug auf Geschlechtergleichheiten von Bedeutung. Diese zwei Felder bilden das Fundament der Analyse und werden im folgenden Kapitel konzeptionell entfaltet und miteinander in Beziehung gesetzt. Die theoretische Rahmung schafft so die Grundlage für die anschließende Anwendung der strukturellen Machtanalyse nach Foucault (1976) und der Critical Frame Analysis (CFA) nach Mieke Verloo (2005), die im Methodenkapitel näher erläutert werden. Diese Methodik soll die diskursiven und machtsstrukturellen Dimensionen sichtbar machen. In Kapitel 4 werden mithilfe der strukturellen Machtanalyse die infrastrukturellen und normativen Machtverhältnisse auf den Plattformen von Meta untersucht. Kapitel 5 ergänzt diese Perspektive durch eine Analyse der öffentlichen Kommunikation Metas mittels der Critical Frame Analyse. Abschließend werden die Ergebnisse beider Analysekapitel zusammengeführt und im Hinblick auf die Forschungsfrage diskutiert.

3 Methodologie

Die theoretische Grundlage und der Forschungsstand dieser Arbeit haben gezeigt, dass algorithmische Moderation nicht nur als technologische Innovation verstanden werden darf, sondern tief in machtvollen Diskursen eingebettet ist, die digitale Souveränität und Geschlechterverhältnisse strukturieren. Um diesen komplexen Zusammenhängen empirisch nachzugehen, bedarf es einer methodischen Herangehensweise, die sowohl diskursive Legitimierungsstrategien als auch strukturelle Machtverhältnisse systematisch sichtbar machen kann.

In diesem Kapitel wird daher erläutert, wie die Kombination einer strukturellen Machtanalyse nach Foucault mit einer Critical Frame Analyse nach Verloo die Untersuchung der Forschungsfrage ermöglicht. Dabei stehen die diskursiven Selbstinszenierungen von Meta im Zentrum der Analyse, deren sprachliche und strategische Muster entlang zentraler thematischer Frames (z. B. Neutralität, Effizienz, Verantwortung) herausgearbeitet und auf ihre politischen, sozialen und geschlechterbezogenen Implikationen hin befragt werden.

Die strukturelle Machtanalyse nach Michel Foucault bildet in dieser Arbeit die erste methodische Säule zur Untersuchung der digitalen Machtverhältnisse, die durch Metas algorithmische Inhaltsmoderation hervorgebracht und legitimiert werden. Foucaults Machtbegriff versteht Macht nicht als Besitz einer Instanz, sondern als Netz relationaler Kräfte, das gesellschaftliche Strukturen, Institutionen und Technologien durchzieht (Foucault, 1976: 38–39). In *Überwachen und Strafen* analysiert Foucault, wie Macht durch Wissen, Normen und technologische Dispositive wirkt – ein Konzept, das algorithmische Systeme als subtile Machtinstrumente entschlüsselbar macht. Einzelne Akteur:innen wie Meta sind dabei nicht als autonome Urheber:innen von Macht, sondern als deren Knotenpunkte innerhalb umfassender gesellschaftlicher Dispositive.

3.1 Theoretische Fundierung

In dem zuvor erwähnten Werk *Überwachen und Strafen* (Foucault, 1976: 221–253) betrachtet Foucault die Disziplinarmacht als eine historische Form der Machtausübung, die durch Überwachung, Normierung und Sanktionierung das Verhalten einzelner Subjekte steuert. In späteren Schriften, insbesondere in *Der Wille zum Wissen* (Foucault, 1986: 161–189), führt Foucault das Konzept der Biopolitik ein, das sich auf die Regulierung ganzer Bevölkerungen durch statistische Steuerungsmechanismen, Normensetzung und institutionelle Planung bezieht.

Im Kontext dieser Arbeit werden diese Konzepte auf die Plattformstrukturen von Meta übertragen. Plattformen wie Meta üben nicht nur ökonomische Macht aus, sondern strukturieren auch gesellschaftliche Normen durch algorithmische Systeme, die darüber entscheiden, welche Inhalte sichtbar, gelöscht oder priorisiert werden (Beer, 2016; Zuboff, 2018). Aus analytischer Perspektive wird in dieser Arbeit argumentiert, dass diese Systeme als moderne biopolitische Dispositive zu begreifen sind, da sie Verhalten, Wissen und Öffentlichkeit über scheinbar objektive technische Infrastrukturen regulieren. Der Begriff „Dispositiv“ bezeichnet nach Foucault ein strategisches Netz heterogener Elemente – wie Diskurse, Institutionen, Gesetze und technische Einrichtungen –, das zur Ausübung von Macht eingesetzt wird (Foucault, 1978a: 119–122).

3.1.1 Methodische Anwendung

Ziel der strukturellen Machtanalyse ist es, die versteckten Machtmechanismen innerhalb der Plattformarchitektur Metas sichtbar zu machen – insbesondere im Hinblick auf digitale Souveränität und Geschlechterverhältnisse. Dafür wurden zentrale Konzepte Foucaults operationalisiert und auf ausgewählte Diskursquellen angewendet.

Dispositive

Dispositive erzeugen und steuern Wissen, indem sie bestimmte Praktiken, Wahrnehmungen und Verhaltensweisen ermöglichen oder ausschließen. Foucault (1978a: 119-123) analysierte Dispositive insbesondere im Hinblick auf Machtverteilung und als Instrumente der Herrschaft, etwa im Sexualitäts- oder Sicherheitsdispositiv. Moderne technische Systeme können somit, als biopolitische Dispositive verstanden werden, die an die von Foucault beschriebenen Machtmechanismen anschließen. In dieser Analyse wird Metas algorithmische Inhaltsmoderation als solches Dispositiv verstanden: Sie vereint Moderationsrichtlinien, algorithmische Klassifikationssysteme und normativ codierte Plattformziele wie "Community Safety" oder "User Engagement".

Gouvernementalität

Mit dem Konzept der Gouvernementalität beschreibt Foucault eine neue Regierungsform, die sich nicht auf direkte Autorität oder staatliche Zwangsmaßnahmen beschränkt, sondern auf die Lenkung von Individuen durch Selbstführung, Techniken und Rationalitäten abzielt. In seinen Vorlesungen zur „Gouvernementalität“ (Foucault, 2006: 173-194) analysiert er, wie Macht durch Wissen, Selbstregulierung und spezifische Regierungstechnologien wirkt, ohne dass der Staat immer als zentraler Akteur auftritt.

Im Fall von Meta zeigt sich eine solche gouvernementale Macht darin, wie das Unternehmen Nutzer:innen über algorithmische Gewichtungen und Incentivierungen zu bestimmten Interaktionen anleitet, ohne direkt autoritär einzugreifen. Die Plattform wird so zu einem Akteur normativer Ordnung im digitalen Raum.

Normalisierung

Ein zentrales Machtinstrument ist die Klassifikation und Kategorisierung von Inhalten durch algorithmische Systeme. Wie Beer (2016) zeigt, erzeugen solche Prozesse Vorstellungen davon, was „normal“, „angemessen“ oder „abweichend“ ist, und etablieren damit digitale Normen, die selten hinterfragt werden. Durch die ständige Bewertung und Sortierung von Beiträgen, Kommentaren und Nutzer:innenverhalten werden bestimmte Sichtweisen und Verhaltensweisen privilegiert, während andere unsichtbar gemacht oder marginalisiert werden. Auf diese Weise tragen algorithmische Systeme nicht nur zur Reproduktion bestehender gesellschaftlicher Normen bei, sondern können auch neue Standards und Erwartungen im digitalen Raum schaffen, die das Handeln und die Selbstwahrnehmung der Nutzer:innen nachhaltig beeinflussen.

3.1.2 Datenbasis und Materialauswahl

Aufgrund der eingeschränkten Transparenz des Unternehmens Meta war ein direkter Zugang zu internen Entscheidungsstrukturen nicht möglich. Die Analyse stützt sich deshalb auf eine gezielte Auswahl von Quellen, die entlang der Forschungsfrage und der theoretischen Perspektive ausgewählt wurden. Eine Übersicht über die analysierten Materialien befindet sich im Anhang A. Es wurden folgende Quellentypen berücksichtigt:

- **Offizielle Unternehmenskommunikation** (z. B. Blogposts, Reden, Statements von Führungspersonen): Diese Dokumente wurden daraufhin analysiert, wie Meta algorithmische Systeme legitimiert und Verantwortung für deren Wirkung kommuniziert (bzw. verschleiert).
- **Technische Dokumentationen und Sekundärliteratur**: Zur Kontextualisierung der Plattformarchitektur wurden wissenschaftliche Analysen algorithmischer Systeme herangezogen.
- **Relevante EU-Regulierungen** (z. B. Digital Services Act): Diese wurden analysiert im Hinblick darauf, wie sie mit Metas Machtstrukturen interagieren oder an deren Regulierung scheitern.

Die methodische Herangehensweise ist dabei qualitativ und fallbezogen. Eine begrenzte Anzahl relevanter Dokumente wurde intensiv untersucht, um tiefergehende Einsichten in die Strukturierungslogiken der Plattform zu gewinnen. Die Auswahl erfolgte theoriegeleitet und entlang der Relevanz für die Forschungsfrage.

3.1.3 Grenzen und Erkenntnispotenzial

Die Untersuchung ist aufgrund der selektiven Quellenlage und der Intransparenz Metas in ihrer Aussagekraft begrenzt. Viele offizielle Dokumente bleiben vage oder strategisch formuliert (Pasquale, 2015). Dennoch erlaubt die strukturelle Machtanalyse eine differenzierte Perspektive auf die diskursive und technische Architektur digitaler Kontrolle. Sie macht sichtbar, wie algorithmische Systeme nicht nur Inhalte moderieren, sondern auch gesellschaftliche Normen formen – und somit digitale Souveränität sowie Geschlechterverhältnisse maßgeblich beeinflussen.

3.2 Critical Frame Analysis

Die CFA wurde in dieser Arbeit als zentrales methodisches Instrument gewählt, um die diskursiven Selbstinszenierungen des Unternehmens Meta im Spannungsfeld zwischen algorithmischer Moderation, digitaler Souveränität und Geschlechterverhältnissen zu untersuchen. Grundlage bildet die Methode nach Verloo (2005), die ursprünglich zur Analyse politischer Gleichstellungsdiskurse entwickelt wurde. Die Anwendung der CFA auf technologiebezogene Diskurse ist durch die einschlägige Literatur legitimiert: Emanuela Lombardo und Petra Meier (2008) zeigen, dass die von Verloo entwickelten Prinzipien der Frame-Analyse auch in anderen thematischen Kontexten fruchtbar gemacht werden können. Ziel ist es, die zentralen Deutungsmuster (Frames), mit denen spezifische gesellschaftliche Probleme konstruiert, legitimiert und naturalisiert werden, systematisch offenzulegen und hinsichtlich ihrer Machtwirkungen zu analysieren (Lombardo & Meier, 2008: 105-107).

3.2.1 Methodologische Grundlagen

Die CFA basiert auf der Annahme, dass politische und mediale Diskurse keine neutralen Beschreibungen von Realität liefern, sondern spezifische Problemdefinitionen, Ursachen, moralische Bewertungen und Handlungsaufforderungen erzeugen. Frames strukturieren die Wahrnehmung gesellschaftlicher Konflikte und beeinflussen, welche Maßnahmen als legitim oder notwendig erscheinen (Verloo, 2005: 20). Die Critical Frame Analysis untersucht, wie das Framing von Geschlechterungleichheit als politisches Problem

bestimmte Sichtweisen und Lösungswege nahelegt und dabei die Rolle verschiedener Akteure sowie die Zuschreibung von Verantwortlichkeiten berücksichtigt. Indem sie die Logik von Policy Frames und die diskursiven Praktiken in politischen Institutionen, der Zivilgesellschaft und den Medien analysiert, legt sie die Machtverhältnisse offen, die bestimmen, wer eine Stimme bei der Problemdefinition und Lösungsfindung erhält und wer ausgeschlossen bleibt (Verloo, 2005: 21-22).

3.2.2 Anwendung in dieser Arbeit und Materialauswahl

Die Critical Frame Analyse wurde in mehreren Schritten operationalisiert, angelehnt an das methodische Vorgehen bei Verloo (2005). Analysiert wurden offizielle Kommunikationsformate von Meta aus den Jahren 2018 bis 2025, darunter Blogposts, Anhörungen und Social-Media-Statements. Die Analyse stützt sich exemplarisch auf drei zentrale Dokumente, die jeweils für eine spezifische Phase in Metas strategischer Kommunikation stehen und die Entwicklung der Diskursstrategien abbilden: ein Blogpost zur KI-Strategie des Unternehmens (Pesenti, 2018, siehe Anhang B), eine öffentliche Senatsanhörung zur Plattformverantwortung während der US-Wahl (U.S. Senate Judiciary Committee, 2020, siehe Anhang B) sowie ein Instagram-Statement von Mark Zuckerberg zur Neuausrichtung der Moderationspolitik (Zuckerberg, 2025, siehe Anhang B). Im Anhang B befindet sich eine Übersicht über die analysierten Quellen. Die Auswahl erfolgte theorie- und forschungsgeleitet nach Relevanz zur Fragestellung: Alle Dokumente thematisieren algorithmische Moderation, Plattformverantwortung, Regulierung oder gesellschaftliche Wirkung. Zudem wurde auf eine zeitliche Streuung geachtet, um diachrone Entwicklungen in den Deutungsmustern nachvollziehen zu können.

Die Materialien decken unterschiedliche Kommunikationsformate sowie zentrale politische und historische Kontexte ab. Dadurch lässt sich nachzeichnen, wie Meta seine Diskursstrategien im Verlauf wichtiger Ereignisse, etwa bei der Einführung neuer KI-Technologien oder bei politischen Anhörungen, situativ anpasst. Aufgrund der Vielzahl potenziell verfügbarer Texte wurde die Auswahl bewusst auf Dokumente beschränkt, die exemplarisch für kommunikative Wendepunkte stehen und somit eine besondere Aussagekraft für die Analyse algorithmischer Moderationsdiskurse und digitaler Souveränität besitzen.

Im ersten Schritt wurde das Material im Rahmen eines Close Readings systematisch auf wiederkehrende Narrative, Schlüsselbegriffe, argumentative Strategien und rhetorische Figuren hin untersucht. Zur besseren Nachvollziehbarkeit sind in Anhang B (Tabellen 3–

6) jeweils drei exemplarische Segmente pro Frame aufgeführt, die die erste Codierung illustrieren. Anschließend wurden diese Passagen in einem kombinierten induktiven und deduktiven codier Verfahren systematisch entlang der vier Frame-Komponenten nach Verloo (2005: 20) erfasst – Problemdefinition, Ursachenzuschreibung, moralische Bewertung und Handlungsaufforderung (siehe Anhang B, Tabelle 2). Dieses Vorgehen folgt der Critical Frame Analysis, die Frames als strukturierte Deutungsmuster mit genau diesen interpretativen Elementen versteht.

Die entstandenen Codes wurden in einem zweiten Schritt verdichtet, sodass konsistente und dominante Frames erkennbar wurden. Für die vertiefte Auswertung wurden vier zentrale Frames herausgearbeitet, da sie das Spannungsfeld zwischen algorithmischer Moderation, Plattformverantwortung und gesellschaftlicher Wirkung besonders prägnant erfassen: der Neutralitäts- und Objektivitätsframe, der Effizienz- und Sicherheitsframe, der Fortschritts- und Innovationsframe sowie der Verantwortungsauslagerungsframe (Responsibility Shifting). Diese Frames spiegeln zentrale narrative Strategien wider, mit denen Meta seine algorithmische Moderation legitimiert und zugleich soziale wie geschlechtliche Machtverhältnisse stabilisiert.

Anschließend wurden die Frames in Bezug auf ihre politischen, sozialen und geschlechtsspezifischen Implikationen analysiert. Zentral war dabei die Frage, wie über die Sprache von Objektivität, Effizienz oder Fortschritt gesellschaftliche Ungleichheiten – insbesondere Gender-Asymmetrien – legitimiert oder verschleiert werden.

Methodisch wurde dabei das Konzept der hegemonialen Männlichkeit von Connell (2015) als kritischer Bezugsrahmen eingesetzt, um die Wechselwirkungen zwischen technologischer Infrastruktur, diskursiven Strategien und kulturell verankerten Geschlechterhierarchien zu analysieren. Dieses Konzept ermöglicht es, die komplexen Machtmechanismen hinter algorithmischer Inhaltsmoderation nicht nur als technische oder ökonomische, sondern als tief in soziale und kulturelle Strukturen eingebettete Prozesse zu verstehen und zu hinterfragen. So kann genauer eruiert werden, inwiefern die identifizierten Frames bestehende Machtverhältnisse stabilisieren oder gegebenenfalls transformieren.

Abschließend wurden die Frame-Strukturen mit aktuellen Regulierungsbemühungen auf EU-Ebene (z. B. DSA, DMA, AI-Act) kontextualisiert. Ziel war es, herauszuarbeiten, dass und wie Metas Diskurse strategisch auf diese politischen Entwicklungen reagieren – sei es durch Akzeptanz, Umdeutung oder Delegitimierung regulatorischer Maßnahmen.

3.2.3 Erkenntnispotenzial der CFA

Die Anwendung der CFA ermöglichte es, die strategische Kommunikationsweise von Meta als Teil einer umfassenden Legitimationsstrategie zu verstehen, die über klassische PR-Ziele hinausgeht. Die analysierten Frames dienen der Herstellung von Akzeptanz für algorithmische Kontrolle, der Externalisierung von Verantwortung sowie der diskursiven Stabilisierung hegemonialer Machtverhältnisse – insbesondere im Hinblick auf Geschlecht, Technik und demokratische Kontrolle.

Die CFA wurde dabei nicht nur als klassisch qualitative Inhaltsanalyse verstanden, sondern im Sinne einer machtkritischen Diskursanalyse, die auch implizite Bedeutungsstrukturen, Auslassungen und politische Rahmungen sichtbar macht. In Verbindung mit der strukturellen Machtanalyse nach Foucault (siehe Abschnitt 3.1) konnte somit eine mehrdimensionale Perspektive auf das Machtwissen algorithmischer Moderation und ihrer gesellschaftlichen Einbettung eröffnet werden.

3.3 Datenbasis & Abgrenzung

Da einige zentrale Dokumente von Meta nicht öffentlich zugänglich sind, werden ergänzend journalistische Recherchen und NGO-Berichte zur Kontextualisierung und zur Ergänzung des Datenkorpus herangezogen. Wissenschaftliche Sekundärquellen dienen darüber hinaus der kritischen Einordnung und sind bereits im Forschungsstand berücksichtigt.

Vor dem Hintergrund dieses Kapitels rückt im Folgenden die strukturelle Machtanalyse in den Fokus. Sie ermöglicht es, die spezifischen Mechanismen und Dynamiken der Machtverteilung innerhalb digitaler Plattformen wie Meta systematisch zu untersuchen und deren Auswirkungen auf gesellschaftliche Teilhabe und demokratische Kontrolle vertieft zu analysieren.

4 Strukturelle Machtanalyse von Meta nach Foucault

Die folgenden Ergebnisse der Analyse struktureller Machtverhältnisse auf digitalen Plattformen werden auf Basis der im Methoden-Kapitel (siehe Kapitel 3.1) erläuterten strukturellen Machtanalyse in Anlehnung an Michel Foucault dargestellt. Im Mittelpunkt steht ein Verständnis von Macht als produktives, dezentral organisiertes Geflecht sozialer Praktiken, normativer Ordnungen und institutioneller Strukturen (Foucault, 1978b: 82-84). Untersucht wird, wie Macht über technische Infrastrukturen, Algorithmen und organisationales Wissen gesellschaftliche Normalitäten herstellt, stabilisiert und kontrolliert.

Im Fokus stehen die Praktiken, mit denen Meta als privatwirtschaftlicher Plattformbetreiber seine Inhaltsmoderation organisiert, normativ absichert und öffentlich legitimiert. Ziel ist es, die Mechanismen sichtbar zu machen, durch die algorithmische Systeme, unternehmerische Leitbilder und regulatorische Rahmenbedingungen zu einem Dispositiv verschmelzen, das die Sichtbarkeit von Informationen, politischen Positionen und gesellschaftlichen Gruppen beeinflusst.

Die empirische Grundlage bildet eine gezielte Dokumentenanalyse von:

- offiziellen Plattformdokumenten und Moderationsrichtlinien (Meta, n.d.),
- politischen Regulierungsdokumenten auf EU-Ebene (z.B. DSA),
- externen zivilgesellschaftlichen Berichten und Evaluierungen (z. B. Corporate Europe Observatory),
- sowie einschlägiger wissenschaftlicher Literatur zur Plattformmacht, digitalen Infrastrukturen und algorithmischer Diskriminierung (z.B. Zuboff, 2018).

Eine Übersicht der untersuchten Materialien befindet sich im Anhang A.

Die Analyse folgt dabei zentralen Leitfragen:

- Welche strukturellen Machtverhältnisse reproduzieren sich über Metas Moderationspraktiken?
- Inwiefern wirken Algorithmen als normative Instrumente der Normalisierung und Kontrolle?
- Wie positioniert sich Meta gegenüber politischen Regulierungsversuchen – und welche Konflikte entstehen daraus in Bezug auf digitale Souveränität?

Die strukturelle Machtanalyse schafft damit eine theoretisch fundierte, empirische Grundlage für die anschließende CFA (Kapitel 5). Sie legt offen, wie tiefgreifend Macht im digitalen Raum in technische Infrastrukturen, ökonomische Interessen und diskursive Ordnungen eingelassen ist – und bereitet so die Analyse jener symbolischen Rahmungen vor, durch die gesellschaftliche Probleme auf Plattformen wie Meta verhandelt werden.

4.1 Plattformmacht und algorithmische Inhaltsmoderation bei Meta

Im Folgenden wird Meta anhand von Foucaults Konzepten des Dispositivs, der Normalisierung und der Gouvernamentalität als zentraler Akteur im Bereich sozialer Medien

analysiert. Dabei zeigt sich, wie Macht durch technische und organisatorische Strukturen ausgeübt wird und zugleich der Eindruck entsteht, Verantwortung liege bei den Nutzer:innen, obwohl die Entscheidungsmacht faktisch beim Unternehmen verbleibt.

Das Internet und insbesondere soziale Medien haben die Verbreitung und Auffindbarkeit von Informationen grundlegend verändert. Ursprünglich galt das Internet als Symbol für Freiheit und Wandel, doch diese Vision wird zunehmend durch die Machtasymmetrie großer Plattformbetreiber unterlaufen (Zuboff, 2018). Am Beispiel Metas wird deutlich: Ein privatwirtschaftliches Unternehmen legt die Regeln der Inhaltsverbreitung fest, während undurchsichtige Algorithmen darüber bestimmen, welche Inhalte sichtbar sind. Die Nutzer:innen haben dadurch nur begrenzte Kontrolle über das, was sie sehen und teilen können, und sind somit in ihrer digitalen Selbstbestimmung eingeschränkt.

Im Zentrum der aktuellen Debatte stehen zwei gegensätzliche Perspektiven: Einerseits fordert die Europäische Union mehr Transparenz und Kontrolle über digitale Plattformen, um demokratische Werte und Rechte zu sichern. Andererseits verteidigt Meta seine eigenen, intern gesetzten Richtlinien als Schutz der Meinungsfreiheit. Diese Spannungen verdeutlichen den grundlegenden Konflikt zwischen der privatwirtschaftlichen Plattformmacht und dem Anspruch demokratischer Regulierung.

Da Meta als privatwirtschaftliches Unternehmen nur begrenzte Einblicke in interne Entscheidungsprozesse, algorithmische Strukturen und Moderationspraktiken gewährt, stützt sich die Analyse zwangsläufig auf öffentlich zugängliche Informationen, insbesondere auf die offizielle Selbstdarstellung des Unternehmens. Unabhängige wissenschaftliche Untersuchungen oder externe Berichte zu den konkreten Abläufen der Inhaltsmoderation sind selten oder nur eingeschränkt verfügbar, da Meta Transparenz nur in dem Maße herstellt, wie es dem eigenen Unternehmensinteresse entspricht. Diese Ausgangslage macht es notwendig, die von Meta kommunizierten Strategien und Strukturen besonders kritisch zu reflektieren, um die bestehenden Transparenzdefizite und die daraus resultierende Machtasymmetrie zwischen Plattformbetreiber und Öffentlichkeit sichtbar zu machen.

Wie Meta seine eigenen Moderationspraktiken nach außen darstellt, lässt sich exemplarisch an der offiziellen Strategie des Unternehmens ablesen: Nach Angaben von Meta (n.d., siehe Anhang A) verfolgt das Unternehmen das Ziel, Menschen eine Stimme geben und sie zugleich schützen zu wollen. Die offizielle Strategie „Entfernen, Verbreitung reduzieren, informieren“ (ebd., Abs. 1) verdeutlicht, dass Algorithmen auf Meta-Plattformen bereits früh eingreifen und Inhalte teilweise noch vor Nutzer:innenmeldungen

entfernen. Dadurch wird die Kontrolle über die Sichtbarkeit von Inhalten weitgehend automatisiert und dem unmittelbaren Einfluss der Nutzer:innen entzogen.

Im zweiten Schritt überprüft ein speziell abgestelltes Review-Team die von der KI entfernten Inhalte manuell. Da die schiere Masse an Beiträgen eine vollständige Kontrolle unmöglich macht, ordnet Meta diese nach Schweregrad, Viralität und vermuteter Regelverletzung, um die Bearbeitung zu priorisieren (Meta, 2024, Abs. 3, siehe Anhang A). Die Bewertungen der Review-Teams fließen wiederum in das Training der Moderations-KI ein und ermöglichen so eine kontinuierliche Anpassung der Algorithmen (ebd., Abs. 5). Mit wachsender Datenbasis entwickelt die KI ihre Fähigkeiten fortlaufend weiter. Die menschlichen Prüfer:innen konzentrieren sich deshalb vor allem auf besonders gravierende, virale oder komplexe Fälle – laut Meta Bereiche, in denen menschliche Bewertung den algorithmischen Verfahren überlegen ist (ebd., Abs. 6).

Ergänzend kooperiert Meta nach eigenen Angaben mit ausgewählten zivilgesellschaftlichen Organisationen als „Trusted Partner“, um lokale Kontexte besser einzubeziehen (Meta, 2023, siehe Anhang A). Allerdings liegt die Auswahl dieser Partner vollständig in der Hand des Unternehmens, was Unabhängigkeit und Vielfalt der Perspektiven einschränken kann.

Die Grundlage der Moderation bilden die von Meta erarbeiteten globalen Gemeinschaftsstandards, die einen Rahmen für das erlaubte Verhalten und die zulässigen Inhalte auf den Plattformen schaffen sollen. Ziel dieser Richtlinien ist es nach Metas Angaben, einen Raum zu schaffen, in dem Menschen selbst entscheiden können, welche Inhalte sie konsumieren (Meta, n.d., Abs. 2, siehe Anhang A). Gleichzeitig betont Meta die Bedeutung der freien Meinungsäußerung, schränkt diese jedoch ein, sobald Missbrauch festgestellt wird (Kaplan, 2025). Hier zeigt sich, dass die Deutungshoheit über die Grenze zwischen legitimer Meinungsäußerung und Missbrauch letztlich bei Meta liegt, wodurch die Machtasymmetrie zwischen Plattform und Nutzer:innen weiter verstärkt wird.

Durch die Analyse, der von Meta bereitgestellten Informationen wird deutlich, dass das Unternehmen zwar Maßnahmen zur Regulierung der Plattformen ergreift, diese jedoch weitgehend intern organisiert und keiner effektiven externen Kontrolle unterliegen. Meta bestimmt eigenständig, welche Expertise und welche Kriterien bei der Weiterentwicklung der Moderationssysteme einfließen. Die genaue Programmierung der Technologien sowie die Zusammensetzung der Review-Teams bleiben intransparent. Zwar gibt Meta an, dass die Review-Teams aus einer Mischung von Vollzeitmitarbeitenden und von Partnerunternehmen beschäftigten Inhaltsprüfer:innen bestehen, die weltweit an über 20 Standorten

tätig sind und mehr als 50 Sprachen sprechen (Meta, 2022b, siehe Anhang A). Dennoch ist für Außenstehende nicht überprüfbar, wie diese Vielfalt tatsächlich umgesetzt wird und nach welchen Kriterien die Prüfer:innen ausgewählt werden. Die Durchsetzung der Gemeinschaftsstandards und die Entwicklung der Moderationstechnologien unterliegen somit keiner unabhängigen Kontrolle, sondern verbleiben vollständig in der Hand des Unternehmens (Meta, 2022b, siehe Anhang A). Dies unterstreicht, dass die behauptete Vielfalt und Offenheit der Moderation letztlich durch unternehmensinterne Interessen und Strukturen dominiert werden.

Daraus ergibt sich eine strukturelle Asymmetrie der Macht: Die Mitarbeitenden der Review-Teams, die über die Sichtbarkeit von Inhalten auf Plattformen wie Instagram, Facebook oder Threads entscheiden, agieren in einem klaren Abhängigkeitsverhältnis zum Unternehmen. Dadurch können ihre Entscheidungen nicht als völlig unabhängig betrachtet werden, sondern orientieren sich vorrangig an den Interessen Metas; externe Maßstäbe wie menschenrechtliche Prinzipien spielen dabei allenfalls eine ergänzende Rolle. Da zugleich wirksame externe Kontrollmechanismen fehlen, bleiben mögliche Verzerrungen in der digitalen Öffentlichkeit schwer erkennbar, was bestehende Machtverhältnisse zementieren kann.

Im weiteren Verlauf des Kapitels wird diese strukturelle Plattformmacht anhand der Analytischen Begriffe Foucaults der Dispositive, Gouvernementalität und Normalisierung auf Grundlage der zuvor dargestellten Unternehmensstrukturen näher untersucht und mit Blick auf digitale Öffentlichkeiten eingeordnet und in Beziehung zur digitalen Souveränität sowie zur Reproduktion von Geschlechterungleichheiten gestellt.

4.2 Algorithmische Inhaltsmoderation als Dispositiv struktureller Macht

Die Ausgestaltung von Metas Inhaltsmoderation lässt sich mit Foucaults Dispositiv-Begriff als ein komplexes, strategisch organisiertes Gefüge institutioneller, technischer und diskursiver Elemente beschreiben (vgl. Foucault, 1978b: 82–84; siehe Kapitel 3.1). Im Fokus steht dabei nicht ausschließlich die Festlegung gesellschaftlicher Normalität, sondern die strukturelle Verknüpfung von Wissensproduktion, Macht und technologischer Infrastruktur, durch die überhaupt erst Möglichkeitsräume von Sichtbarkeit, Verhalten und Kommunikation entstehen.

Meta organisiert seine Plattformen über ein Ineinandergreifen technischer Algorithmen, interner Moderationsrichtlinien und personeller Kontrollinstanzen (Review-Teams), das gezielt Handlungsoptionen, Entscheidungswege und Formen der Beteiligung strukturiert.

Die algorithmische Inhaltsmoderation folgt dabei nicht nur der technischen Logik der Datenverarbeitung, sondern ist – wie im Dispositivverständnis Foucaults – auf eine umfassende Steuerung sozialer Praktiken, Kommunikationswege und Wahrnehmungsweisen gerichtet (vgl. Foucault, 1983: 90–98; Foucault, 2006: 119–122).

Im Unterschied zur Normalisierung, die vorrangig auf die Festlegung und Durchsetzung gesellschaftlicher Standards zielt (vgl. Kapitel 4.3), meint das Dispositiv in Bezug auf Metas Unternehmensstrukturen ein Regelungs- und Produktionsnetz, das die „Spielregeln“ des digitalen Raums vorgibt: Wer erhält Sichtbarkeit? Welche Resonanzräume entstehen? Wer entscheidet über Grenzziehungen und Deutungshoheit? Die Moderationspraxis ist somit Ausdruck einer Dispositivmacht, weil sie technische Artefakte, organisatorische Prozesse und diskursive Normierungen so verschränkt, dass eine spezifische Ordnung sozialer Realität erzeugt wird – jenseits bloßer Normdurchsetzung.

Indem Meta Aspekte wie „Remove, Reduce, Inform“ (Meta, n.d., siehe Anhang A) zum Prinzip erhebt, legitimiert es diese Steuerung nach außen mit dem Verweis auf Meinungsfreiheit und Nutzer:innenschutz, verankert die eigentliche Definitionsmacht über Relevanz, Sichtbarkeit und Ausschluss jedoch fest im Unternehmen selbst. So entsteht ein Zugriff auf zentrale Infrastrukturen der digitalen Öffentlichkeit, bei der demokratische Rechenschaft kaum greift und Kontrolle diskret, aber dauerhaft durchgesetzt wird (vgl. Foucault, 1978b: 82–84; Pasquale, 2015: 207).

Die Wirkung des Dispositivs entfaltet sich darin, dass Nutzer:innen ihre Handlungsspielräume, Kommunikationsmöglichkeiten und Selbstinszenierungen aktiv an von Meta geschaffene Strukturen anpassen – ohne sich der strukturellen Vorprägungen im digitalen Raum immer bewusst zu sein.

4.3 Normalisierung: Technische Reproduktion gesellschaftlicher Normen

Normalisierung bezeichnet bei Foucault einen Machtmechanismus, der aktiv definiert, was als „normal“ gilt und was als „Abweichung“ markiert wird (Foucault, 2006: 87–98). Überträgt man diesen Begriff auf Metas Plattformen, so zeigt sich: Nicht explizite Verbote regeln den digitalen Raum, sondern algorithmische Verfahren, die laufend festlegen, welche Inhalte sichtbar bleiben, an den Rand gedrängt oder unsichtbar gemacht werden. So entstehen technische Prozesse, die bestehende gesellschaftliche Wertungen nicht nur widerspiegeln, sondern fortlaufend verstärken und Abweichungen unsichtbar machen. Meta fungiert damit als Akteur, der über algorithmische Klassifikationen aktiv an der Herstellung dessen beteiligt ist, was als „normal“ wahrgenommen wird.

Während das Dispositiv (siehe Kapitel 4.2) die infrastrukturellen und organisatorischen Möglichkeitsbedingungen digitaler Kommunikation beschreibt, bezeichnet Normalisierung die algorithmische Bewertung und Klassifizierung von Inhalten nach unternehmensinternen Maßstäben. Maschinelles Lernen, historische Interaktionsdaten und automatisierte Filterprozesse verstärken dabei bestimmte gesellschaftliche Wertungen, indem sie Abweichungen unsichtbar machen und gängige Muster als Standard hervorheben (vgl. Kayser-Bril, 2021; Foucault, 2006: 98).

Dies wird besonders bei geschlechterspezifischen Inhalten deutlich: Studien zeigen, dass algorithmische Systeme häufig heteronormative und männlich dominierte Perspektiven privilegieren, während alternative oder marginalisierte Positionen benachteiligt werden (siehe Kapitel 1.3, Connell, 2015; Noble, 2018: 10). Unternehmensinterne Strukturen wie männlich geprägte Entwicklerteams oder eine auf Effizienz ausgerichtete Firmenkultur (McKinsey & Company, 2021; Nobel 2018: 1) prägen dabei den Rahmen, innerhalb dessen die technische Inhaltsmoderation entwickelt wird. Die eigentliche Normalisierung vollzieht sich jedoch auf der Ebene der algorithmischen Prozesse selbst. Sie greifen auf Trainingsdaten zurück, die historisch bereits hegemonial männlich geprägt sind, und verstärken diese dominanten Muster systematisch – ein Vorgang, der im foucaultschen Sinne nicht bloße Reproduktion, sondern aktive Herstellung von Normalität bedeutet.

Die Prozesse der Normalisierung erfolgen dabei subtil und unsichtbar: Anstelle expliziter Zensur wirken technologische Optimierung, feingliedrige Kategorisierung und die ständige Anpassung von Sichtbarkeitsgrenzen. So wird, ganz im foucaultschen Sinne, aktiv hergestellt, was als „normal“ gilt. Meta fungiert damit nicht nur als Regelgeber, sondern als Akteur, der über technologische und organisatorische Verfahren alltäglich gesellschaftliche Realitäten strukturiert und die Handlungsspielräume insbesondere marginalisierter Gruppen einschränkt (vgl. Foucault, 2006: 87–98).

4.4 Gouvernamentalität: Steuerung durch Plattformlogiken

Die Macht, die Plattformen wie Meta ausüben, vollzieht sich nicht durch offene Zensur oder autoritäre Kontrolle, sondern – ganz im Sinne Foucaults – über subtile, verhaltenslenkende Mechanismen. Gouvernamentalität bezeichnet genau diese Form der indirekten Macht: Sie operiert nicht über Repression, sondern über die aktive *Gestaltung von Handlungsspielräumen*, in denen Nutzer:innen scheinbar frei agieren, ihre Möglichkeiten aber durch technische Ordnungen, normative Vorgaben und ökonomische Anreizsysteme strukturiert werden (Foucault, 2006: 173–194).

Übertragen auf Plattformen wie Meta bedeutet dies: Die Gouvernamentalität der Plattformen entfaltet sich vor allem über die „*Regierung der Möglichkeiten*“ – also über das algorithmische Festlegen dessen, was sichtbar, belohnt oder sanktioniert wird. Algorithmische Empfehlungssysteme sind hier das zentrale Instrument: Sie sortieren, priorisieren und personalisieren Inhalte anhand ökonomischer Logiken wie Engagement, Klicks oder Werbeerträgen – nicht nach demokratischen oder gesellschaftlich-ethischen Kriterien (Muhammad et al., 2019: 23; Kayser-Bril, 2021). Nutzer:innen begegnet damit keine neutrale Öffentlichkeit, sondern eine algorithmisch strukturierte Erfahrungswelt, die ihr Verhalten über Feedback-Schleifen formt und in produktive Bahnen lenkt (Beer, 2019: 127; Gillespie, 2018: 121).

Gerade aus gouvernementalitätstheoretischer Perspektive wird sichtbar, dass diese Steuerung so wirksam ist, weil sie Freiheit nicht einschränkt, sondern *durch Freiheit wirkt*: Über Interface-Designs, Community-Richtlinien und algorithmische Gewichtungen werden Handlungsräume nicht autoritär begrenzt, sondern unsichtbar organisiert und durch ökonomische Anreize in bestimmte Richtungen gelenkt. Governance im digitalen Raum vollzieht sich so nicht mehr primär durch Gesetze, sondern über Softwarecodes und Datenlogiken – ein genuiner Ausdruck dessen, was Foucault als *liberale Gouvernamentalität* beschrieben hat: Herrschaft durch Selbststeuerung und Anreizsysteme (Foucault, 2006: 261).

Der Gouvernamentalitätsbegriff bietet eine analytische Betrachtungsweise, die aufzeigt, dass Plattformen wie Meta nicht einfach politisch kontrollierende Macht ausüben, sondern Nutzer:innenverhalten durch ökonomische Rationalitäten und technische Anreizstrukturen hervorbringen. Dieses Spannungsverhältnis wird besonders sichtbar im Kontext der digitalen Souveränität Europas: Zwar setzen gesetzgeberische Maßnahmen wie DSA und DMA normative Rahmenbedingungen, die konkrete Steuerungsmacht bleibt jedoch in Metas Händen. Denn Gouvernamentalität bedeutet hier, dass Plattformen Verhalten auf der Mikroebene technisch formen, während Regulierung auf der Makroebene über Recht schwer an diese Mechanismen heranreicht. Praktiken wie algorithmische Blackboxes, selektive Transparenz oder technische Umgehungstaktiken sind damit nicht nur Unternehmensstrategien, sondern Ausdruck eines gouvernementalen Wissens- und Machtregimes: Meta entscheidet, welche „Trusted Partner“ als Wissensinstanz zugelassen werden, welche Daten sichtbar werden, und welche Grenzen regulatorischer Eingriffe bestehen bleiben (Corporate Europe Observatory, 2023, siehe Anhang A; Pasquale 2015).

4.5 Digitale Souveränität und die Grenzen regulatorischer Kontrolle

Die bisherigen Abschnitte haben verdeutlicht, dass Macht auf Plattformen wie Meta weniger mittels offener Regulierung als vielmehr durch strukturelle Dispositive, Normalisierungsmechanismen und gouvernementale Steuerungslogiken wirkt (vgl. Foucault, 2006: 87–98, 173–194). Die Unternehmensstrukturen von Meta (Kap. 4.1) schaffen eine hierarchische, weitgehend intransparente Ordnung, die demokratischen Prinzipien der Rechenschaft und Kontrolle entgegensteht. Im Sinne Foucaults manifestiert sich Macht hier nicht nur in expliziten Regeln, sondern vor allem in der technischen und organisatorischen Gestaltung von Sichtbarkeit, Zugang und Teilhabe (siehe 4.2).

Der Digital Services Act (DSA), die Gleichstellungsstrategie und andere regulatorische Maßnahmen der EU dienen dem Ziel, digitale Souveränität und demokratische Werte in Europa zu sichern, Diskriminierung zu bekämpfen und die Sichtbarkeit sowie Sicherheit marginalisierter Gruppen – insbesondere von Frauen und Mädchen – zu stärken (Europäische Kommission, 2022: 8; Europäische Kommission, 2020, siehe Anhang A). So ist es etwa Ziel, Frauen und Mädchen gezielt vor digitaler Gewalt und Belästigung zu schützen und ihnen mehr IT-Kompetenzen zu ermöglichen. Auch werden Plattformen verpflichtet, gegen schädliche und missbräuchliche Inhalte vorzugehen – selbst, wenn diese nicht im engen Sinne rechtswidrig sind (Europäische Kommission, 2020, siehe Anhang A).

Doch wie die Analyse in den Kapiteln 4.2 bis 4.4 gezeigt hat, sind diese regulatorischen Eingriffe bislang nur bedingt wirksam. Dies zeigt sich zum Beispiel darin, dass Facebook (heute Meta) trotz offizieller Privacy-Einstellungen wie "Friends Only" wiederholt durch Datenpannen auffiel: Drittanbieter-Apps erhielten Zugriff auf Daten, obwohl dies laut Datenschutzeinstellungen ausgeschlossen war, und Daten wurden auch nach Löschung noch weiterverarbeitet (Pasquale, 2015: 144–145). Gleichzeitig berichten Studien und zivilgesellschaftliche Organisationen regelmäßig von algorithmisch verstärkter Unsichtbarkeit marginalisierter Gruppen, etwa wenn feministische Beiträge, Posts von LGBTQ+-Communities oder Berichte zu antirassistischer Arbeit von algorithmischen Moderationen besonders häufig entfernt oder eingegrenzt werden (Kayser-Bril, 2021; Noble, 2018; Geschke et al., 2019).

Insbesondere die Studie von Daniel Geschke et al. (2019) verdeutlicht, dass Hassrede, sexistische Kommentare und rechtsextreme Inhalte auf Plattformen regelmäßig nur unzureichend moderiert werden und lange sichtbar bleiben. Besonders betroffen sind junge Menschen, die vermehrt angeben, sich aufgrund der Dominanz von Hassrede und diskriminierenden Äußerungen aus öffentlichen Debatten zurückzuziehen (Geschke et al.,

2019: 28). Auf diese Weise verstetigen digitale Plattformstrukturen bestehende gesellschaftliche Machtverhältnisse – sie bilden diese nicht nur ab, sondern festigen sie technisch und organisatorisch.

Die Definitionsmacht über Sichtbarkeit, Kategorisierung und Ausschluss verbleibt somit weitgehend bei Meta selbst, das als Dispositivbetreiber zugleich die strukturellen und organisatorischen Bedingungen gesellschaftlicher Kommunikation schafft und sich so der demokratischen Kontrolle entzieht (Pasquale, 2015: 207). Hinzu kommen Praktiken wie die strategische Auswahl von "Trusted Partnern" sowie gezielte Lobbyarbeit, die Teilhabe an regulatorischen Prozessen sichern und oft industriefreundliche Regelungen begünstigen (Corporate Europe Observatory, 2023: 4–5, siehe Anhang A).

Das Problem struktureller Intransparenz zeigt sich besonders deutlich im Datenschutzregime: Selbst, wenn Nutzer:innen formal einwilligen, bleibt meist unklar, wie Daten tatsächlich verarbeitet und weitergegeben werden – etwa bei personalisierten Werbeanzeigen oder der Verknüpfung von Profil- und Verhaltensdaten über verschiedene Dienste hinweg (Pasquale, 2015: 144–145).

Entscheidend ist, dass Macht bei Meta nicht allein von außen korrigierbar ist. Die eigentliche Steuerung geschieht über die subtile Verknüpfung technischer, organisatorischer und diskursiver Prozesse. Was als akzeptabel, sichtbar oder schützenswert gilt, wird im Alltag der Moderationspraxis permanent durch algorithmische Selektion und Normalisierung bestimmt – ein Vorgehen, das nach Foucaults Ansatz als weit über die Einhaltung von Gesetzen hinausgehend und gesellschaftliche Wirklichkeiten technisch wie symbolisch herstellend, interpretiert werden kann (Foucault, 2006: 98, 173–194).

So zeigt sich, dass digitale Souveränität mehr sein muss als rechtliche Regulierung: Sie ist – in foucaultscher Perspektive – stets auch eine Frage symbolischer Ordnungsmacht, der Aneignung von Deutungshoheit und Sichtbarkeitsbedingungen. Solange die tiefenstrukturellen Machtmechanismen der Plattformökonomie unangetastet bleiben, ist die tatsächliche Selbstbestimmung der Nutzer:innen und der Gesellschaft insgesamt auf digitale Öffentlichkeiten begrenzt (Helberger et al., 2017: 10–11).

Gerade in Krisensituationen – wie etwa während der Corona-Pandemie, als Falschinformationen und Verschwörungserzählungen algorithmisch besonders stark verbreitet wurden, oder im Kontext politischer Polarisierung, in dem geschlechterfeindliche Angriffe in Kommentarspalten häufig sichtbar blieben – werden die Konsequenzen dieser Machtasymmetrien besonders deutlich (Cinelli, 2020: 1; Geschke et al., 2019: 11).

Diese Erkenntnisse bilden die Grundlage für die folgende Critical Frame Analyse, die genauer beleuchtet, wie Geschlechterverhältnisse und gesellschaftliche Deutungsmacht in der Praxis von Meta und vergleichbaren Plattformen (re-)produziert werden.

4.6 Zwischenfazit

Die vorige Analyse der Moderationspraxis und Plattformmacht von Meta macht deutlich, wie tiefgreifend digitale Öffentlichkeiten von privatwirtschaftlichen Akteuren und deren intransparenter, algorithmischer Strukturierung geprägt werden. Ausgehend von einer an Foucault orientierten Perspektive zeigt sich, dass die Macht von Plattformen nicht in punktuellen Eingriffen, sondern in dauerhaft wirksamen Dispositiven, Formen der Gouvernamentalität und subtilen Prozessen der Normalisierung besteht (vgl. Foucault, 1978; 2005). Meta gestaltet nicht nur technische Infrastrukturen, sondern schafft die Voraussetzungen dessen, was als sagbar, sichtbar und „normal“ gilt – und damit die symbolischen Spielräume gesellschaftlicher Teilhabe.

Die Regulierung durch die Europäische Union – etwa in Form des DSA, AI-Act und Gleichstellungsinitiativen – adressiert diese Problematiken mit wichtigen Vorgaben für Transparenz, Antidiskriminierung und Nutzer:innenschutz. Doch wie die Analyse der letzten Kapitel gezeigt hat, bleibt deren Reichweite begrenzt, solange Schlüsselmechanismen der Plattformmacht wie algorithmische Blackboxes, unternehmensinterne Normsetzung und die (Re-)Produktion gesellschaftlicher Ungleichheiten – etwa durch Normalisierung hegemonialer Männlichkeit – weitgehend unberührt bleiben. Die praktische Definitionsmacht über Sichtbarkeit, Diskurspositionen und Ausschluss verbleibt größtenteils bei Meta und entzieht sich oft der rechtlichen und demokratischen Kontrolle.

Aus intersektionaler Perspektive wird zudem deutlich, dass Diskriminierung in digitalen Räumen mehrdimensional wirkt: verschiedene soziale Differenzlinien wie Geschlecht, Race oder Klasse können sich in algorithmischen Selektions- und Ausschlussverfahren verschränken und soziale Ausschlüsse verstärken (Buolamwini & Gebru, 2018). Studien belegen, dass insbesondere marginalisierte Gruppen unter technischen Verzerrungen, mangelndem Schutz und algorithmisch bedingter Unsichtbarkeit leiden.

Methodisch offenbart die machtanalytische Herangehensweise nach Foucault einerseits zentrale Dynamiken von Dispositiven, Gouvernamentalität und Normalisierung, bleibt aber auf der Ebene struktureller Analysen verhaftet – längsschnittliche, subjektive Widerstandspraxen und individuelle Handlungsspielräume werden kaum erfasst. Die

Abhängigkeit von öffentlich verfügbaren Quellen erschwert zudem Einblicke in interne Konzernentscheidungen.

Insgesamt konnte ich zeigen: Technische und rechtliche Maßnahmen greifen allein zu kurz. Digitale Souveränität und Gerechtigkeit erfordern eine umfassende Auseinandersetzung mit den symbolischen, technischen und kulturellen Konfigurationen von Machtstrukturen im digitalen Raum. Die anschließende Critical Frame Analyse nimmt deshalb in den Blick, wie Meta seine Machtverhältnisse diskursiv legitimiert und welche Deutungsmuster für Geschlecht und Diversität im öffentlichen Plattformdiskurs dominant sind.

5 Critical Frame Analyse

Um nachvollziehen zu können, wie sich die Macht von Meta nicht nur über Infrastrukturen, sondern ebenso über kommunikative Praktiken stabilisiert, wird in diesem Kapitel die CFA angewandt (vgl. Verloo, 2005; siehe Kapitel 3.2.2). Sie dient dazu zu untersuchen, mit welchen diskursiven Strategien Meta seine algorithmische Inhaltsmoderation – etwa als objektiv, effizient oder unvermeidlich – darstellt und wie dadurch gesellschaftliche Deutungen von Geschlecht, Diversität und Verantwortung reproduziert oder verschleiert werden.

Mit diesem Perspektivwechsel von der strukturellen auf die diskursive Ebene rückt die Frage in den Mittelpunkt, durch welche Narrative und Argumentationslogiken Plattformunternehmen ihre Handlungsmacht legitimieren und welche Effekte diese Diskurse im Kontext digitaler Öffentlichkeit entfalten.

Die Analyse richtet sich, wie in Kapitel 3.2.2 erläutert, auf vier zentrale Frames: Neutralität/Objektivität, Effizienz/Sicherheit, Fortschritt/Innovation sowie Verantwortungslagerung (Responsibility-Shifting). Diese verdeutlichen das Spannungsfeld zwischen algorithmischer Moderation, Plattformverantwortung und gesellschaftlicher Wirkung. Als Grundlage dienen drei Schlüsseltexte aus unterschiedlichen politischen Kontexten (Pescetti, 2018; U.S. Senate Judiciary Committee, 2020; Zuckerberg, 2025; siehe Anhang B, Tabelle 1), die im Close Reading untersucht wurden. Anschließend erfolgte eine kombinierte induktiv-deduktive Kodierung (siehe Anhang B, Tabelle 3-6) entlang der vier Frame-Komponenten nach Verloo (2005: 20) – Problemdefinition, Ursachenzuschreibung, moralische Bewertung und Handlungsaufforderung (siehe Anhang B, Tabelle 2). Auf dieser Basis ließen sich dominante narrative Strategien rekonstruieren, mit denen

Meta seine Moderationspraxis legitimiert und zugleich bestehende Macht- und Geschlechterverhältnisse stabilisiert.

Die Frames sind analytisch getrennt dargestellt, ihre geschlechtsspezifischen Wirkungen stützen sich jedoch gegenseitig. Gerade weil Neutralitäts-, Effizienz-, Fortschritts- und Responsibility-Shifting-Frame ineinandergreifen, ergeben sich Überschneidungen in ihren Auswirkungen: Sie stabilisieren unterschiedliche Aspekte derselben hegemonialen Machtstrukturen.

5.1 Neutralitäts- und Objektivitätsframe

Das Narrativ von Neutralität und Objektivität technologischer Systeme – insbesondere von KI-gestützten Moderationssystemen – ist historisch in westlichen Gesellschaften etabliert und basiert auf einem technikdeterministischen Weltbild. Es wurzelt in einem technikdeterministischen Weltbild, das technologische Entwicklung als Inbegriff von Fortschritt, Effizienz und objektiver Entscheidungsfindung versteht. Bereits während der Aufklärung und Industrialisierung galt Rationalität als Motor gesellschaftlicher Emanzipation, und technischen Mitteln wurde eine neutralisierende, gerechtigkeitsstiftende Kraft zugesprochen (Crawford, 2021: 197-198, 214; Historisches Lexikon der Schweiz, 2014).

Wie Kate Crawford (2021: 211) aufzeigt, bestand die Erwartung, Technologie ermögliche überlegene, entemotionalisierte und damit automatisch gerechtere Entscheidungen. Diese Annahmen verkennen jedoch, dass technische Systeme nie losgelöst von gesellschaftlichen Machtverhältnissen entstehen: Algorithmen sind stets das Resultat menschlicher Entscheidungen und werden in spezifischen politischen und ökonomischen Kontexten entwickelt (Crawford, 2021: 211; vgl. Kapitel 3).

Das Bild der vermeintlichen „neutralen“ Technologie ist zudem eng mit patriarchalen Strukturen verwoben: Eigenschaften wie Objektivität und Rationalität werden kulturell als männlich und überlegen, Subjektivität und Emotionalität dagegen als weiblich und minderwertig markiert (Harding, 2016: 3; Keller, 1995: 76-77). Dies trägt dazu bei, dass hegemoniale Männlichkeitsnarrative – und damit patriarchale Machtverhältnisse – durch die Inszenierung neutraler Technologien unsichtbar bleiben (vgl. Connell, 2015).

Ein zentrales technisches Merkmal, das diese Scheinobjektivität verstärkt, ist die Intransparenz algorithmischer Systeme. Der sogenannte „Black-Box“-Effekt (Pasquale, 2015) bedeutet, dass deren Funktionsweisen nur schwer nachvollziehbar sind. Beer (2019: 127) betont, dass diese Undurchsichtigkeit Plattformen zusätzliche Macht verleiht, da Entscheidungen weder sichtbar noch rechenschaftspflichtig gemacht werden.

Empirisch zeigt sich: Algorithmen wirken keineswegs neutral, sondern können bestehende Diskriminierungen etwa entlang von Gender und Race verstärken (siehe Kapitel 2.2). Auch Daten, die als Grundlage für KI-Systeme dienen, sind keine neutralen Fakten, sondern spiegeln gesellschaftliche Vorannahmen wider (Mützel et al., 2018: 2). Durch algorithmische Weiterverarbeitung, etwa im Rahmen personalisierter Werbung, werden diese Biases oft verstärkt (Cheney-Lippold, 2011: 168). Crawford (2021) und Staab (2019) unterstreichen, dass algorithmische Systeme nicht nur in bestehende Machtverhältnisse eingebettet sind, sondern diese auch aktiv reproduzieren und gesellschaftliche Wahrnehmungen mitprägen.

Im Folgenden werden exemplarische Beispiele aus Metas öffentlicher Kommunikation auf den Frame der Neutralität und Objektivität hin analysiert. Damit wird sichtbar, wie Meta diese Rahmung strategisch einsetzt, um seine Rolle bei Fragen der digitalen Souveränität, algorithmischen Kontrolle und gesellschaftlichen Verantwortung zu gestalten. Die Analyse beginnt mit einem Blogposts von Jerome Pesenti (2018, siehe Anhang B).

5.1.1 Blogpost: Objektivität durch Forschung und Technik

Der analysierte Blogpost mit dem Titel „Facebook F8 2018 – KI: Offene Frameworks & verantwortungsvolle Entwicklung“ wurde 2018 von Pesenti, damals Vizepräsident für Künstliche Intelligenz bei Facebook (heute Meta), veröffentlicht. Der Beitrag erschien im Kontext der F8, einer jährlich stattfindenden Entwicklerkonferenz von Facebook, die Entwickler:innen, Unternehmen und technikinteressierte Personen über aktuelle technologische Fortschritte informiert. Im Mittelpunkt steht die Rolle Künstlicher Intelligenz bei der Verbesserung des digitalen Raums. Dabei werden verschiedene neue Tools und Technologien vorgestellt, wobei immer wieder betont wird, wie wichtig offene Werkzeuge, der Transfer von Forschungsergebnissen durch Open Source und ethische Standards für Facebook sind. Im abschließenden Abschnitt wird zudem die Vermeidung algorithmischer Verzerrungen hervorgehoben, verbunden mit dem Bekenntnis zu vielfältigen Teams und einer fairen KI (Pesenti, 2018, siehe Anhang B).

Besonders auffällig im Zusammenhang des hier untersuchten Neutralitäts- und Objektivitätsframes ist der Einsatz zahlreicher technischer Fachbegriffe wie „Betaversion“ oder „PyTorch language library“ (Pesenti, 2018, Abs. 3, siehe Anhang B). Damit wird ein sachlich-technischer Sprachstil erzeugt, der Emotionalität ausblendet und den Eindruck von Wissenschaftlichkeit und Objektivität verstärkt. Dieser Effekt wird zusätzlich durch wiederkehrende Verweise auf Wissenschaft, Forschung und externe Ingenieur:innen gestützt.

So wird etwa erwähnt, dass eine neue Webseite insbesondere Forscher:innen und Entwickler:innen unterstützen soll (ebd.: Abs. 3). Hier manifestiert sich das Narrativ, dass die enge Zusammenarbeit mit der Wissenschaft – verstanden als faktenbasierte, nachvollziehbare Arbeit – die technologische Entwicklung von Meta in Richtung Neutralität und Objektivität vorantreibt.

Gleichzeitig entsteht durch die Betonung des Engagements für die KI-Community der Eindruck, dass die Unterstützung dieser Gemeinschaft die Hauptmotivation von Meta bei der Entwicklung neuer Technologien sei (ebd.: Abs. 2). Kritisch bleibt jedoch, dass wirtschaftliche Gewinninteressen als treibende Kraft globaler Großkonzerne wie Meta dabei ausgeblendet werden.

Während der Blogpost eine positive, technikzentrierte Perspektive vermittelt, zeigt die Analyse einer Senatsanhörung aus dem Jahr 2020 eine andere Seite von Metas Diskurs: Dort werden algorithmische Systeme als Überwachungsinstrumente dargestellt, die zugleich als unverzichtbare technische Kontrollinstanzen legitimiert werden.

5.1.2 Senatsanhörung: Technik als Überwachungsinstanz

Am 17. November 2020 äußerte sich Mark Zuckerberg, CEO von Meta (ehemals Facebook), im Rahmen einer Anhörung vor dem US-Senat zur Rolle sozialer Medien bei der US-Präsidentchaftswahl 2020. Die Sitzung fand vor dem Hintergrund wachsender Kritik an der Einflussnahme digitaler Plattformen auf öffentliche Diskurse statt – insbesondere durch algorithmische Moderation und die Verbreitung von Desinformation. Für die Critical Frame Analyse wurden jene Passagen ausgewählt, in denen Zuckerberg zu Moderation, Plattformkontrolle, Verantwortung und Regulierungsmaßnahmen Stellung nimmt, da diese den Kern der Forschungsfrage berühren.

Im Zentrum der Befragung stand die Frage, ob Meta demokratische Prozesse beeinflusse und wie dafür Rechenschaft eingefordert werden kann (U.S. Senate Judiciary Committee, 2020, siehe Anhang B). Zuckerberg bedient konsequent den Frame der Neutralität und Objektivität: Er inszeniert die algorithmischen Systeme als rational, emotionslos und urteilsfrei und blendet aus, dass deren Entwicklung und Training menschlich und sozial situiert sind.

Dieser Neutralitätsframe stabilisiert Metas Macht, indem suggeriert wird, „neutrale Technologien“ entzogen sich demokratischer Kontrolle. Diese Logik lässt sich mit Connells Konzept hegemonialer Männlichkeit (2015) verbinden: Technik wird als rational, überlegen und frei von Emotionalität dargestellt – typisch männlich konnotierte Eigenschaften,

die technokratische Autorität legitimieren und entpolitisieren. Entpolisierung bedeutet hier, dass Entscheidungen über Sichtbarkeit und Kontrolle nicht mehr als gesellschaftlich und demokratisch auszuhandelnde Fragen erscheinen, sondern als scheinbar objektive Resultate technischer Systeme.

Zuckerbergs Behauptung, Facebook treffe keine parteipolitische Bewertung von Inhalten, verstärkt den Eindruck technologischer Objektivität. Er hebt hervor, dass KI-Systeme mindestens ebenso zuverlässig oder sogar verlässlicher als menschliche Prüfer:innen seien (U.S. Senate Judiciary Committee, 2020: 03:05:13, siehe Anhang B). Damit wird technische Überwachung idealisiert und technischer Objektivität eine Geschlechterhierarchie eingeschrieben, die Rationalität und Kontrolle als männliche Tugenden naturalisiert.

Auch die Aussage, das Produkt sei „möglichst nützlich und bedeutsam gestaltet“ (ebd.: 36:36), suggeriert, technische Architektur garantiere Neutralität – eine zusätzliche Unsichtbarmachung gesellschaftlicher Interessen, die von bestehenden Machtverhältnissen profitiert. Die Betonung unabhängiger Faktenprüfer:innen (ebd.: 25:31) dient der Externalisierung von Verantwortung, während Meta selbst die Definitionsmacht behält und institutionelle Neutralität nur vortäuscht.

Zuckerberg betont außerdem, dass die Facebook-Community keine Falschinformationen wünsche und Facebook selbst nicht entscheide, was richtig oder falsch sei. Als Reaktion darauf sei ein neues Programm entwickelt worden, das branchenweit führend sei und darauf abziele, Informationsqualität sicherzustellen (ebd.: 01:08:23 ebd.: 01:08:23). Technologie wird als übergeordnete, wertfreie Kontrollinstanz dargestellt, wodurch der Bedarf an demokratischer oder diversitätssensibler Kontrolle relativiert wird.

Insgesamt trägt Meta mit dieser Kommunikation zur Naturalisation technokratischer wie geschlechterhierarchischer Machtstrukturen bei und legitimiert stereotype Muster männlicher Dominanz im digitalen Raum. Diese naturalisierten Machtmuster bilden den Nährboden, vor dem sich spätere strategische Verschiebungen in Metas Kommunikationspraxis – wie der Wandel im Instagram-Statement von 2025 – vollziehen.

5.1.3 Instagram-Statement: Metas Strategiewechsel

Am 7. Januar 2025 veröffentlichte Mark Zuckerberg ein offizielles Statement auf allen Meta-Plattformen – inmitten einer politisch aufgeladenen Lage: Donald Trump trat seine zweite Amtszeit als US-Präsident an, nachdem die Präsidentschaftswahl im November 2024 zu einer deutlichen Verschiebung der politischen Mehrheitsverhältnisse geführt

hatte. Der politische Diskurs in den Vereinigten Staaten war zunehmend geprägt von einem Narrativ der vermeintlich eingeschränkten Meinungsfreiheit durch sogenannte „Big Tech“-Konzerne – insbesondere vorangetrieben durch republikanische Akteur:innen. Zuckerbergs Stellungnahme kann als strategische Reaktion auf diesen politischen Druck gelesen werden (McMahon, Kleinman & Subramanian, 2025): Er kündigte die Abschaffung des bisherigen Drittanbieter-Faktenprüfprogramms, die Vereinfachung der Inhaltsrichtlinien sowie eine stärkere Betonung der Meinungsfreiheit auf den Meta-Plattformen an (Zuckerberg, 2025, Abs. 8, siehe Anhang B). Zugleich markiert das Statement von 2025 eine politische Zäsur: Erstmals bezieht Meta offen Stellung gegen internationale Regulierungsbemühungen und positioniert sich sichtbar an der Seite einer konservativen Regierung (McMahon, Kleinman & Subramanian, 2025):

Zentral ist dabei ein diskursiver Perspektivwechsel: Während Meta zuvor algorithmische Kontrolle mit gesellschaftlicher Verantwortung legitimierte, verschiebt sich nun der Fokus hin zur Betonung individueller Freiheit. Diese Neuausrichtung wird begleitet von einer expliziten Kritik an internationalen Regulierungsinitiativen, insbesondere der Europäischen Union. Regulierungen der EU werden in Zuckerbergs Statement nicht als Maßnahmen zur demokratischen Kontrolle, sondern als Ausdruck staatlicher Überregulierung und potenzieller Zensur verstanden (ebd., Abs. 13). Damit konstruiert Meta einen transatlantischen Wertekonflikt: Auf der einen Seite „amerikanische Werte“ wie Redefreiheit und unternehmerische Autonomie – auf der anderen Seite europäische Regulierungsversuche, die als Einschränkung individueller und unternehmerischer Freiheit gerahmt werden.

Zuckerberg inszeniert Meta dabei weiterhin als neutrale Instanz im digitalen Raum. Die Distanzierung von einer redaktionellen Rolle wird rhetorisch durch die Infragestellung der Glaubwürdigkeit externer Faktenprüfer gestützt, denen politische Voreingenommenheit unterstellt wird. Diese Akteure hätten, so Zuckerberg, mehr Vertrauen zerstört als geschaffen (ebd., Abs. 8). Meta selbst erscheint dadurch im Kontrast als vermeintlich unparteiischer Akteur. Diese Verschiebung von Verantwortung stellt eine wichtige Schnittstelle zum nachfolgenden Verantwortungsframe (Kapitel 5.4) dar, der die Inszenierung technischer Neutralität weiter absichert und stützt.

Besonders bemerkenswert ist auch Zuckerbergs Bezug auf die technologische Komplexität der Plattformen. Er argumentiert, dass selbst eine geringe Fehlerquote von nur einem Prozent bei Milliarden Nutzer:innen zu millionenfach unbeabsichtigter „Zensur“ führen könne (ebd., Abs. 5). Diese Darstellung entpolitisiert algorithmische Eingriffe: Zensur

erscheint nicht als strukturelles Machtverhältnis, sondern als technisches Restrisiko. Objektivität wird nicht mehr durch Präzision und Faktenprüfung legitimiert, sondern durch die vermeintlich unvermeidbare Fehlertoleranz komplexer Systeme.

Zuckerbergs Ankündigung, Moderationsteams künftig von Kalifornien nach Texas zu verlagern, besitzt symbolische Relevanz. Der Schritt wird als Versuch dargestellt, politische Voreingenommenheit zu verringern und die Plattform institutionell zu „entpolitisieren“ (ebd., Abs. 12). Diese Maßnahme kann als strategische Distanzierung von Kalifornien als Standort mit spezifischer politischer Ausrichtung verstanden werden, wodurch der Neutralitätsanspruch performativ betont wird. Gleichzeitig betont diese Maßnahme den performativen Charakter des Neutralitätsanspruchs. Neutralität wird dabei weniger faktisch belegt als symbolisch inszeniert.

Im Vergleich mit den vorangegangenen Quellen wird deutlich, wie flexibel Meta den Frame der Neutralität und Objektivität einsetzt, um auf sich wandelnde politische Rahmenbedingungen zu reagieren. Während der F8-Blogpost von 2018 technokratische Begriffe und wissenschaftliche Autorität nutzte, um algorithmische Systeme als rational und ethisch verantwortungsvoll zu präsentieren, und die Senatsanhörung von 2020 algorithmische Entscheidungen als objektiver und effizienter als menschliche Eingriffe legitimierte, erfolgt im Instagram-Statement von 2025 eine strategische Neupositionierung: Nicht mehr Kontrolle und Verantwortung stehen im Fokus, sondern Freiheit und Deregulierung.

Dennoch bleibt die Rhetorik technischer Objektivität zentral: Nun aber nicht mehr zur Legitimation von Eingriffen, sondern zur Entlastung von Verantwortung. Fehler werden als unvermeidlich, aber kontrollierbar dargestellt. Kritik an Intransparenz und Machtasymmetrien wird dabei nicht ausgeräumt, sondern diskursiv umgedeutet: Sie erscheint als Angriff auf individuelle Freiheit – nicht als Forderung nach demokratischer Verantwortlichkeit.

5.1.4 EU- und Geschlechterperspektive zum Frame Neutralität & Objektivität

Der Frame der Neutralität und Objektivität algorithmischer Systeme steht in einem direkten Spannungsfeld zur europäischen Regulierung, insbesondere dem Digital Services Act (DSA), welcher seit 2024 strenge Vorgaben zu Transparenz, Nachvollziehbarkeit und Fairness von Plattformen wie Meta vorsieht. Die EU betont in ihren Regulierungen explizit Transparenz- und Offenlegungspflichten für algorithmische Systeme (vgl. Kapitel 2.3.4; Europäische Kommission, n.d.-c, siehe Anhang A; Wagner et al., 2022), um

demokratische Kontrolle zu stärken, Nutzer:innenrechte zu sichern und intransparente Moderationspraktiken einzudämmen. Gleichzeitig zeigt die Analyse des Frames, wie Plattformen wie Meta den Neutralitätsframe nutzen, um politischen und regulatorischen Druck durch die Behauptung technischer Objektivität abzumildern.

In einer geschlechterkritischen Lesart erscheint der Neutralitätsdiskurs als Strategie der Verschleierung: Indem Technik als objektiv inszeniert wird, werden geschlechtsspezifische Verzerrungen naturalisiert, während zugleich die sozialen, kulturellen und politischen Voraussetzungen technologischer Infrastrukturen verdeckt werden. Auf diese Weise tragen vermeintlich neutrale Systeme zur Stabilisierung patriarchaler Machtstrukturen im digitalen Raum bei. Die EU-Politik fordert deshalb – etwa im Rahmen der EU-Gleichstellungsstrategie – eine kritische und reflektierte Hinterfragung solcher Technologien und ihrer Wirkungen auf Diversität und Gleichstellung (vgl. Kapitel 2.3).

Indem Neutralität als technisches Attribut inszeniert wird, werden Fragen nach gesellschaftlichen Machtverhältnissen, etwa geschlechtsspezifischen Verzerrungen oder Diskriminierungen, marginalisiert. Dieses Framing erschwert eine inklusive und gerechte Regulierung algorithmischer Systeme, da er die normative Forderung nach einer transparenten und verantwortlichen Gestaltung von Technologien konterkariert. Die EU unterstreicht im Regulierungsdesign die Notwendigkeit, technische Prozesse nicht nur als technische Gegebenheiten, sondern auch als sozial bedeutungsvolle und politisch gestaltbare Systeme zu verstehen, die einer demokratischen Kontrolle unterliegen müssen (vgl. Kapitel 2.3; Europäische Kommission, 2022, siehe Anhang A).

Gerade in der digitalen Souveränität spielt die kritische Reflexion von Neutralität eine Schlüsselrolle, da sie ermöglicht, koloniale, patriarchale und ökonomische Machtasymmetrien sichtbar zu machen und aufzulösen. Dies ist essenziell, um eine nachhaltige und geschlechtergerechte Digitalisierung zu fördern, die soziale Differenzen anerkennt und verhindert, dass technische Systeme bestehende Ungleichheiten unreflektiert reproduzieren.

5.2 Effizienz- und Sicherheitsframe

Im Diskurs von Meta spielen Effizienz und Sicherheit zentrale Rollen, um den Einsatz algorithmischer Systeme zu legitimieren. Diese werden als notwendige Instrumente dargestellt, um Nutzer:innen vor Bedrohungen wie Hate Speech, Diskriminierung oder Terrorpropaganda zu schützen. Bereits seit dem ausgehenden 19. Jahrhundert ist Effizienz ein zentrales Ordnungsprinzip im Technikdiskurs, eng verbunden mit gesellschaftlicher

Modernisierung und Kontrolle (vgl. Staab, 2019: 46-52). Mit der Digitalisierung wurde dieses Paradigma zunehmend durch Sicherheitsdiskurse ergänzt, sodass die Verschränkung von Effizienz und Sicherheit heute einen prägenden Rahmen für die Legitimierung algorithmischer Systeme bildet (van Dijck, Poell & de Waal, 2018: 25–27).

Wie in der strukturellen Machtanalyse gezeigt wurde (siehe Kapitel 4), liegt die Definitionsmacht über problematische Inhalte – etwa Hate Speech – letztlich bei Meta selbst. Dies schränkt die Einflussmöglichkeiten demokratischer Institutionen wie der EU ein und begünstigt die Reproduktion bestehender gesellschaftlicher Machtverhältnisse. Der Effizienz- und Sicherheitsframe trägt dazu bei, algorithmische Systeme als „natürliche“ Antwort auf komplexe Probleme zu inszenieren.

5.2.1 Blogpost: Technologische Machbarkeit und moralischer Fortschritt

Bereits im F8-Blogpost von 2018 konstruiert Meta algorithmische Systeme als ethisch motivierte, technisch überlegene Lösung. Es fällt auf, dass gleich zu Beginn des Beitrags Formulierungen wie „verantwortungsvoll“ und „ethische Entwicklung“ verwendet werden (Pesenti, 2018, siehe Anhang B), um eine moralische Aufladung der Technik zu etablieren. Gleichzeitig werden potenzielle Risiken zwar angedeutet, jedoch sofort relativiert durch die Aussicht, mit KI-Systemen die Sicherheit der Nutzer:innen erhöhen zu können (ebd.)

In den technischen Abschnitten des Blogposts tritt der Effizienz-Frame noch stärker hervor. Es ist wiederholt von „modernster Genauigkeit und Effizienz“, „schnellen“, „skalierbaren“ Systemen die Rede. Immer wieder werden große Zahlen (z. B. „zwei Milliarden Menschen“, „3,5 Milliarden Fotos“, „85,4 Prozent Genauigkeit“) als Belege für Leistungsfähigkeit genannt. Auch das Mensch-gegen-Maschine-Narrativ wird aufgegriffen – ein KI-System schlägt im Go-Spiel mehrere Top-Spieler „14:0“, angeblich ohne menschliche Hilfe (ebd.). Dadurch wird eine technologische Überlegenheit postuliert, die algorithmische Kontrolle, als überlegen gegenüber menschlicher Einschätzung erscheinen lässt. Das Mensch-gegen-Maschine-Narrativ, ist ein verbreitetes Deutungsmuster im Technologiediskurs: Es inszeniert künstliche Intelligenz als überlegen, effizienter und leistungsfähiger gegenüber menschlichen Akteuren. Solche Darstellungen stärken die gesellschaftliche Akzeptanz algorithmischer Kontrolle und delegitimieren Zweifel an deren Vertrauenswürdigkeit (Crawford, 2021: 5-7). Gesellschaftlich birgt diese Rahmung die Gefahr, komplexe soziale Probleme ausschließlich als technologische Herausforderungen

zu deuten, wodurch politische und ethische Fragestellungen reduziert und kritische Debatten über Macht, Kontrolle und Gerechtigkeit marginalisiert werden.

Abschließend wird in dem Blogpost betont, Meta sei sich der Risiken von KI bewusst und bemühe sich um ethische, faire und vertrauenswürdige Entwicklung (ebd.) – eine Strategie, um Kritik an algorithmischer Diskriminierung vorzubeugen. Der Effizienz- und Sicherheitsframe wird hier also nicht nur technologisch, sondern auch rhetorisch aufgebaut. Im folgenden Abschnitt wird deutlich, wie sich dieses Narrativ in der öffentlichen Darstellung durch Mark Zuckerberg auf politischer Ebene weiter konsolidiert und vor allem durch den Verweis auf messbare Erfolge und institutionelle Maßnahmen gefestigt wird.

5.2.2 Senatsanhörung: Konsolidierung des Frames Effizienz & Sicherheit

In der US-Senatsanhörung am 17. November 2020 verstärkt Mark Zuckerberg den Effizienz- und Sicherheitsframe. Er verweist auf Kooperationen mit zertifizierten Faktenprüfungsorganisationen und betont große Investitionen in Sicherheitsmaßnahmen wie das „Voting Information Center“ und die Rekrutierung von Wahlhelfer:innen. Algorithmische Systeme werden so als zentrale Infrastruktur demokratischer Prozesse dargestellt (U.S. Senate Judiciary Committee, 2020, Zuckerberg: 36:36, 27:32, siehe Anhang B).

Besonders markant ist die wiederholte Hervorhebung hoher Erkennungsraten – 99 % bei terroristischen Inhalten, 94 % bei Hassrede –, ein rhetorisches Mittel, das bereits im F8-Blogpost verwendet wurde (ebd.: 53:48). Diese Zahlen sollen die Effektivität und Präzision algorithmischer Moderation gegenüber menschlicher Kontrolle unterstreichen.

Das vermittelt das Bild, dass algorithmische Systeme als unverzichtbare Schutzinstanzen fungieren, deren Versagen nicht nur technische Mängel, sondern Risiken für die gesellschaftliche Ordnung wären. Zugleich legitimiert die Betonung nahezu perfekter Erkennungsraten technokratische Autorität und schwächt Kritik an Intransparenz und Fehleranfälligkeit ab. Das bestehende Spannungsfeld zwischen dem Anspruch auf vollständige Sicherheit und der Akzeptanz systemischer Fehler wird genutzt, um regulatorische Eingriffe zusätzlich zu erschweren.

Darüber hinaus festigt die Darstellung algorithmischer Systeme als demokratische Infrastruktur die Macht von Meta in der digitalen Öffentlichkeit. So konzentriert sich die Definitionsmacht über Effizienz und Sicherheit in der Unternehmenshand, was kritisch im Hinblick auf Transparenz, Verantwortlichkeit und gesellschaftliche Machtstrukturen reflektiert werden muss.

Insgesamt erzeugt Meta ein doppeltes Narrativ: Einerseits rechtfertigt die vermeintlich enorme Datenmenge den KI-Einsatz, andererseits wird Vertrauen durch Transparenzversprechen und Lernfähigkeit hergestellt, während eine kritische Auseinandersetzung mit der Definitionsmacht des Unternehmens weitgehend ausbleibt.

Vor diesem Hintergrund entwickelt Meta mit dem Statement von Zuckerberg im Jahr 2025 den Effizienz- und Sicherheitsframe strategisch weiter, indem einerseits die Unvermeidbarkeit algorithmischer Fehler betont und andererseits eine gezielte Neujustierung der Moderationspraktiken präsentiert wird, die Effizienzsteigerung mit einer stärkeren Verschiebung der Verantwortung verbindet.

5.2.3 Instagram-Statement: Neujustierung von Effizienz & Sicherheit

Im Januar 2025 kündigt Zuckerberg eine Neuausrichtung der Moderationsstrategie an, die den Effizienz- und Sicherheitsframe weiterentwickelt und bedeutende Verschiebungen in der Verantwortungszuschreibung beinhaltet. Dabei wird zunächst die Komplexität und Fehleranfälligkeit der Systeme betont, etwa durch den Hinweis, dass „nur 1 % der Beiträge versehentlich zensiert“ werden (Zuckerberg, 2025, Abs. 5, siehe Anhang B). Diese Betonung der Unvermeidbarkeit von Fehlern dient nicht nur dazu, Zweifel an der Zuverlässigkeit der algorithmischen Moderation zu relativieren, sondern legitimiert zugleich ihre fortwährende Anwendung als unverzichtbaren Schutzmechanismus für die Plattformgemeinschaft.

Parallel dazu verschiebt Zuckerberg die Verantwortung teilweise auf die Nutzer:innen, indem er einen Rückbau der Moderation bei „weniger schwerwiegenden Verstößen“ ankündigt und die Priorisierung auf gravierende Delikte wie Terrorismus oder Kindesmissbrauch legt (Zuckerberg, 2025, Abs. 4, 14–15, siehe Anhang B). Diese strategische Fokussierung verspricht eine ressourcenschonende Effizienzsteigerung, während der Sicherheitsaspekt weiter zugespitzt wird.

Zugleich grenzt sich Meta durch Kritik an früheren Faktenprüfungsmechanismen ab, die als „zu politisch voreingenommen“ dargestellt werden (ebd., Abs. 8). Diese Differenzierung festigt das Bild eines lernfähigen und effizienteren Akteurs, der externe Kontrolle ablehnt und stattdessen auf eigene Kontrollinstanzen setzt. Symbolische Maßnahmen, wie die Verlagerung der Moderationsteams nach Texas, dienen dabei der performativen Wiedereinordnung von Vertrauen und institutioneller Neutralität.

In der Gesamtdarstellung zeigt sich eine strategische Anpassung und Verschiebung im Effizienz- und Sicherheitsframe: Während in den vorangegangenen Phasen vor allem

technische Überlegenheit und hohe Erkennungsraten als Legitimation dienen, wird im Statement von 2025 zusätzlich die Unvermeidbarkeit technischer Fehler hervorgehoben und die Verantwortung auf Nutzer:innen ausgelagert. Dadurch wird einerseits die Notwendigkeit algorithmischer Kontrolle betont, andererseits aber auch deren Grenzen und die damit verbundene Verantwortung kommunikativ neu justiert.

Diese Entwicklung setzt den bisherigen Diskurs über technologische Machbarkeit und institutionelle Sicherheit fort, ohne die zugrundeliegenden Legitimationsmuster grundlegend zu verändern. Vielmehr ergänzt und flexibilisiert Meta den Frame, um auf politischen Druck zu reagieren und die eigene Definitionsmacht in einem sich wandelnden Regulierungs- und Öffentlichkeitskontext zu sichern.

5.2.4 EU- und Geschlechterperspektive zum Frame Effizienz & Sicherheit

Der Effizienz- und Sicherheitsframe erhält im Lichte des europäischen Digital Services Act (DSA) eine neue Relevanz, da diese großen Plattformen verpflichtet, wirksame und skalierbare Maßnahmen gegen Risiken durch illegale oder schädliche Inhalte zu ergreifen (siehe Kapitel 2.3.4; Europäische Kommission, n.d.-c, siehe Anhang A). Ergänzend spielt die Datenschutz-Grundverordnung (DSGVO, siehe Kapitel 2.3.2) eine zentrale Rolle, indem sie klare Anforderungen an den Schutz personenbezogener Daten sowie an Transparenz und Rechenschaftspflicht bei der Verarbeitung durch Plattformen stellt. Meta nutzt den Effizienz- und Sicherheitsframe gezielt, um algorithmische Moderation als technisch unverzichtbare Schutzmaßnahme darzustellen und sich selbst als einzigen Akteur zu positionieren, der diese Risiken wirksam kontrollieren kann. Gleichzeitig diskreditiert das Unternehmen regulatorische Eingriffe wiederholt als innovationshemmend und ineffizient, womit der DSA diskursiv entwertet und die Notwendigkeit demokratischer, externer Kontrolle in Frage gestellt wird. Durch diese Rhetorik stellt sich Meta bewusst gegen die von der EU im DSA und DSGVO betonte Forderung nach Transparenz, demokratischer Rechenschaftspflicht und dem Schutz grundlegender Rechte (vgl. Wagner et al., 2022; siehe Kapitel 2.3.2, 2.3.4) und setzt sich somit außerhalb zentraler europäischer Kontrollinteressen.

Betrachtet man den Effizienzframe aus Genderperspektive, wird deutlich, dass die Fixierung auf Geschwindigkeit und Skalierbarkeit vor allem komplexe Ausdrucksformen benachteiligt. Gerade feministische und queere Inhalte geraten dadurch überproportional ins Visier automatisierter Filter und verschwinden häufiger aus der Sichtbarkeit. Das liegt nicht zuletzt daran, dass solche Ausdrucksweisen bewusst von dominanten

Geschlechternormen abweichen und alternative Identitäten betonen, wodurch sie stärker vom algorithmisch trainierten „Normalfall“ abweichen und damit häufiger falsch klassifiziert werden (siehe Kapitel 1.3, 2.2). Auf diese Weise verstärkt die technokratische Orientierung an Effizienz bestehende strukturelle Ungleichheiten: Marginalisierte Gruppen – insbesondere Frauen* und intersektional diskriminierte Personen – werden besonders stark durch algorithmische Fehlentscheidungen betroffen und verfügen zugleich über kaum partizipative Kontrollmöglichkeiten. Damit widerspricht die Effizienzheterik den Anforderungen einer geschlechter- und diversitätssensiblen Risikoanalyse, wie sie die EU im Rahmen ihrer Antidiskriminierungsvorgaben, des AI-Act (siehe Kapitel 2.3.5) und des Datenschutzes (siehe Kapitel 2.3.2) vorsieht, und erschwert deren Umsetzung.

Im EU-Kontext gewinnt diese Kritik besondere Bedeutung, da der DSA, AI-Act und die DSGVO nicht nur den Schutz vor illegalen Inhalten und Datenmissbrauch adressieren, sondern auch die Verhinderung von Diskriminierung und die Förderung fairer sowie inklusiver digitaler Räume zum Ziel haben. Die enge Verschränkung von Effizienz und Sicherheit darf somit nicht zur Unsichtbarmachung sozialer Ungleichheiten führen, sondern muss durch eine demokratische, geschlechter- und datenschutzsensible Risikobewertung ergänzt werden, um den Schutz vulnerabler Gruppen zu gewährleisten und digitale Souveränität inklusive sozialer Gerechtigkeit zu sichern (vgl. Kapitel 2, Einleitung; Europäische Kommission, 2022, siehe Anhang A).

Die Rhetorik von Meta reproduziert hier eine technokratische Machtlogik, die ökonomische und operative Leistung über gesellschaftliche Verantwortung stellt und dadurch bestehende Machtasymmetrien verschärft – ein Ansatz, der mit den normativen Zielen der EU-Digitalpolitik grundlegend in Konflikt steht.

5.3 Fortschritts- und Innovationsframe

Indem Meta algorithmische Moderation als notwendigen und logischen Schritt darstellt, mit dem gesellschaftliche und technische Herausforderungen bewältigt werden sollen, rahmt sie diese als technologischen Fortschritt (vgl. Crawford, 2021: 8). Dieses Innovationsnarrativ betont „Innovation“, „Effizienzsteigerung“ und „Zukunftsfähigkeit“ und dient sowohl der Legitimation als auch der Abwehr kritischer Stimmen, die sonst als innovationsfeindlich gelten könnten (vgl. Crawford, 2021: 217).

Das Fortschrittsnarrativ überschneidet sich dabei mit Effizienz- und Sicherheitsrahmungen, die algorithmische Systeme als moralisch verantwortliche Schutzmechanismen und effiziente Lösungen darstellen (siehe Kapitel 5.2). Ebenso stehen Fortschritts-, Effizienz-

und Sicherheitsframes in Verbindung mit Objektivitäts- und Neutralitätsframes, die Algorithmen als technisch neutrale und unvoreingenommene Instrumente inszenieren. Dennoch grenzt sich der Fortschrittsframe dadurch ab, dass er vor allem die kulturelle und gesellschaftliche Akzeptanz technologischer Neuerungen betont und technologische Innovation als unvermeidliche und erwünschte Entwicklung darstellt, während Effizienz- und Sicherheitsframes stärker operative Notwendigkeiten und moralische Legitimation hervorheben.

Aufbauend auf den Überlegungen von Noble (2018) und Klonick (2018) (siehe Kapitel 1.3.2, 2.1) lässt sich zeigen, dass alle Frames gemeinsam dazu beitragen, regulatorische Kritik zu entwerten und bestehende Machtstrukturen in digitalen Plattformen zu festigen, indem sie algorithmische Moderation als alternativlos, gerecht und modern positionieren. Dadurch erschweren sie eine kritische gesellschaftliche Auseinandersetzung und die Durchsetzung einer digitalen Souveränität der EU.

5.3.1 Blogpost: Technologische Fortschritte in Metas Kommunikationsstrategie

Bereits zu Beginn des analysierten Blogposts von Pesenti (2018: Abs. 1, siehe Anhang B) wird von Technologien gesprochen, die durch ihren „Fortschritt“ Leben verändern und neue digitale Erfahrungen ermöglichen sollen. Für deren Entwicklung seien grundlegende Investitionen und strategische Ausbauten der technologischen Visionen notwendig. Ein großer Teil des Blogposts widmet sich der ausführlichen Aufzählung und Erklärung neuer Technologien, die Meta entwickelt hat. Formulierungen wie „hochoptimierbar“, „schneller“, „nützlicher“, „intelligenter“ oder „Durchbruch“ betonen dabei wiederholt die Bedeutung von Fortschritt und Innovation (ebd.). Diese fortschrittsorientierte Sprache inszeniert technische Entwicklung als durchweg positiven und notwendigen Prozess, wodurch technologische Innovationen als Kern der Unternehmensidentität gerahmt werden.

Zudem wird technologische Innovation als unabdingbar dargestellt, indem Forscher und Ingenieure als gezwungen beschrieben werden, neue Techniken zu entwickeln, wobei KI ihnen bei der Arbeit unterstützend zur Seite stehe. Die häufige Bezugnahme auf die Zukunft verstärkt den Eindruck, dass KI und technischer Fortschritt die „Grundlage“ für kommenden Wandel bilden (ebd.: Abs. 6). Diese Rhetorik personalisiert und vermenschlicht die technologische Entwicklung, indem Forscher:innen als getriebene Akteure und KI als unterstützende Kraft inszeniert werden. Dadurch wird der Eindruck eines

unvermeidlichen Innovationsdrucks erzeugt, der sowohl sozial-ökonomische Zwänge als auch politische Interessen hinter der Technologieentwicklung überdeckt.

Abschließend unterstreicht ein Beispiel, in dem ein KI-System die Blutspende in verschiedenen Ländern vereinfacht habe, den positiven Effekt von Innovation. Dieses Beispiel ist so platziert, dass es im Gedächtnis der Leser:innen haften bleibt und gleichzeitig ethische Bedenken, die im gleichen Abschnitt thematisiert werden, überschattet. Das innovationsrechtfertigende Narrativ suggeriert zudem, dass Meta trotz möglicher ethischer Risiken alles unternehme, um negative Auswirkungen zu minimieren (z.B. durch Investitionen in Mitarbeiter, Prozesse und Tools) — ethische Bedenken würden jedoch immer bestehen (ebd.: Abs. 12-14). Hier nutzt Meta gezielt positive Praxisbeispiele, um ethische Risiken zu relativieren und sich als verantwortungsbewusster Akteur darzustellen, ohne mögliche Schattenseiten vertieft zu reflektieren

Insgesamt zielt Metas Fortschrittsrahmung hier darauf ab, technologische Neuerungen als zeitlichen und moralischen Imperativ zu etablieren. Fortschritt wird dadurch als unabdingbar, erwünscht und unkritisch darstellbar inszeniert. Diese Strategie erschwert eine umfassende kritische Auseinandersetzung, da technologische Innovationen mit Zukunftsfähigkeit und gesellschaftlichem Nutzen aufgeladen sind und Gegenpositionen als innovationsfeindlich delegitimiert werden. Diese Diskursive Strategie Metas wird auch in der US-Senatsanhörung von 2020 fortgesetzt.

5.3.2 Senatsanhörung: Fortschrittsrhetorik Zuckerbergs

In der US-Senatsanhörung 2020 nutzt Mark Zuckerberg zentrale Elemente des Innovationsnarrativs, um algorithmische Moderation als fortschrittlichen, unverzichtbaren Mechanismus darzustellen. Er betont Meta's proaktive Verantwortung zur Sicherung der Wahlintegrität und kommuniziert einen kontinuierlichen Verbesserungsprozess im Umgang mit gesellschaftlichen Herausforderungen wie Wahlmanipulation (U.S. Senate Judiciary Committee, 2020: 26:33, siehe Anhang B). Diese Rhetorik verschiebt Verantwortung für Fehler in die Zukunft und relativiert so Forderungen nach sofortiger Rechenschaft. Technologischer Fortschritt wird als alternativlose Lösung präsentiert, wodurch normative Kritik jenseits des Fortschrittsdiskurses marginalisiert wird.

Zuckerbergs Verweis auf Nutzer:innenzahlen, wie die 140 Millionen Besucher:innen des Voting Information Centers (U.S. Senate Judiciary Committee, 2020: 27:32, siehe Anhang B), verstärkt dieses Fortschrittsbild zusätzlich durch die Betonung gesellschaftlicher Wirksamkeit. Die Fokussierung auf den gesellschaftlichen „Impact“ wird dabei

strategisch genutzt, um regulatorische Eingriffe als mögliche Hemmnisse für eine vermeintlich ausschließlich positive Entwicklung zu rahmen – ein diskursives Muster, das bereits im analysierten Blogpost von Pesenti (2018: Abs. 12, siehe Anhang B), deutlich wird. So zeigt sich, dass Meta sowohl in der Außendarstellung als auch in politischen Kontexten wiederholt dieselbe Strategie verfolgt, um ethische Risiken zu relativieren und Gegenpositionen zu delegitimieren.

Aussagen wie „So we continue to improve with each election“ zementiert darüber hinaus das Bild eines ethisch verantwortlichen, fortwährenden Fortschritts, der unvermeidbare Fehler einplant und zugleich Kritik an Machtstrukturen umgeht (U.S. Senate Judiciary Committee, 2020: 26:33, siehe Anhang B). Gerade durch dieses Narrativ des stetigen Fortschritts und der permanenten Innovation wird der Eindruck erzeugt, dass kontinuierliche Entwicklung unausweichlich und grundsätzlich wünschenswert sei. Dadurch werden bestehende Machtverhältnisse verdeckt, während Kritik an strukturellen Problemen als Teil eines fortlaufenden Optimierungsprozesses neutralisiert und auf technische Lösbarkeit reduziert wird.

Letztlich entschärft diese Fortschrittsrhetorik Kritik durch den Verweis auf stetige technologische Verbesserung, während soziale und politische Folgen algorithmischer Entscheidungen kaum thematisiert werden. Innovation wird so als positiver Schutzmechanismus inszeniert, der normative Debatten und demokratische Kontrolle in den Hintergrund drängt. Dieses Beispiel verdeutlicht exemplarisch, wie technologische Macht diskursiv abgesichert und regulatorische Kritik marginalisiert wird. Auch im Instagram Statement von Zuckerberg, lässt sich diese diskursive Legitimierungsstrategie wiederfinden.

5.3.3 Instagram-Statement: Neuausrichtung des Frames

Das Statement bedient zentrale Elemente des Fortschritts- und Innovationsframes, um Metas Neuausrichtung in der Inhaltsmoderation rhetorisch zu legitimieren. Gleich zu Beginn betont er die Rückkehr zu den Wurzeln der freien Meinungsäußerung und verweist auf die komplexen Systeme, die Meta aufgebaut habe (Zuckerberg, 2025: Abs. 1–5, siehe Anhang B). Diese Rhetorik verbindet eine Rückbesinnung auf ursprüngliche Plattformwerte mit dem Status quo technologischer Kompetenz. Die Beschreibung der „vielen komplexen Systeme“ dient nicht nur als Nachweis technologischer Fortschrittlichkeit, sondern inszeniert algorithmische Moderation als Produkt einer kontinuierlichen,

evolutionären Entwicklung. Die vermeintliche Komplexität wird nicht als Problem, sondern als Fortschrittsbeleg dargestellt.

Des Weiteren grenzt sich Zuckerberg ausdrücklich von früheren Moderationspraktiken ab, die er als zu restriktiv und innovationshemmend beschreibt. Er zeichnet ein Bild einer „überregulierten“ Vergangenheit, in der eine ursprünglich inklusive Bewegung zunehmend zur Unterdrückung genutzt wurde (ebd.: Abs. 9). Während die früheren Quellen vor allem den technologischen Fortschritt und die Innovationskraft Meta als unverzichtbaren, positiven Entwicklungsprozess inszenieren, legt Zuckerberg in der Senatsanhörung zusätzlich einen stärkeren Fokus auf die Abgrenzung von „überregulierten“ und „restriktiven“ Moderationspraktiken der Vergangenheit. Dadurch verschiebt sich die Rhetorik von der bloßen Hervorhebung technologischer Neuerungen hin zu einer Verteidigung einer offenen, dynamischen Plattform, bei der Kritiker:innen als Fortschrittsverhinderer diskreditiert und Meta als Vorkämpfer für Innovation und Freiheit positioniert wird.

Gesellschaftliche Umbrüche, insbesondere politische Ereignisse wie jüngste Wahlen, werden von Zuckerberg direkt mit einem technologischen Reformbedarf verknüpft: Diese Entwicklungen kennzeichnet er als „kulturellen Wendepunkt“, der den Fokus auf Fehlerreduktion und interne Reformen nötig mache (ebd.: Abs. 6). Technologie erscheint so als flexible, lernfähige Antwort auf gesellschaftliche Herausforderungen, deren Wandel den Fortschritt legitimiert.

International positioniert Zuckerberg technologischen Fortschritt explizit als Produkt westlich-liberaler Werte und hebt dabei die USA als Innovationsmotor hervor. Europa kritisiert er scharf wegen institutionalisierter Zensur, die Innovation erschwere, und ruft zur globalen Zusammenarbeit unter US-Führung auf (ebd.: Abs. 13). Dieses binäre Narrativ entwertet europäische Regulierungsansätze als bürokratisch und rückständig und legitimiert eine innovationsfreundliche US-Zentrierung.

Zum Abschluss erzeugt die Aussicht auf ein „nächstes Kapitel“ einen zukunftsorientierten Ton, der technologischen Wandel als dynamischen, stetigen und legitimierten Prozess rahmt (ebd.: Abs. 16). Die zukunftsgerichtete Sprache verstärkt das Bild von technologischem Fortschritt als positiver, unverzichtbarer Kraft.

Die Analyse hat verdeutlicht, wie technologische Innovation strategisch eingesetzt wird, um legitimatorisch einen Bruch mit der Vergangenheit zu markieren und gleichzeitig Kritik an Machtstrukturen subtil zu entkräften. Der Fortschrittsframe wird so zum Schlüsselement der Selbstdarstellung und Machtabstärkung des digitalen Plattformunternehmens.

5.3.4 EU- und Geschlechterperspektive zum Frame Fortschritt & Innovation

Im Rahmen der europäischen Digitalpolitik wird Innovationsförderung ausdrücklich anerkannt, jedoch stets verbunden mit sozialer Verantwortung und Grundrechtsschutz. Dies manifestiert sich exemplarisch im EU AI Act, der Unternehmen wie Meta verpflichtet, die Funktionsweise ihrer KI-Systeme offen zu legen, Risiken systematisch zu adressieren und gesellschaftliche Auswirkungen transparent zu machen (Europäisches Parlament, 2023; siehe Kapitel 2.3.5). Die EU verfolgt damit konsequent das Leitbild, dass technologische Entwicklung nicht als autonomen Fortschritt zu verstehen ist, sondern im Dienst demokratischer Kontrolle und kollektiver digitaler Souveränität stehen muss (siehe Kapitel 2.3.4; 2.3.3).

Meta hingegen nutzt den Fortschrittsframe, um regulatorische Anforderungen als innovationshemmend, bürokratisch und wettbewerbsfeindlich zu diskreditieren. Innovation erscheint dabei als fragile, nur durch größtmögliche Freiheit gewährleistete Weiterentwicklung, während gesellschaftliche Risiken marginalisiert und demokratische Steuerung rhetorisch relativiert werden. Diese Strategie bedeutet, dass Meta Regulierung diskursiv als Gegenmodell zu Innovation entwirft, was in einem klaren Gegensatz zu den Zielen der EU steht, die eine verantwortungsvolle Regulierung und die Eindämmung monopolistischer Plattformstrukturen fordern (vgl. Europäische Kommission, 2022, siehe Anhang A; Wagner et al., 2022; siehe Kapitel 2.3).

Aus Sicht einer geschlechteranalytischen Perspektive ist der Fortschrittsframe vor allem deshalb problematisch, weil er Diversität rhetorisch vereinnahmt. Inklusion wird zum Innovationslabel umgedeutet, während die zugrunde liegenden Ausschlussmechanismen unangetastet bleiben. Die positive Aufladung technischer Entwicklung fördert symbolisch eine männlich konnotierte Innovationskraft, die als dominant und überlegen inszeniert wird. Während Diversität auf rhetorischer Ebene vereinnahmt wird, gilt ihre tatsächliche Umsetzung vielfach als hinderlich für technologische Innovation und bleibt damit marginalisiert.

Die Gleichstellungspolitik der EU fordert eine konsequente und kritische Einbindung von Genderaspekten in die Entwicklung und Regulierung digitaler Technologien, um die Reproduktion und Verstärkung struktureller Geschlechterungleichheiten entgegenzuwirken (Europäische Kommission, 2022, siehe Anhang A; siehe Kapitel 2.3.5). Nur durch eine solche kritische Perspektive können Innovationen nicht nur technisch, sondern auch sozial gerecht und inklusiv gestaltet werden, was zentrale Voraussetzung für eine demokratisch legitimierte digitale Souveränität ist.

5.4 Vermeidung von Verantwortung (Responsibility Shifting)

Obwohl Meta die Kontrolle über Infrastruktur und Entscheidungsprozesse innehat, externalisiert das Unternehmen häufig die Verantwortung für soziale, politische und ethische Folgen – sei es auf Nutzer:innen, staatliche Instanzen oder vermeintlich neutrale Algorithmen (vgl. Gillespie, 2018; Gorwa & Garton Ash, 2020). So wird technische Komplexität und individuelle Eigenverantwortung der Community betont, um unternehmerische Haftung abzuschwächen.

Der Responsibility-Shifting-Frame ist eng mit anderen Strategien verwoben, etwa der Behauptung von Neutralität („Wir sind nur die Plattform“) und der Delegation an automatisierte Systeme oder externe Partner (vgl. U.S. Senate Judiciary Committee, 2020: 01:08:23, siehe Anhang B). Fehler und problematische Inhalte werden als unvermeidbare Begleiterscheinungen komplexer Technik dargestellt, während regulatorische Eingriffe regelmäßig als innovationshemmend abgewertet werden (vgl. van Dijck, Poell & de Waal, 2018: 21–28). In der Kombination mit zuvor analysierten Frames sichert Meta so seine diskursive Positionierung ab.

Im Gegensatz zu den vorherigen Frames verlagert diese Strategie Deutungsmacht, indem Meta die technische Kontrolle intern wahrt, jedoch Verantwortung diffus oder entpolitisiert nach außen delegiert. Dies erschwert nicht nur soziale Rechenschaft, sondern verstärkt auch bestehende Machtasymmetrien zwischen Plattform, Nutzer:innen und Regulierungsinstanzen (vgl. Helberger et al., 2017: 10–11).

Für die Beantwortung der Forschungsfrage ist diese Verantwortungsdiffusion von besonderer Bedeutung: Durch die Auslagerung auf externe Akteur:innen und die Hervorhebung technischer Komplexität erschwert Meta die Durchsetzung digitaler Souveränität in der EU. Die Dezentralisierung der Kontrollbefugnisse erzeugt intransparente Strukturen, die regulatorische Eingriffe erschweren und unklare Verhältnisse im Umgang mit geschlechterspezifischer und anderer sozialer Diskriminierung schaffen. Die daraus resultierende Verantwortungszersplitterung verstärkt bestehende Ungleichheiten und behindert demokratische Kontroll- und Regulierungsprozesse.

Dieser Frame bildet die Grundlage für die folgende Analyse der konkreten Ausprägungen von Responsibility Shifting bei Meta und deren Funktionen in digitalen Macht- und Verantwortungsstrukturen.

5.4.1 Blogpost: Verantwortungsdiffusion durch Open Source und Kollaboration

Im Blogpost von Pesenti (2018: Abs. 2, siehe Anhang B) hebt Meta die Offenheit und Zusammenarbeit bei der Entwicklung seiner KI-Frameworks hervor. Diese werden als Beiträge zu Innovation und Transparenz inszeniert, während zugleich auf die inhärente Komplexität von KI-Systemen verwiesen wird. So relativiert sich die Verantwortung für problematische Folgen implizit, indem ethische Herausforderungen als fortlaufendes, gemeinschaftliches Problem dargestellt werden. Diese Strategie verteilt Verantwortung breit auf eine diffuse Akteursgruppe – bestehend aus der globalen Forschung, Entwickler:innen und der Open-Source-Community. Meta tritt dabei als Werkzeuggeber auf, nicht als ethisch handelndes Unternehmen, wodurch direkte Verantwortung bewusst ausgeklammert bleibt (ebd.: Abs. 2–3).

Ein konkretes Beispiel ist die Bereitstellung von Open-Source-Frameworks auf Facebook.ai, mit dem argumentiert wird, dass problematische KI-Effekte gemeinschaftlich bearbeitet werden müssen. Verantwortung wird so zum „Community-Problem“, nicht zum Versagen des Unternehmens. Die Betonung technischer Komplexität, etwa des StarCraft-Bots in einer „komplexen und teilweise beobachtbaren Umgebung“, verstärkt die Verschiebung der Verantwortung zusätzlich (Pesenti, 2018: Abs. 10–11, siehe Anhang B). Dieses Framing verschleiert Metas erheblichen Einfluss auf gesellschaftliche Normen und Realitäten, da Designentscheidungen und technische Implementierungen weitreichende Folgen haben – ohne dass dem Unternehmen uneingeschränkte Verantwortlichkeit zugeschrieben wird. Meta nutzt im Blogpost gezielt die Strategie der Verantwortungsdiffusion, um sich aus direkter ethischer Verantwortung zu entziehen und kritische Debatten in technische, gemeinschaftliche Herausforderungen umzudeuten. So wird Verantwortung verlagert und problematische Praktiken legitimiert.

5.4.2 Senatsanhörung: Ablehnung redaktioneller Verantwortung

Während Meta im Blogpost Verantwortung diffundiert und ethische Debatten technisiert, nutzt Zuckerberg in der Senatsanhörung die Strategie, Verantwortung für Inhaltsmoderation rhetorisch zu entpolitisierten und abzuwälzen. Zuckerberg lehnt es bspw. ab, dass Facebook („Meta“) die Rolle eines „Wahrheitswächters“ übernehmen sollte, und verweist auf Nutzer:innen Rückmeldungen, die einerseits ein starkes Bedürfnis nach Desinformationsbekämpfung zeigen, andererseits aber auch eine Ablehnung einer alleinigen Wahrheitskontrolle durch Meta ausdrücken. Diese Verantwortung bezeichnet Zuckerberg persönlich als unangemessen (U.S. Senate Judiciary Committee, 2020: 01:21:09, siehe

Anhang B). Mit dieser Ablehnung übernimmt Meta formal keine redaktionelle Verantwortung, obwohl seine Algorithmen faktisch zentrale Macht über Sichtbarkeit und Reichweite von Inhalten besitzen. Stattdessen präsentiert sich das Unternehmen als neutraler Vermittler, der den „Willen der Community“ abbildet. Dadurch wird der Anschein erweckt, unternehmerische Macht und Verantwortung lägen bei den Nutzer:innen, obwohl sie tatsächlich beim Unternehmen verbleiben. (ebd.: 01:21:09).

Zudem propagiert Zuckerberg ein Narrativ, wie auch in dem F8-Blogpost (Pesenti, 2018: Abs. 2-3, siehe Anhang B), wissenschaftlicher Offenheit, indem er erklärt, Meta stelle Forschungsergebnisse aus Kooperationen mit externen Wissenschaftler:innen öffentlich und unverändert bereit (ebd.: 01:56:54). So wird auch hier Verantwortung für potenziell problematische Inhalte oder Interpretationen auf die akademische Gemeinschaft verlagert, während Meta seine wissenschaftliche Autorität wahrt und sich einer aktiven Steuerungsfunktion entzieht.

Analytisch zeigt sich, dass Ablehnung redaktioneller Verantwortung, die Betonung von Neutralität und wissenschaftlicher Offenheit Teil eines umfassenden Diskurses der Entpolitisierung und Externalisierung von Verantwortung sind. Diese Rhetorik zielt darauf ab, regulatorische und gesellschaftliche Kontrollansprüche als problematisch darzustellen und die Rolle des Unternehmens als verbindliche Instanz zu unterminieren (vgl. van Dijck, Poell & de Waal, 2018). Gleichzeitig wird die faktische Deutungshoheit und Inhaltssteuerung hinter einer vermeintlichen Neutralität verborgen. Diese Diskursstrategie der Ablehnung redaktioneller Verantwortung kann als Teil eines größeren Musters verstanden werden, in dem Meta Verantwortung zunehmend externalisiert und entpolitisiert – ein Muster, das sich auch in der Verlagerung von Moderationsaufgaben auf die Community im Instagram-Statement 2025 manifestiert.

5.4.3 Instagram-Statement: Delegation an Community-basierte Systeme

Mark Zuckerbergs Ankündigung des Community-Notes-Programms markiert einen bewussten Strategiewechsel in Metas Moderationspolitik: Die Verantwortung für die Bewertung von Inhalten wird zunehmend auf die Nutzer:innen delegiert. Durch die Integration der Community in den Moderationsprozess tritt Meta formal als direkt verantwortlicher Akteur zurück und überträgt die Kontrolle auf kollektiv organisierte Systeme (Zuckerberg, 2025: Abs. 8, siehe Anhang B).

Diese Delegation ist Teil einer umfassenden Strategie der Externalisierung und Entpolitisierung von Verantwortung, die sich in verschiedenen kommunikativen Praktiken zeigt.

Indem Zuckerberg frühere Faktenprüfungsnetzwerke als „politisch voreingenommen“ diskreditiert, kanalisiert er Fehler und Kritik auf externe Akteure, während sich Meta rhetorisch als lernfähiger und reformorientierter Innovator positioniert.

Die Betonung, wie auch im F8-Blogpost (Pesenti, 2018: Abs. 2–3, siehe Anhang B), technischer Komplexität und Fehleranfälligkeit KI-gestützter Systeme fungiert dabei als legitimierende Erklärung: Fehlschläge gelten nicht als unternehmensinternes Versagen, sondern als unvermeidbare Grenzen technischer Innovation (Zuckerberg, 2025: 5, siehe Anhang B). So wird Verantwortung symbolisch an die Algorithmen und Nutzer:innen delegiert und gleichzeitig die unternehmerische Haftung abgeschwächt.

Wie bereits beschrieben, kündigt Zuckerberg an, dass künftig eine verstärkte Fokussierung auf schwerwiegende und illegale Inhalte erfolgen werde, während die Kontrolle über weniger gravierende Inhalte den Nutzer:innen und der öffentlichen Debatte überlassen bleiben soll (ebd.: Abs. 10). Diese Priorisierung stellt eine bewusste Form der Verantwortungsreduktion dar, durch die unternehmerische Entscheidungsverantwortung verschoben und eine effektive Regulierung erschwert wird.

Ein weiteres Beispiel für diese Strategie zeigt sich in der Verlagerung von Moderationsteams von Kalifornien nach Texas, die rhetorisch als Reaktion auf kulturelle Verzerrungen dargestellt wird. Hier werden unternehmerische Fehlentscheidungen indirekt dem kulturellen Kontext zugeschrieben, wodurch interne strukturelle Verantwortlichkeiten systematisch ausgeklammert werden (ebd.: Abs. 12).

Insgesamt inszeniert Meta sich als technisch neutraler Vermittler, während die sozialen und politischen Implikationen der Moderationspraxis diskursiv auf externe Partner, Nutzer:innen und komplexe technische Rahmenbedingungen ausgelagert werden. Community-basierte Moderationsprogramme fungieren als Mechanismen strategischer Verantwortungsdiffusion und erschweren dadurch eine kritische Auseinandersetzung sowie demokratische Kontrollmöglichkeiten (vgl. van Dijck, Poell & de Waal, 2018).

Im Kontext europäischer Digitalpolitik, insbesondere der Debatten um digitale Souveränität und den Schutz marginalisierter Gruppen, manifestiert sich diese Delegation als erneute Entpolitisierung von Macht und Verantwortung durch Meta. Dies unterstreicht die Notwendigkeit einer intensiveren Regulierung und Transparenz im Umgang mit algorithmischer Inhaltsmoderation.

5.4.4 EU- und Geschlechterperspektive zum Frame Responsibility-Shifting

Der Responsibility-Shifting-Frame, mit dem Meta Verantwortung für Fehler, problematische Inhalte oder ethische Herausforderungen auf Nutzer:innen, technische Systeme oder externe Partner abschiebt, kennzeichnet eine deutliche Ablehnung der Übernahme unternehmerischer und regulatorischer Verantwortung. Dieser Diskurs dient der Abwehr klarer Rechenschaftspflichten und entzieht sich einer direkten Auseinandersetzung mit den gesellschaftlichen Folgen der Plattformaktivitäten. Im europäischen Kontext stehen Meta damit die verbindlichen Anforderungen des Digital Services Act (DSA) und des AI-Act entgegen, die nicht nur eine transparente Offenlegung algorithmischer Prozesse, sondern auch eigenständige Risikoabschätzungen, Wirksamkeitskontrollen und umfassende Rechenschaftslegung von Plattformbetreibern fordern (siehe Kapitel 2.3.4, 2.3.5).

Die EU verfolgt mit diesen Rechtsrahmen explizit das Ziel, klar definierte Verantwortlichkeiten und Kontrollmechanismen zu etablieren, um strukturelle Machtasymmetrien zu adressieren und den Schutz marginalisierter Gruppen sowie Gleichstellungsziele durchzusetzen (siehe Kapitel 1, 2.3). Indem Meta jedoch Verantwortung diffundiert und auf diffuse, technische oder gemeinschaftliche Mechanismen delegiert, wird diese notwendige regulatorische Kontrolle erheblich erschwert. Diese Delegitimierung unternehmerischer Haftung perpetuiert bestehende Diskriminierungen, insbesondere entlang geschlechtsspezifischer Linien, und steht im direkten Widerspruch zu den normativen Ansprüchen der EU an digitale Souveränität und soziale Gerechtigkeit (vgl. Gillespie, 2018; van Dijck, Poell & de Waal, 2018; siehe Kapitel 2.3.1).

Wird der Responsibility-Shifting-Frame genderbezogen betrachtet, zeigt sich, dass Verantwortung ungleich verteilt wird: Frauen* und intersektional diskriminierte Gruppen tragen die Hauptlast, da sie zusätzliche Ressourcen aufbringen müssen, um sich gegen diskriminierende Inhalte zu wehren. Die Verschiebung von Verantwortung auf Nutzer:innen und Technik verhindert eine transparente und faire politische Kontrolle und unterminiert damit sozial gerechte Regulierung, wie sie die EU in ihrer Gleichstellungs- und Antidiskriminierungspolitik fordert (siehe Kapitel 2.3.5).

Die EU-Regulierungsansätze betonen, dass algorithmische Systeme nicht nur technisch, sondern ebenso sozialpolitisch gestaltet sind und deshalb einer demokratischen Kontrolle und klaren Verantwortungszuschreibung unterliegen müssen, um geschlechter- und gruppenspezifische Diskriminierungen aktiv zu bekämpfen (siehe Kapitel 2.3.4, 2.3.5). Indem Meta diese Verantwortung fragmentiert und externalisiert, wird die notwendige

gesellschaftliche und politische Auseinandersetzung mit algorithmischer Diskriminierung systematisch erschwert. Dies trägt dazu bei, dass marginalisierte Gruppen – insbesondere Frauen* und intersektional benachteiligte Personen – weiterhin strukturell benachteiligt bleiben, da die erforderlichen regulatorischen Interventionen blockiert oder verwässert werden.

Diese Verantwortungsexternalisierung ist somit nicht nur eine rhetorische Strategie zur Abwehr regulatorischer Kontrolle, sondern auch ein Mechanismus, der bestehende Machtungleichheiten im digitalen Raum reproduziert und demokratische Regulierungsansprüche konterkariert. Die EU-Politik sieht in der Übernahme klarer unternehmerischer Verantwortung und der Schaffung transparenter, inklusiver Kontrollmechanismen einen zentralen Hebel zur Förderung digitaler Souveränität und sozialer Gerechtigkeit – deren Realisierung durch das Responsibility Shifting von Meta aktuell gefährdet wird (Gillespie, 2018; van Dijck, Poell & de Waal, 2018; siehe Kapitel 2.3.1).

5.5 Zwischenfazit

Zusammenfassend konnten die Ergebnisse der Critical Frame Analysis (CFA) zu den verschiedenen Frames, mit denen Meta seine algorithmische Inhaltsmoderation und damit verbundene Machtverhältnisse legitimiert, herausgearbeitet werden. Im Zusammenspiel mit den zuvor beschriebenen infrastrukturellen und normativen Machtstrukturen zeichnet sich ein vielschichtiges Bild strategischer Verantwortungsverschiebung und Legitimationsmuster ab.

Die Analyse der Kommunikationsstrategien macht deutlich, dass Meta durch Frames wie Neutralität, Effizienz, Fortschritt und Responsibility Shifting nicht nur regulatorische Eingriffe abwehrt, sondern zugleich technokratische Männlichkeitsmuster stabilisiert und gesellschaftliche Machtasymmetrien reproduziert. Algorithmische Systeme werden dabei als technisch rational und weitgehend fehlerunabhängig inszeniert, was individuelle und kollektive Kontrolle erschwert. Gleichzeitig wird Verantwortung diffus verteilt – insbesondere auf Nutzer:innen oder automatisierte Systeme –, was die demokratische Rechenschaftspflicht massiv unterminiert.

In der strukturellen Machtanalyse (siehe Kapitel 4) wurde bereits Metas ambivalente Machtposition als Plattformunternehmen verdeutlicht: Einerseits kontrolliert es zentrale Infrastrukturen und Datenströme, andererseits verhandelt es regulatorische Zwänge und gesellschaftliche Erwartungen mit hoher Diskursmacht. Die in diesem Kapitel

herausgearbeiteten Frames ergänzen diese strukturelle Perspektive um eine tiefere Ebene der Bedeutungszuschreibung und Legitimationspraxis.

Insgesamt macht die Analyse deutlich, dass Meta durch die Kombination aus technischer Infrastrukturmacht und diskursiver Rahmung die digitale Souveränität der EU herausfordert, indem regulatorische Anforderungen entpolitisiert, Verantwortung vermieden und bestehende soziale Ungleichheiten, insbesondere im Bereich der Geschlechtergerechtigkeit, stabilisiert werden. Diese Befunde unterstreichen die Bedeutung einer verstärkten, kritischen Regulierung und gesellschaftlichen Auseinandersetzung. Zugleich wird sichtbar, dass die einzelnen Frames nicht isoliert wirken, sondern sich gegenseitig verstärken und dadurch ähnliche Konsequenzen zeitigen, was die Persistenz dieser Machtstrukturen zusätzlich erklärt.

Im folgenden Kapitel 6 wird vor diesem Hintergrund die Implikation dieser Erkenntnisse für die weiterführende Diskussion zu reflektiert.

6 Diskussion der Ergebnisse

6.1 Zusammenführung der Erkenntnisse

Ziel dieses Kapitels ist es, die zentralen Erkenntnisse der strukturellen Machtanalyse (siehe Kapitel 4) mit der Critical Frame Analyse (siehe Kapitel 5) zusammenzuführen, um im folgenden Kapitel 7 dann die Forschungsfrage zu beantworten.

Die in Kapitel 5 ermittelten Frames – *Neutralität & Objektivität, Effizienz & Sicherheit, Fortschritt & Innovation* sowie *Responsibility Shifting* – bilden das diskursive Fundament, mit dem Meta algorithmische Moderation als alternativlose, sachlich neutrale und zukunftsorientierte Lösung positioniert (vgl. Tabelle 2 in Anhang B). Diese Rahmungen wirken nicht isoliert, sondern in einem komplementären Zusammenspiel: Effizienz- und Sicherheitsargumente schaffen Legitimität für technologische Ausweitungen, die durch Fortschrittsrhetorik flankiert und durch Verantwortungsauslagerung gegen Kritik abgeschirmt werden.

Entscheidend ist, dass diese diskursive Legitimation stets auf die strukturelle Plattformmacht aufsetzt, wie sie in Kapitel 4 beschrieben wurde: Meta kontrolliert maßgebliche digitale Infrastrukturen, algorithmische Bewertungs- und Sichtbarkeitsregeln sowie normative Standards für Inhalte (vgl. Staab 2019, Kapitel 4.2–4.5). Damit verschiebt sich die Festlegung gesellschaftlicher Kommunikationsnormen weg von demokratisch legitimierten Verfahren hin zu einem privatwirtschaftlich gesetzten Ordnungsrahmen.

Die Critical Frame Analyse zeigt, wie diese materiellen Machtpositionen rhetorisch stabilisiert werden: Indem technische Systeme als wertfrei, unvermeidlich und innovationsgetrieben inszeniert werden, erscheinen regulatorische Eingriffe als politisch motiviert oder innovationshemmend. Dadurch entsteht eine wechselseitige Verstärkung: Die technische Kontrolle Metas setzt Grenzen des Sag- und Zeigbaren, während die Diskurse diese Grenzen als notwendig und unpolitisch normalisieren.

Dieses Zusammenspiel aus materieller Infrastrukturkontrolle und diskursiver Absicherung bildet die Grundlage für die spezifischen Wirkungen, die in den folgenden Unterkapiteln auf die digitale Souveränität (Kapitel 6.2) und die Reproduktion von Geschlechterungleichheiten (Kapitel 6.3) vertieft diskutiert werden. Während Kapitel 6.3 den Fokus auf die Implikationen dieser Machtkonstellationen für Gendergerechtigkeit legt, diskutiert Kapitel 6.2 zunächst, wie sich das Wechselspiel von struktureller Macht und diskursiver Legitimation im politischen Handlungsraum der Europäischen Union auswirkt, insbesondere vor dem Hintergrund der Ergebnisse der Analysen (siehe Kapitel 4, 5).

6.2 Digitale Souveränität und die Rolle der EU

Die Analyse in Kapitel 4 verdeutlicht, dass Meta weit über die Rolle eines reinen Dienstleisters hinaus als zentraler Machtakeur über digitale Kommunikationsinfrastrukturen agiert. Diese strukturelle Macht manifestiert sich wesentlich in folgenden Dimensionen: Metas Macht ist auf einem Dispositiv gegründet, das Infrastruktur, Regeln und Wissen strategisch verbindet; gouvernementale Steuerungspraktiken entwickelt, die Formen von politischer Kontrolle und Eigenverantwortung miteinander verzahnen; und mittels Normalisierungsmechanismen gesellschaftliche Kommunikationsstandards durchsetzt und stabilisiert – ganz im Sinne der foucaultschen Analyse der Macht (siehe Kapitel 4). Diese materiell fundierte Machtbasis verschiebt Entscheidungskompetenzen zunehmend von demokratisch legitimierten Institutionen hin zu privatwirtschaftlichen Plattformunternehmer:innen. Die EU steht hier vor der Herausforderung, dass regulatorische Instrumente wie der Digital Services Act (DSA), der Digital Markets Act (DMA), die DSGVO und der AI-Act zwar rechtlich wirksame Vorgaben bieten, ihre praktische Durchsetzungskraft jedoch stark von der Kooperationsbereitschaft und technischen Offenheit der Plattformen abhängt (vgl. Hulkó et al., 2025: 1-2).

Die Erkenntnisse aus der CFA zeigen ergänzend, wie Meta diese strukturelle Macht durch strategisch diskursive Rahmen absichert: Frames wie Neutralität & Effizienz sowie Fortschritt & Innovation fungieren als rhetorische Schutzschilde gegen regulatorische

Einmischung, indem potenzielle Kritik als innovationsfeindlich und technisch naiv abgetan wird. Dieses Legitimationsmuster erschwert den politischen Handlungsspielraum der EU, indem es die Diskussion um Plattformkontrolle auf eine vermeintlich wertfreie technologische Sachzwangsebene verlagert (siehe Kapitel 5).

Das Zusammenspiel von materieller Plattformmacht und diskursiver Abschirmung schafft ein wirkmächtiges Dispositiv: Auf technischer Ebene verhindern geschlossene, proprietäre Infrastrukturen und Black-Box-Algorithmen eine wirksame regulatorische Einflussnahme (siehe Kapitel 4.2, 4.5) während auf sozialer Ebene die Inszenierung als objektiv zwingend und alternativlos die Diskussion über digitale Regulierung und geschlechtergerechte Teilhabe entpolitisiert (siehe Kapitel 5, 6.3).

Diese Mechanismen wirken nicht losgelöst, sondern in einem komplexen, transnationalen Machtgeflecht, in dem die US-Regierung eine bedeutende Rolle spielt. Ihre industrieorientierte Politik setzt stark auf Deregulierung und schützt amerikanische Tech-Giganten auch im globalen Wettbewerbsumfeld, indem sie bei EU-Regulierungsbestrebungen wie dem AI-Act diplomatisch interveniert, um diese abzuschwächen oder zu verzögern (Pamuk, 2025; Santos, 2025). Diese transatlantische Dynamik verstärkt die asymmetrische Machtverteilung zugunsten US-amerikanischer Konzerne wie Meta und erschwert der EU die Realisierung effektiver digitaler Souveränität.

Die europäische Digitalpolitik sucht dagegen nach Wegen, sich gegenüber dieser Machtfülle zu behaupten und die digitale Souveränität zu stärken. Die regulatorischen Eckpfeiler – DSA, DMA, DSGVO und AI-Act – bilden zwar eine beispiellose gesetzliche Infrastruktur mit zum Teil extraterritorialer Wirkung (vgl. Hulkó et al., 2025: 6). Doch insbesondere bei der praktischen Umsetzung und Durchsetzung stoßen diese Regulierungen an Grenzen. Gründe hierfür sind neben internationalen Machtverhältnissen auch:

- Uneinheitliche Handhabung und Umsetzung in den Mitgliedstaaten, die hinsichtlich Kapazitäten und politischen Prioritäten stark variieren (vgl. ebd.: 6).
- Technische Abhängigkeiten von proprietären Plattforminfrastrukturen, die daten- und Algorithmus basierte Kontrollzugänge erschweren (vgl. Staab, 2019: 134; siehe Kapitel 4).
- Strategische Diskursverschiebungen von Meta, die regulatorischen Initiativen als innovationshemmend empfinden lassen und so gesellschaftliche Unterstützung schwächen (siehe Kapitel 5).

Vor diesem Hintergrund zeigt sich, dass Digitale Souveränität nicht nur technologische Souveränität über Infrastruktur und Daten umfasst, sondern auch eine politische Durchsetzungsfähigkeit, die in der Lage ist, über Diskurs und Regulierung gegen transnationale Plattformmacht anzutreten (siehe Kapitel 1.1).

Besonders anschaulich wird dies am Beispiel der Umsetzungspflichten im DSA: Während die Verordnung klare Transparenzanforderungen an sehr große Plattformen festlegt, sind die tatsächlichen Möglichkeiten für Regulierungsbehörden, algorithmische Entscheidungsprozesse zu kontrollieren, durch die technische Komplexität und fehlende Zugangsmöglichkeiten faktisch begrenzt (vgl. Hulkó et al., 2025: 8). Hier zeigt sich, wie Metas Infrastrukturen zur faktischen Auslagerung politischer Steuerung führen – eine zentrale, in Kapitel 4 erarbeitete Problemdimension.

Die digitale Souveränität der EU wird durch das enge Zusammenspiel proprietärer technischer Infrastrukturen und deren diskursiver Legitimationsstrategien strukturell eingeschränkt. Die Analyse macht deutlich, dass nachhaltige Regulierung der EU nur dann realisierbar ist, wenn diese gleichzeitig Zugang und Kontrolle über wesentliche technische Bausteine (bspw. Datenportabilität, Transparenz der Algorithmen, API-Zugänge) sicherstellen kann und die diskursive Legitimation bestehender Machtstrukturen kritisch dekonstruiert wird, um den politischen Handlungsspielraum wiederherzustellen.

Durch diese doppelte Herangehensweise kann die EU ihre digitale Souveränität nicht nur formal juristisch, sondern auch praktisch und gesellschaftlich stärken.

Insgesamt weisen die Ergebnisse deutliche Parallelen zu bisherigen Studien auf, die die strukturelle Abhängigkeit politischer Regulierungsbemühungen von privat kontrollierten Infrastrukturen herausarbeiten. Allerdings lässt sich die Analyse in dieser Arbeit in zweierlei Hinsicht erweitern: Erstens, indem sie die foucaultsche Perspektive auf Machtdiskurse mit konkreten legitimierenden Strukturen der Plattformbetreiber verbindet. Zweitens, indem sie den institutionellen Kontext der EU-spezifischen Digitalpolitik mit einbezieht. Damit liefert die Arbeit nicht nur eine Bestätigung bekannter Machtasymmetrien, sondern schärft deren Erklärung durch die integrierte Betrachtung von Dispositiv-, Normalisierungs- und Gouvernementalitätsbegriffen.

6.3 Reproduktion von Geschlechterungleichheit durch algorithmische Systeme

Ein weiteres zentrales Ergebnis der Analyse ist die systematische Entpolitisierung und Unsichtbarmachung geschlechtsspezifischer Ungleichheiten in Metas Diskursen zur algorithmischen KI-Inhaltsmoderation – ein Befund, der erst im Zusammenspiel von

Critical Frame Analyse und struktureller Machtanalyse seine volle Bedeutung entfaltet. Die bei Metas öffentlicher Kommunikation verwendeten Frames, legitimieren KI-basierte Moderationssysteme, indem sie diese als objektiv, alternativlos und zukunftsweisend darstellen. Diskriminierung wird dabei rhetorisch als Einzelfall oder technische Komplexität entpolitisiert und aus dem öffentlichen Diskurs herausgehalten (siehe Kapitel 5; Anhang B: Tabelle 2).

Diese diskursive Entpolitisierung ist eng mit der Plattforminfrastruktur verwoben: Die technische Gestaltung – von Moderationsalgorithmen bis zu Sichtbarkeitsregeln – reproduziert gesellschaftliche Ungleichheiten im digitalen Raum. Die unsichtbare Ordnung, die sich daraus ergibt, normalisiert bestehende Machtverhältnisse, während sie gleichzeitig eine vermeintliche Wertfreiheit suggeriert (siehe Kapitel 4, 5).

Betroffen sind insbesondere Frauen* und LGBTIQ-Personen, deren Inhalte algorithmisch häufiger marginalisiert, als unangemessen klassifiziert oder blockiert werden. Binäre oder stereotype Trainingsdaten führen dazu, dass geschlechtliche und sexuelle Vielfalt im System als Abweichung erscheint – mit demokratischen Folgen: Die algorithmische Praxis beschränkt digitale Sichtbarkeit und Teilhabe und wirkt somit direkt auf Repräsentationsrechte (siehe Kapitel 2.2).

Die als „neutral“ und „effizient“ gerahmten Technologien verschleiern somit, dass algorithmische Diskriminierung kein technischer Fehler, sondern Ausdruck bestehender Machtverhältnisse ist. Proprietäre Systeme, Black-Box-Algorithmen und intransparente Datensätze verfestigen normierende Effekte, ohne dass Betroffene oder Regulierungseinrichtungen eingreifen können (vgl. Pasquale, 2015; siehe Kapitel 4.2 - 4.5). Wie bereits bei Noble (2018) oder Buolamwini & Gebru (2018) dargelegt, ist der Ausschluss marginalisierter Perspektiven nicht zufällig, sondern strukturell angelegt.

Ein zentraler Analysebegriff ist hierbei hegemoniale Männlichkeit (vgl. Connell 2015: 130, siehe Kapitel 1.3.2). Die Frame-Analyse zeigt, dass Werte wie Rationalität, Kontrolle und technischer Fortschritt – klassisch mit Männlichkeit assoziiert (Wajcman 1991; Harding 1986) – sowohl die Diskurse als auch die technische Architektur prägen. Diese männlich konnotierten Technikkulturen bestimmen, wie Systeme gebaut, bewertet und reguliert werden. Die Abwertung relationaler, emotionaler oder fürsorgender Praktiken (Care) verstärkt nicht nur symbolische, sondern auch infrastrukturelle Ausschlüsse (siehe Kapitel 4.2–4.4). Der Fokus auf Fehlerquoten, Skalierbarkeit und Effizienz verdrängt Gerechtigkeitsfragen, die als irrational oder irrelevant gelten (siehe Kapitel 5).

Das Responsibility Shifting fügt sich ebenfalls in diese Machtarchitektur ein: Indem Meta Verantwortung auf Technik, Nutzer:innen oder externe Faktenprüfer:innen überträgt, wird sowohl die materielle Wirkung der Infrastruktur als auch die politische Verhandbarkeit struktureller Diskriminierung verschleiert (siehe Kapitel 5.4).

Zusammenfassend tragen die verschiedenen Frames, durch Mechanismen wie Verschleierung, Overblocking, symbolische Vereinnahmung und Lastenverschiebung, auf unterschiedliche Weise zur Stabilisierung dieser Ungleichheiten bei, stützen sich jedoch gegenseitig und entfalten dadurch kumulative Wirkung.

Das Konzept von Connell bietet einen Rahmen, um zu verstehen, warum feministische und intersektionale Perspektiven im digitalen Machtfeld oft keinen Anschluss finden: Technologische Männlichkeitsnormen prägen institutionelle Logiken innerhalb Metas und lassen alternative Ordnungsvorstellungen kaum zu. Hegemoniale Männlichkeit stabilisiert sich somit nicht nur im Diskurs, sondern als dispositives Prinzip, das technologische, normative und organisatorische Praktiken miteinander verknüpft (siehe Kapitel 4.4).

Daraus folgt: Die Reproduktion von Geschlechterungleichheit durch algorithmische Moderation ist weder rein technisch noch rein diskursiv erklärbar – sie ist eine machtvolle Verschränkung von Deutung, Struktur und Normativität.

Im Ergebnis zeigt sich Connells Konzept der Hegemonialen Männlichkeit nicht nur als soziale Ordnung, sondern als prägendes Machtgefüge im digitalen Raum: Meta ist eines von wenigen globalen Konzernen, welche digitale Infrastrukturen gestalten und kontrollieren können. Dabei gelten hegemonial männliche Ordnungsvorstellungen als selbstverständlich (siehe Kapitel 4, 5). Das entscheidet darüber, ob digitale Souveränität und Geschlechtergerechtigkeit überhaupt verhandelbar sind.

Plattformen wie Meta entziehen durch ihre strukturelle Macht sowohl demokratischen Institutionen als auch zivilgesellschaftlichen Bewegungen die Möglichkeit, gerechte Bedingungen für den digitalen Raum einzufordern. Die diskursive wie infrastrukturelle Normsetzung marginalisiert geschlechtergerechte Anliegen systematisch – und koppelt so digitale Souveränität von Kritik an Geschlechterungleichheiten ab.

Die kritische Diskussion der Ergebnisse zeigt, dass nur eine intersektional und machtanalytisch informierte Regulierung algorithmischer Systeme sowohl das technokratische Legitimationsnarrativ als auch die patriarchal strukturierten Machtverhältnisse effektiv herausfordern kann. Ein bloßes „Bias Monitoring“ oder formale

Antidiskriminierungsrichtlinien greifen zu kurz, solange die grundlegende diskursive und infrastrukturelle Ordnung hegemonialer Männlichkeit unangetastet bleibt.

Im Gegensatz zu Noble (2018), die den Fokus primär auf kulturelle und mediale Repräsentationen legt, zeigt die vorliegende Arbeit, dass nicht nur die algorithmische/technische Architektur, sondern auch die Unternehmensarchitektur und Kommunikationsstrategien wesentliche Träger patriarchaler Strukturen sind, die gesellschaftliche Ungleichheiten reproduzieren. Anschließend an die Erkenntnisse von Buolamwini & Gebru (2018), die algorithmischen Verzerrungen im KI-Kontext analysierten, konnte außerdem hervorgehoben werden, wie diese Verzerrungen durch diskursive Machtpraktiken bei Meta verschleiert, zementiert und als technologisch neutral dargestellt werden.

Diese Arbeit zeigt somit, dass politische und regulatorische Strategien multidimensional ansetzen müssen – sowohl auf der Ebene der technischen Gestaltung als auch im Bereich der gesellschaftlichen Bedeutungsproduktion –, um wirkliche Geschlechtergerechtigkeit im digitalen Raum zu erreichen.

Die Analyse verdeutlicht also, dass digitale Souveränität und Geschlechtergerechtigkeit nicht als getrennte Politikfelder behandelt werden können. Die in Kapitel 4 beschriebenen infrastrukturellen Kontrollmechanismen und die in Kapitel 5 aufgezeigten Legitimationsframes wirken im Genderkontext doppelt einschränkend: Einerseits reduzieren intransparente Moderationsalgorithmen die politische Durchsetzbarkeit europäischer Regulierungsansprüche (vgl. Hulkó, Kálmán & Lapsánszky, 2025), andererseits marginalisieren sie systematisch feministische und intersektionale Positionen, indem über die Reichweite von entsprechenden Inhalten von voreingenommenen KI-Moderationssystemen entschieden wird (vgl. Noble, 2018; Buolamwini & Gebru, 2018). Damit wird deutlich, dass jede Strategie zur Stärkung digitaler Souveränität auch geschlechterpolitische Perspektiven integrieren muss, um nicht zu einer formalrechtlichen, aber inhaltlich exklusiven Souveränität zu werden. Eine Machtanalyse ohne Genderbezug würde zentrale Ausschlussmechanismen im digitalen Raum übersehen und bestehende Ungleichheiten in die Zukunft fortschreiben – genau jene Gefahr, die Connells Konzept hegemonialer Männlichkeit auf den digitalen Raum angewendet sichtbar macht.

6.4 Grenzen und Implikationen der Analyse

Die in der vorliegenden Masterarbeit verwendete qualitative Methodik, basierte wie zuvor in Kapitel 2 ausführlich erläutert, auf einer kritischen Frame-Analyse nach Verloo (2005) und einer an Foucault (1976) orientierten strukturellen Machtanalyse. Ziel war es,

diskursive Legitimationen algorithmischer Kontrolle bei Meta sichtbar zu machen und deren Auswirkungen auf digitale Souveränität sowie Geschlechterverhältnisse in der EU zu untersuchen. Interne Dokumente oder technische Audits standen nur begrenzt zur Verfügung, sodass insbesondere die strukturelle Machtanalyse sich primär auf von Meta selbst veröffentlichten Materialien und nur limitiert auf infrastrukturelle Aspekte stützen konnte.

Die starke Orientierung der strukturellen Machtanalyse an diskurstheoretischen Konzepten eröffnet wertvolle Einsichten in produktive Machtbeziehungen, erschwert jedoch die umfassende Erfassung materieller Machtaspekte technischer Systeme, wie sie sich in Algorithmen, Datenbasis oder Hardware manifestieren. Diese technische Intransparenz stellt eine zentrale Limitation dar, die eine tiefgehende Analyse algorithmischer Funktionsweisen und damit auch der materiellen Dimension struktureller Macht begrenzt (vgl. Kapitel 4).

Neben diesen methodischen Einschränkungen lassen sich auch inhaltliche Begrenzungen feststellen. Die Konzentration auf einen dominanten Plattformakteur ermöglicht zwar eine vertiefte Analyse, blendet jedoch Wechselwirkungen mit anderen Plattformen und Akteuren wie TikTok, X/Twitter oder YouTube weitgehend aus. Darüber hinaus fehlt eine systematische Einbeziehung von Nutzer:innen- oder zivilgesellschaftlichen Perspektiven, sodass nicht erfasst werden kann, wie die untersuchten Frames in unterschiedlichen Öffentlichkeiten tatsächlich wirken. Auch intersektionale Analysen konnten in dieser Arbeit nur eingeschränkt berücksichtigt werden. Der Schwerpunkt liegt entsprechend auf Gender-Aspekten, während andere Dimensionen wie Ethnizität, sozioökonomischer Status oder Behinderung lediglich am Rand in die Analyse eingeflossen sind. Auch der regulatorische Kontext ist hochdynamisch, weshalb sich Befunde zur Wirksamkeit aktueller politischer Maßnahmen – etwa zu DSA oder AI-Act – in kurzer Zeit verändern können. Die aufgeführten Aspekte hätten den Umfang dieser Arbeit überstiegen, eröffnen jedoch Potenzial für weiterführende Forschungsansätze.

Trotz dieser Grenzen bestätigt die Analyse, dass Meta durch gezielte diskursive Frames algorithmische Kontrolle als alternativlos und unpolitisch legitimiert und damit wirksame regulatorische Eingriffe erschwert. Gleichzeitig werden strukturelle Benachteiligungen, wie geschlechtsspezifische Diskriminierungen, die sich in hegemonialen Männlichkeitsmustern manifestieren, verschleiert. Daraus ergeben sich zentrale Implikationen für Regulierung, Forschung und gesellschaftspolitische Strategien. Künftige europäische Politik sollte die technische wie auch die diskursive Macht von Plattformen adressieren, indem

beispielsweise verpflichtende Daten- und Algorithmus Transparenz mit einer kritischen Beobachtung diskursiver Machtpraktiken verbunden wird. Für die Forschung bedeutet dies, dass nur eine Kombination aus technischer Analyse, betroffenenzentrierter Perspektive und plattformübergreifenden Vergleichen ein vollständiges Bild der Machtstrukturen ermöglichen kann. Für Gender- und Demokratiepoltik wiederum ist entscheidend, digitale Souveränität so zu gestalten, dass geschlechtergerechte und inklusive Perspektiven integraler Bestandteil regulatorischer Strategien werden (vgl. Hulkó, Kálmán & Lapsánszky, 2025; Noble, 2018).

Insgesamt macht die Arbeit deutlich, dass digitale Souveränität und Geschlechtergerechtigkeit eng miteinander verknüpft sind. Die infrastrukturelle Vormacht von Plattformen wie Meta beschränkt demokratische Steuerungs- und Gestaltungsfähigkeiten und wirkt zugleich als Verstärker bestehender gesellschaftlicher Hierarchien. Eine wirksame Regulierung muss daher immer sowohl die strukturellen als auch die diskursiven Machtaspekte berücksichtigen, intersektional ausgerichtet sein und die technische Realität proprietärer Systeme anerkennen.

7 Beantwortung der Forschungsfrage

Auf Grundlage, der in Kapitel 6 dargestellten Befunde wird im Folgenden die zentrale Forschungsfrage abschließend beantwortet und die Bedeutung der Ergebnisse im Hinblick auf digitale Souveränität und Geschlechtergerechtigkeit eingeordnet. Es wurde, wie in Kapitel 2 dargelegt, die Forschungsfrage untersucht, inwiefern Metas KI-gestützte Moderationssysteme die digitale Souveränität der EU sowie die Reproduktion von Geschlechterungleichheiten beeinflussen. Des Weiteren wurden die Unterfragen behandelt, inwieweit algorithmische Bias in Metas Inhaltsmoderation Geschlechterungleichheiten verstärken, wie Metas Kontrolle über algorithmische Entscheidungen die digitale Souveränität Europas untergräbt und welche regulatorischen Maßnahmen der EU diese Herausforderungen adressieren sowie wie wirksam diese sind. Die Analyse verdeutlicht, dass Metas Unternehmensstruktur ein intransparentes und asymmetrisches Machtgefüge hervorbringt, das nicht nur technologische, sondern auch politische Handlungsspielräume europäischer Regulierung massiv einschränkt. Kapitel 4 hat gezeigt, dass sich diese Macht auf mehreren Ebenen entfaltet: durch die infrastrukturelle Organisation digitaler Öffentlichkeiten, durch Prozesse algorithmischer Normalisierung gesellschaftlicher Wertungen und durch gouvernementale Steuerungslogiken, die Nutzer:innenverhalten subtil formen. Meta agiert damit de facto als globale Plattforminfrastruktur, deren

algorithmische Systeme primär unternehmensinternen Kriterien folgen – ohne effektive demokratische Kontrolle oder institutionelle Regulierungseinbindung (Klonick, 2018: 1607–1610; siehe Kapitel 2.1). Die bestehenden regulatorischen Maßnahmen der EU, wie der Digital Services Act, Digital Markets Act, die DSGVO und der AI-Act, bleiben in ihrer Wirkung begrenzt, da ihre Durchsetzung stark von der Kooperationsbereitschaft der Plattformen abhängt und zugleich durch die komplexe politische Struktur der EU erschwert wird (vgl. Hulkó et al., 2025: 8; siehe Kapitel 5, 6.2). Zugleich entzieht sich Meta normativer Kritik durch eine Kombination aus rhetorischer Verschleierung, Verzögerungsstrategien und juristischer Komplexität (siehe Kapitel 2.3, 5).

Die diskursive Konstruktion algorithmischer Moderation als „neutral“, „objektiv“, „fortschritt“ und „effizient“ hat sich als zentrales Legitimationsinstrument erwiesen (siehe Kapitel 5). Diese Frames delegitimieren regulatorische Eingriffe und stärken die narrative Position der unternehmerischen Autonomie. Dadurch wird der politische und gesellschaftliche Charakter von Technologie ausgeblendet und die digitale Souveränität der EU systematisch unterminiert.

Bezüglich der Unterfrage, ob algorithmische Bias in Metas Inhaltsmoderation zur Verstärkung von Geschlechterungleichheiten beitragen, hat die Analyse gezeigt: Metas Moderationssysteme zementieren und verstärken bestehende gesellschaftliche Hierarchien. Die hochgradig zentralisierte Unternehmensstruktur ermöglicht es, technologische Designentscheidungen und normative Standards zentral festzulegen und bestehende Biases in Trainingsdaten sowie geringe Diversität in Entwickler:innen-Teams führen zu einer systematischen Benachteiligung von Frauen*, People of Color und queeren Communities (siehe Kapitel 4). Proprietäre Community Standards institutionalisieren diskriminierende Strukturen und marginalisieren kritische Perspektiven.

Diskursiv wird diese Problematik durch Strategien wie die Betonung von „technischer Komplexität“ und „Fehlerquoten“ von KI-Moderationssystemen entpolitisiert und die Verantwortung Metas abgeschwächt, wodurch Diskriminierung als technisches, kaum vermeidbares Phänomen dargestellt wird (siehe Kapitel 5). Diese Rhetorik schützt hegemoniale, für Geschlechterungleichheiten blinde Annahmen der Tech-Industrie vor Kritik und sichert so die strukturelle Macht von Meta zusätzlich ab.

Die Verknüpfung dieser Befunde mit Connells Konzept der hegemonialen Männlichkeit offenbart, dass technologische Rationalität und Effizienzheterik Teil eines männlich kodierten Herrschaftssystems sind. Meta tritt demnach nicht nur als wirtschaftlicher Gigant,

sondern auch als kultureller Verstärker patriarchaler Ideale auf, die intersektionale Perspektiven systematisch delegitimieren.

In der Summe lassen sich zwei zentrale Mechanismen festhalten, mit denen Meta die digitale Souveränität und Geschlechtergerechtigkeit unterläuft: Einerseits reproduziert die Unternehmensarchitektur durch technische Abschottung und unternehmenszentrierte Entscheidungsprozesse die asymmetrischen Machtverhältnisse und diskriminierenden Strukturen. Andererseits stabilisiert die strategische Kommunikation diese Macht durch Entpolitisierung algorithmischer Systeme, Externalisierung von Verantwortung und die Diskreditierung regulatorischer Eingriffe.

Obwohl EU-Regulierungen erste Schritte auf dem Weg zu einer demokratisch kontrollierten Plattformordnung darstellen, bleibt deren Durchsetzung begrenzt wirksam. Solange Meta und vergleichbare Unternehmen normative Kritik durch wirksame diskursive Frames abwehren, fehlt es an wirksamen Mechanismen zur Durchsetzung geschlechtergerechter und intersektionaler Fairness in der Technologieentwicklung.

Die Antwort auf die Forschungsfrage lautet daher: Metas KI-gestützte Moderationssysteme gefährden sowohl die digitale Souveränität der EU als auch die Geschlechtergerechtigkeit, weil sie strukturelle und diskursive Machtverhältnisse verstärken und damit politische sowie soziale Transformationen blockieren. Plattformmacht operiert nicht isoliert, sondern im Kontext eines technopolitischen Feldes, das von ökonomischen Interessen, patriarchalen Normen und politischer Rhetorik geprägt ist.

Während gesetzgeberische Maßnahmen wie DSA, DMA, DSGVO und AI-Act wichtige Grundlagen schaffen, ist eine weitergehende politische Debatte über die kulturellen und geschlechtsspezifischen Fundamente algorithmischer Macht unabdingbar. Insbesondere muss die institutionalisierte Einbindung intersektionaler Perspektiven in die Technikentwicklung, Regulierung und gesellschaftliche Kommunikation gewährleistet werden.

Die politische Dimension der digitalen Souveränität ist untrennbar mit der sozialen Frage der Gleichstellung verbunden: Wer digitale Plattformmacht demokratisch begrenzen will, muss auch die kulturellen Normen hinterfragen, die diese Macht legitimieren.

Die dargelegten Befunde beantworten nicht nur die eingangs formulierte Forschungsfrage, sondern verdeutlichen zugleich, dass die Analyse weit über den spezifischen Fall von Meta hinausweist. Sie legt grundlegende Muster offen, in denen technische Infrastrukturen, wirtschaftliche Interessen und kulturelle Normen zusammenwirken, um Machtverhältnisse im digitalen Raum zu stabilisieren. Damit liefert die Arbeit nicht nur eine Fallstudie, sondern auch Ansatzpunkte für eine breitere Diskussion über die

demokratische Gestaltung von Plattformmacht. Auf dieser Grundlage werden im folgenden Fazit die zentralen Schlussfolgerungen gebündelt, kritisch eingeordnet und mit einem Ausblick auf künftige Forschungs- und Handlungsfelder verknüpft.

8 Fazit

Die Analyse zeigt, dass Metas KI-gestützte Moderationssysteme durch technische Intransparenz, zentralisierte Entscheidungsstrukturen und gezielte Diskursstrategien sowohl die digitale Souveränität der EU schwächen als auch bestehende Geschlechterungleichheiten verfestigen. Ausgehend von Foucaults Machtbegriffen und Connells Konzept hegemonialer Männlichkeit wurde ein methodischer Ansatz gewählt, der strukturelle Plattformmacht und ihre diskursive Absicherung gleichermaßen in den Blick nimmt. Die Kombination aus struktureller Machtanalyse und Critical Frame Analyse ermöglichte es, sowohl technische als auch narrative Mechanismen der Plattformmacht zu erfassen (siehe Kapitel 3).

Die Untersuchung macht deutlich, dass digitale Souveränität und Fragen der Geschlechtergerechtigkeit nicht isoliert betrachtet werden können, sondern in einem komplexen Wechselverhältnis stehen (siehe Kapitel 7). Metas Rolle als globaler Infrastrukturakteur zeigt, dass technische Intransparenz, proprietäre Standards und gezielte Diskursstrategien nicht nur regulatorische Eingriffe erschweren, sondern auch geschlechterbezogene Ungleichheiten fortschreiben. Damit wird deutlich, dass rechtliche Regulierung allein nicht ausreicht, solange strukturelle und kulturelle Machtgrundlagen unangetastet bleiben. Wirksame Regulierung setzt daher voraus, dass technische Transparenz und verbindliche Zugriffsmöglichkeiten auf relevante Systeme geschaffen werden, um demokratische Kontrolle überhaupt zu ermöglichen. Ebenso bedarf es verpflichtender Mechanismen, die intersektionale Fairness in der Entwicklung, im Betrieb und in der Aufsicht algorithmischer Systeme sicherstellen. Nicht zuletzt muss die europäische Handlungsfähigkeit gestärkt werden, indem bestehende Rechtsinstrumente koordiniert und konsequent umgesetzt werden.

Künftige Forschung könnte insbesondere unterschiedliche Plattformen vergleichend untersuchen, technische Audits stärker einbeziehen und Nutzer:innen zentrierte Perspektiven ausbauen, um die gesellschaftlichen Wirkungen algorithmischer Moderation in ihrer gesamten Breite zu erfassen.

KI-gestützte Moderationssysteme sind nicht nur Werkzeuge zur Inhaltskontrolle, sondern machtvolle Akteure im digitalen Raum. Ihre Regulierung muss daher technologisch

fundiert, politisch entschlossen und gesellschaftlich inklusiv gestaltet sein, um bestehende Asymmetrien nachhaltig zu verändern.

Auf Basis der in den in dieser Arbeit herausgearbeiteten strukturellen und diskursiven Mechanismen lassen sich folgende Handlungsschwerpunkte ableiten: Bestehende EU-Regelwerke wie DSA, DMA und AI-Act sind ein wichtiger Schritt, greifen jedoch in entscheidenden Punkten zu kurz. Um digitale Souveränität und Geschlechtergerechtigkeit wirksam zu sichern, sollten Transparenzpflichten gezielt um die Offenlegung von Trainingsdaten, algorithmischen Gewichtungen und Moderationslogiken erweitert werden. Die Durchsetzung dieser Normen erfordert deutlich mehr technische, personelle und finanzielle Ressourcen in den Aufsichtsbehörden. Zudem sind verbindliche intersektionale Prüfmechanismen sowie unabhängige Monitoring-Stellen nötig, um algorithmische Verzerrungen zu minimieren und gesellschaftliche Effekte kontinuierlich zu bewerten.

Literaturverzeichnis

- AnitaB.org. (2021). Top companies for women technologists: 2021 results. <https://anitaB.org/research-and-impact/top-companies/2021-results/>
- Art. 83 DSGVO, Allgemeine Bedingungen für die Verhängung von Geldbußen (Absatz 5). (n.d.). In *DSGVO-Gesetz*. Abgerufen am 25. August 2025, von <https://dsgvo-gesetz.de/art-83-dsgvo/>
- Beck, S., Grunwald, A., Jacob, K., & Matzner, T. (2019). *Künstliche Intelligenz und Diskriminierung: Herausforderungen und Lösungsansätze*. Plattform Lernende Systeme. https://elearning.hfm-nuernberg.de/pluginfile.php/2807/mod_forum/attachment/330/KI%20und%20Diskriminierung.pdf
- Beer, D. (2016). The social power of algorithms. *Information, Communication & Society*, 20(1), 1–13. <https://doi.org/10.1080/1369118X.2016.1216147>
- Beer, D. (2019). *The data gaze: Capitalism, power and perception*. Sage.
- Bendiek, A., & Stürzer, I. (2022). *Die digitale Souveränität der EU ist umstritten: Warum die EU dennoch im EU-US Handels- und Technologierat auf den Brüssel-Effekt setzen sollte* (SWP-Aktuell, No. 30/2022). Stiftung Wissenschaft und Politik (SWP). <https://doi.org/10.18449/2022A30>
- Bennett, C. J., & Raab, C. D. (2020). Revisiting the governance of privacy: Contemporary policy instruments in global perspective. *Regulation & Governance*, 14(3), 447–464. <https://doi.org/10.1111/rego.12222>
- Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211017308>
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency (FAccT '18)*, 81, 77–91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Cheney-Lippold, J. (2011). A new algorithmic identity: Soft biopolitics and the modulation of control. *Theory, Culture & Society*, 28(6), 164–181. <https://doi.org/10.1177/02632764114244>

- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10, 16598. <https://doi.org/10.1038/s41598-020-73510-5>
- Connell, R. (2015). *Der gemachte Mann: Konstruktion und Krise von Männlichkeiten* (4. Aufl.). Springer VS.
- Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Criado Perez, C. (2019). *Invisible women: Exposing data bias in a world designed for men*. Chatto & Windus.
- Europäische Kommission. (2021). *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>
- Europäische Kommission. (n.d.-a). *Europa für das digitale Zeitalter*. Europäische Kommission. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age_de
- Europäische Kommission. (n.d.-b). *About the DMA*. Abgerufen am 28. Juni 2025, von https://digital-markets-act.ec.europa.eu/about-dma_en
- Europäische Organisation für Gaia-X. (2023, Dezember 4). *About Gaia-X*. Abgerufen am 30. Juli 2025, von <https://gaia-x.eu/what-is-gaia-x/about-gaia-x/>
- Europäische Union. (2025). *Wettbewerbsregeln in der EU*. Ihr Europa. Abgerufen am 28. Juni 2025, von https://europa.eu/youreurope/business/selling-in-eu/competition-between-businesses/competition-rules-eu/index_de.htm
- Europäischer Datenschutzausschuss. (n.d.). *European Data Protection Board*. https://www.edpb.europa.eu/about-edpb/who-we-are/european-data-protection-board_de
- Europäisches Parlament und Rat der Europäischen Union. (2022b). Verordnung (EU) 2022/1925 des Europäischen Parlaments und des Rates vom 14. September 2022 über die Achtung des Privatlebens und den Schutz personenbezogener Daten in

- der elektronischen Kommunikation und zur Aufhebung der Richtlinie 2002/58/EG (Verordnung über Privatsphäre und elektronische Kommunikation). *Amtsblatt der Europäischen Union*, L 337, 1-102. <https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX:52017PC0010>
- Europäisches Parlament und Rat der Europäischen Union. (2022c). Richtlinie (EU) 2022/2555 des Europäischen Parlaments und des Rates vom 14. Dezember 2022 über Maßnahmen für ein hohes gemeinsames Cybersicherheitsniveau in der Union, zur Änderung der Verordnung (EU) Nr. 910/2014 und der Richtlinie (EU) 2018/1972 sowie zur Aufhebung der Richtlinie (EU) 2016/1148 (NIS-2-Richtlinie). *Amtsblatt der Europäischen Union*, L 33/80. <https://eur-lex.europa.eu/legal-content/DE/ALL/?uri=CELEX%3A32022L2555>.
- Europäisches Parlament. (2023, Juni 1). *KI-Gesetz: Erste Regulierung der künstlichen Intelligenz*. Europäisches Parlament. <https://www.europarl.europa.eu/topics/de/article/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz>
- European Commission. (2018, April 25). Artificial Intelligence for Europe (COM(2018) 237 final). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0237>
- European Institute for Gender Equality. (2022, November 28). Artificial intelligence, platform work and gender equality (Report No. 978-92-9482-939-9). Publications Office of the European Union. <https://doi.org/10.2839/372863>
Abgerufen am 1. Juli 2025, von <https://eige.europa.eu/publications/artificial-intelligence-platform-work-and-gender-equality-report>
- Feast, J. (2019, November 7). 4 ways to address gender bias in AI. *Harvard Business Review*. Abgerufen am 1. Juli 2025, von <https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai>
- Floridi, L. (2020). The Fight for Digital Sovereignty: What It Is, and Why It Matters, Especially for the EU. *Philosophy & Technology*, 33, 369–378. <https://doi.org/10.1007/s13347-020-00423-6>
- Foucault, M. (1976). *Überwachen und Strafen: Die Geburt des Gefängnisses* (W. Seitter, Üb.). Suhrkamp. (Original erschienen 1975)

- Foucault, M. (1978a). *Ein Spiel um die Psychoanalyse. Gespräch mit Angehörigen des Département de Psychanalyse der Universität Paris VIII in Vincennes*. In *Dispositive der Macht. Über Sexualität, Wissen und Wahrheit* (118–176). Merve Verlag.
- Foucault, M. (1978b). *Recht der Souveränität/Mechanismus der Disziplin Vorlesung vom 14. Januar 1976*. In *Dispositive der Macht. Über Sexualität, Wissen und Wahrheit* (75–86). Merve Verlag.
- Foucault, M. (1983). *Der Wille zum Wissen: Sexualität und Wahrheit I* (W. Seitter, Üb.). Suhrkamp. (Original erschienen 1976)
- Foucault, M. (2006). *Geschichte der Gouvernementalität I: Sicherheit, Territorium, Bevölkerung*. Vorlesung am Collège de France 1977–1978 (C. Pialoux & J. Schröder, Hrsg.; J. Schröder, Üb.). Suhrkamp. (Original erschienen 2004)
- Freudenschuss, M. (2014). Digitalisierung: Eine feministische Baustelle. *Femina Politica*, 23(2), 9–21.
- Geschke, D., Klaben, A., Quent, M., & Richter, C. (2019). *#Hass im Netz: Der schleichende Angriff auf unsere Demokratie. Eine bundesweite repräsentative Untersuchung*. Institut für Demokratie und Zivilgesellschaft. https://www.idz-jena.de/fileadmin/user_upload/_Hass_im_Netz_-_Der_schleichende_Angriff.pdf
- Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Glasze, G., Odzuck, E., & Staples, R. (Hrsg.). (2022). *Was heißt digitale Souveränität? Diskurse, Praktiken und Voraussetzungen "individueller" und "staatlicher Souveränität" im digitalen Zeitalter* (Politik in der digitalen Gesellschaft, Bd. 3). transcript Verlag. <https://doi.org/10.14361/9783839458273>
- Goanta, C., Bertaglia, T., & Iamnitchi, A. (2022). The case for a legal compliance API for the enforcement of the EU's Digital Services Act on social media platforms. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22)* (Article 111). ACM. <https://doi.org/10.1145/3531146.3533190>

- Gorwa, R., & Ash, T. G. (2020). Democratic Transparency in the Platform Society. In N. Persily & J. A. Tucker (Hrsg.), *Social Media and Democracy* (286–312). Cambridge University Press. <https://doi.org/10.1017/9781108890960>
- Grieser, C. (2021). *Plattformen als infrastrukturzentrierte Märkte: Eine Integration von Markt- und Infrastrukturperspektiven auf Plattformen*. SSOAR. <https://doi.org/10.21241/ssoar.99662>
- Harding, S. (1986). *The science question in feminism*. Cornell University Press. <https://doi.org/10.1177/027046768600600481>
- Harding, S. (2016). *Whose science? Whose knowledge?: Thinking from women's lives*. Cornell University Press. <https://doi.org/10.7591/9781501712951>
- Helberger, N., Pierson, J., & Poell, T. (2017). Governing online platforms: From contested to cooperative responsibility. *The Information Society*, 34(1), 1–14. <https://doi.org/10.1080/01972243.2017.1391913>
- Historisches Lexikon der Schweiz (HLS). (2014, September 11.). *Technischer Fortschritt*, Abgerufen am 21. Juli 2025, von <https://hls-dhs-dss.ch/de/articles/013843/>
- Hulkó, G., Kálmán, J., & Lapsánszky, A. (2025). The politics of digital sovereignty and the European Union's legislation: Navigating crises. *Frontiers in Political Science*, 7, Article 1548562. <https://doi.org/10.3389/fpos.2025.1548562>
- Jamali, L. (2025, April 10). *Meta whistleblower alleges work with China on censorship*. BBC News. <https://www.bbc.com/news/articles/c4grrwvn1lyo>
- Kaplan, J. (2025, Januar 7). *More speech and fewer mistakes*. <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>
- Katzenbach, C., & Ulbricht, L. (2019). Algorithmic governance. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1424>
- Kaufmann, T. (Hrsg.). (2015). *Geschäftsmodelle in Industrie 4.0 und dem Internet der Dinge: Der Weg vom Anspruch in die Wirklichkeit*. Springer Fachmedien.
- Kayser-Bril, N. (2021, November 16). Automatisierte Diskriminierung: Facebook verwendet grobe Stereotypen, um die Anzeigenschaltung zu optimieren.

AlgorithmWatch. <https://algorithmwatch.org/de/automatisierte-diskriminierung-facebook-verwendet-grobe-stereotypen-um-die-anzeigenschaltung-zu-optimieren/>

Keller, E. F. (1995). *Reflections on gender and science*. Yale University Press.

Kiersch, G. (1977). Souveränität. In W. Woyke (Hrsg.), *Handwörterbuch internationale Politik* (281–284). Leske + Budrich.

Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 131(6), 1598–1670.

Köhler, P., & Schmid, G. (2022). *Der Digital Services Act – ein Überblick*. Taylor Wessing. <https://www.taylorwessing.com/de/insights-and-events/insights/2022/11/digital-services-act-ein-ueberblick>

Lombardo, E., & Meier, P. (2008). Framing gender equality in the European Union political discourse. *Social Politics: International Studies in Gender, State & Society*, 15(1), 101–129. <https://doi.org/10.1093/sp/jxn001>

McKinsey & Company. (2021). *Diversity wins: How inclusion matters*. <https://www.mckinsey.com/business-functions/organization/our-insights/diversity-wins-how-inclusion-matters>

McMahon, L., Kleinman, Z. & Subramanian, C. (2025, Januar 7). Meta to replace 'biased' fact-checkers with moderation by users. *BBC News*. <https://www.bbc.com/news/articles/cly74mpy8klo>

Meta. (2022a). *Annual Diversity Report*. https://about.fb.com/wp-content/uploads/2022/07/Meta_Diversity-Data-Summary-Report_2022.pdf

Moerel, L., & Timmers, P. (2021, Januar). *Reflections on digital sovereignty*. (Research in Focus, 4-32) Tilburg University. Abgerufen am 1. Juli 2025, von <https://eucyberdirect.eu/research/reflections-on-digital-sovereignty>

Muhammad, A., Sapieżyński, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019, November). Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), Article 199. <https://doi.org/10.1145/3359301>

- Mützel, S., Saner, P., & Unternährer, M. (2018). Schöne Daten! Konstruktion und Verarbeitung von digitalen Daten. In D. Houben & B. Prietl (Hrsg.), *Datengesellschaft* (111–132). transcript Verlag. <https://doi.org/10.1515/9783839439579-005>
- Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- Pamuk, H. (2025, August 7). Exclusive: Rubio orders US diplomats to launch lobbying blitz against Europe's tech law. *Reuters*. <https://www.reuters.com/sustainability/society-equity/rubio-orders-us-diplomats-launch-lobbying-blitz-against-europes-tech-law-2025-08-07/>
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Perez, S. (2025, April 9). Meta whistleblower Sarah Wynn-Williams says company targeted ads at teens based on their emotional state. *TechCrunch*. <https://techcrunch.com/2025/04/09/meta-whistleblower-sarah-wynn-williams-says-company-targeted-ads-at-teens-based-on-their-emotional-state/>
- Podszun, R. (2023). From competition law to platform regulation – Regulatory choices for the Digital Markets Act. *Economics. The Open-Access, Open-Assessment e-Journal*, 17(1). <https://doi.org/10.1515/econ-2022-0037>
- Pohle, J., Thüer, L., Dammann, F., & Winkler, J. (2023). Das Subjekt im politischen Diskurs zu „digitaler Souveränität“. In N. Kersting, J. Radtke & S. Baringhorst (Hrsg.), *Handbuch Digitalisierung und politische Beteiligung* (1–23). Springer VS. https://doi.org/10.1007/978-3-658-31480-4_12-1
- Ranking Digital Rights. (2022). *The 2022 Big Tech Scorecard*. <https://rankingdigitalrights.org/bts22/>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big data & society*, 6(1), 1-12. <https://doi.org/10.1177/2053951718820549>
- Santos, B. (2025, Juli 18). In the Trump era, deregulation has become a foreign policy directive. *TechPolicy Press*. <https://www.techpolicy.press/in-the-trump-era-deregulation-has-become-a-foreign-policy-directive/>

- Sherman, N. (2025, Januar 11). Meta and Amazon scale back diversity initiatives. *BBC News*. <https://www.bbc.com/news/articles/cgmy7xpw3pyo>
- Srnicek, N. (2018). *Platform-Kapitalismus* (U. Schäfer, Üb.). Hamburger Edition, HIS.
- Staab, P. (2019). *Digitaler Kapitalismus: Markt und Herrschaft in der Ökonomie der Unknappheit*. Suhrkamp.
- UNESCO. (2022). *The effects of AI on the working lives of women*. UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000380861>
- Van Dijck, J., Poell, T., & de Waal, M. (2018). The platform society: Public values in a connective world (7–30). Oxford University Press.
<https://doi.org/10.1093/oso/9780190889760.003.0002>
- Verloo, M. (2005). Mainstreaming gender equality in Europe: A critical frame analysis approach. *The Greek Review of Social Research*, 117(B), 11-34. <https://dx.doi.org/10.12681/grsr.9555>
- Wagner, B., Husovec, M., Kuklis, L., & Pirkova, E. (2022, Februar 23). The next step towards auditing intermediaries. *Verfassungsblog*.
<https://doi.org/10.17176/20220224-001235-0>
- Wajcman, J. (1991). *Feminism confronts technology*. Polity Press.
- Wiegerling, K., Nerurkar, M., & Wadephul, C. (2018). Ethische und anthropologische Aspekte der Anwendung von Big-Data-Technologien. In B. Kolany-Raiser, R. Heil, C. Orwat, & T. Hoeren (Hrsg.), *Big Data und Gesellschaft. Technikzukünfte, Wissenschaft und Gesellschaft / Futures of Technology, Science and Society*. 1-66. Springer VS. https://doi.org/10.1007/978-3-658-21665-8_1
- Wolford, B. (n.d.). What is GDPR? Abgerufen am 28. Juni 2025, von <https://gdpr.eu/what-is-gdpr/>
- Zuboff, S. (2018). *Das Zeitalter des Überwachungskapitalismus*. Campus.

Transparenzhinweis:

Für die sprachliche Überarbeitung dieser Arbeit wurde das KI-Lektorat- und Korrekturlesetool Scribbr eingesetzt, das mit Hilfe künstlicher Intelligenz das gesamte Word-Dokument gescannt und stilistische sowie orthografische Verbesserungsvorschläge erarbeitet hat. Diese wurden im neuen Dokument über die Word-Nachverfolgung als Kommentare am Seitenrand angezeigt, um Lesbarkeit und Ausdruck präzise zu optimieren. Das Tool wurde am 26. August 2025 genutzt. Die Anwendung umfasste sowohl orthografische Korrekturen als auch stilistische Anpassungen, wobei die Vorschläge von der Verfasserin überprüft und teilweise übernommen wurden.

Zur Nachvollziehbarkeit der eigenständigen Leistung sind exemplarische Korrekturvorschläge dokumentiert und auf Anfrage einsehbar.

Alle inhaltlichen Ausführungen und Bewertungen dieser Arbeit stammen vollständig von der Verfasserin. Das KI-Tool wurde ausschließlich zur sprachlichen Unterstützung und Qualitätssicherung eingesetzt.

Anhang A

Im Anhang A befinden sich die in der strukturellen Machtanalyse in Kapitel 4 analysierten Quellen.

Quellen der strukturellen Machtanalyse (Kapitel 4)

Corporate Europe Observatory. (2023, Februar 23). *The lobbying ghost in the machine: Big Tech's covert defanging of Europe's AI Act*. <https://corporateeurope.org/en/2023/02/lobbying-ghost-machine>

Europäische Kommission. (2020, März 5). *Eine Union der Gleichheit: Strategie für die Gleichstellung der Geschlechter 2020–2025 (COM(2020) 152 final)*. <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=COM%3A2020%3A152%3AFIN>

Europäische Kommission. (2022). *Digitalstrategie der Europäischen Kommission: Digitale Kommission der nächsten Generation*. https://commission.europa.eu/publications/european-commission-digital-strategy_de

Europäische Kommission. (n.d.-c). *Digital Services Act*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act_de

Europäisches Parlament und Rat der Europäischen Union. (2022a). Verordnung (EU) 2022/1925 über bestreitbare und faire Märkte im digitalen Sektor und zur Änderung der Richtlinien (EU) 2019/1937 und (EU) 2020/1828 (Gesetz über digitale Märkte). *Amtsblatt der Europäischen Union*, L262/1 <https://eur-lex.europa.eu/legal-content/DE/TXT/?uri=CELEX:32022R1925>

Meta. (2022b, Januar 19). *Die Menschen hinter den Review-Teams von Meta*. <https://transparency.meta.com/de-de/enforcement/detecting-violations/people-behind-our-review-teams/>

Meta. (2023, Januar 18). *So bringen wir lokalen Kontext in unsere globalen Standards ein*. <https://transparency.meta.com/de-de/policies/improving/bringing-local-context/>

Meta. (2024, November 12). *So priorisiert Meta Inhalte zur Überprüfung*. <https://transparency.meta.com/de-de/policies/improving/prioritizing-content-review/>

Meta. (n.d.). *Transparenzberichte*. Meta. Abgerufen am 3. Juli 2025, von <https://transparency.meta.com/de-de/>

Anhang B

Im Anhang B sind die analysierten Quellen der Critical Frame Analyse (CFA) aufgelistet sowie, die zentralen Arbeitsschritte dokumentiert. Die Tabellen zeigen sowohl die Auswahl der Dokumente (Tabelle 1), die systematische Rekonstruktion der Frame-Komponenten (Tabelle 2) als auch exemplarische Textsegmente und deren erste Codierungen für jeden Frame (Tabellen 3–6). Sie dienen der Transparenz und Nachvollziehbarkeit des methodischen Vorgehens.

Material der CFA (Kapitel 5)

Pesenti, J. (2018, Mai 2). AI at F8 2018: Open frameworks and responsible development. *Meta for developers*. Abgerufen am 1. Juli 2025, von <https://ai.meta.com/blog/ai-at-f8-2018-open-frameworks-and-responsible-development/>

U.S. Senate Judiciary Committee. (2020, November 17). Hearing on Censorship, Suppression, and the 2020 Election. Testimony of Mark Zuckerberg, CEO of Facebook, and Jack Dorsey, CEO of Twitter [Transkript]. *U.S. Government Publishing Office*. Abgerufen am 1. Juli 2025, von <https://www.rev.com/transcripts/mark-zuckerberg-jack-dorsey-testimony-transcript-senate-tech-hearing-november-17>

Zuckerberg, M. [@zuck]. (2025, Januar 7). It's time to get back to our roots around free expression. We're replacing fact checkers with Community Notes, simplifying our policies and focusing on reducing mistakes. Looking forward to this next chapter. [Instagram-Video, Transkript]. *Instagram*. Abgerufen am 1. Juli 2025, von <https://www.americanrhetoric.com/speeches/markzuckerbergnewmetaccontentmoderationpolicies.htm>

Tabelle 1: Analyisierte Quellen der CFA

Quelle	Medium/Typ	Verfasser:in	Veröffentlichungsdatum	Kontext/Kurzbeschreibung	Zugangs-/Fundstelle
Pesenti, 2018	Unternehmensblogpost	Nick C. Pesenti (Meta)	2018	Blogpost zur Einführung algorithmischer Systeme bei Meta; Darstellung technologischer Effizienz & ethischer Ansprüche	Offizieller Blog von Meta (Facebook.ai)
US-Senatsanhörung, 2020	Öffentliche Anhörung	Mark Zuckerberg, U.S. Senate Judiciary Committee	17. November 2020	Aussagen Zuckerbergs zu algorithmischer Moderation und Zusammenarbeit mit Faktenprüfungsorganisationen	Video und Transkript (öffentlicher YouTube-Kanal, Politik-Archiv)
Zuckerberg, 2025 Instagram Statement	Social Media Posting	Mark Zuckerberg	Januar 2025	Statement zu Neuerungen in der Inhaltsmoderation, Einführung von Community-Notes und Moderationsteams-Verlagerung	Instagram-Post auf offiziellen Meta-Kanälen und Transkript

Tabelle 2: Rekonstruktion der Frames nach Verloo (2005) für CFA

Frame	Problemldefinition	Ursachen	Moralische Bewertung	Lösungsansatz	Beispielzitat
Neutralität & Objektivität	Kritik an Algorithmen als unbegründet dargestellt, indem die Systeme als technisch neutral und frei von Vorurteilen dargestellt werden	Technische Systeme aufgrund ihrer Rationalität und Wissenschaftlichkeit angeblich von menschlichen Vorurteilen entkoppelt	Technik erscheint vertrauenswürdiger und moralisch überlegen gegenüber menschlichen Entscheidungen.	Verantwortung an Technik delegieren; Regulierung erscheint unnötig.	„And I think it's worth noting that our people reviewers are not, in all cases, always more accurate than the technical systems.“ (U.S. Senate Judiciary Committee, 2020, 03:05:13)
Effizienz & Sicherheit	Online-Plattformen sind durch Hassrede, Terror-Propaganda oder Desinformation gefährdet.	Menge und Komplexität der Inhalte, die menschlich nicht kontrollierbar sind.	Schutz der Gesellschaft vor Gefahr und Chaos.	KI-gestützte, automatisierte Moderation; Legitimation massiver technischer Eingriffe.	„We built sophisticated systems to protect against election interference [...]. We've taken down more than 100 networks of bad actors [...].“ (U.S. Senate Judiciary Committee, 2020, 25:31)
Fortschritt & Innovation	Regulatorische Maßnahmen als Hemmnisse für Fortschritt. Kritik an algorithmischen Systemen erscheint als innovativ-unsensibel.	Überregulierung durch Politik, die Dynamiken von Technik nicht versteht.	Fortschritt und Innovation sind intrinsisch wertvoll und notwendig für wirtschaftliche Wettbewerbsfähigkeit, Regulierung gefährdet diese Werte.	Forderung nach Deregulierung, Schutz der Innovationsfreiheit, Betonung technoprimistischer Narrative.	„The U.S. has the strongest constitutional protections for free expression in the world. Europe has an ever-increasing number of laws institutionalizing censorship and making it difficult to build anything innovative there.“ (Zuckerberg, 2025, Abs. 13)
Responsibility Shifting	Verantwortung wird diffundiert und ausgelagert, wodurch mangelnde politische und gesellschaftliche Rechenschaft verdeckt wird.	Inhärenten Komplexität KI-basierter Systeme und die Vielzahl beteiligter Akteure.	Verantwortungslosigkeit erscheint als strukturelles Nebenprodukt; Schuld wird diffundiert.	Delegation an Nutzer:innen, Kombination aus Community-Praktiken und Maschinen – Regulierung bleibt diffus.	„We've open-sourced House3D, and will be releasing the data [...] in the near future, to help the entire AI research community to make faster progress toward creating smarter, more genuinely autonomous intelligent assistants.“ (Pesenti, 2028, Abs. 9)

Tabelle 3: Codierung Neutralitäts- und Objektivitätsframe

Quelle	Textausschnitt	Erste Codierung
Meta Blogpost (2018)	„To help accelerate and optimize this process, we are announcing our plan for the next version of our open source AI framework, PyTorch 1.0 [...] With PyTorch 1.0, developers will be able to experiment rapidly in a flexible, immediate execution mode, and then seamlessly transition to a highly optimizable, graph-based mode for deployment.“ (Pesenti, 2018, Abs.3)	„optimize“, „flexible“, „highly optimizable“ → Technik erscheint neutral, wertfrei, objektiv.
Instagram Statement (2025)	„We tried in good faith to address those concerns without becoming the arbiters of truth. But the fact checkers have just been too politically biased and have destroyed more trust than they’ve created, especially in the U.S.“ (Zuckerberg, 2025, Abs. 8)	„not becoming the arbiters of truth“, „politically biased“ → Technik/Plattform stellt sich selbst als neutral, unparteiisch, objektiv dar, im Gegensatz zu den angeblich voreingenommenen Fact Checkern.
Senats- anhö- rung (2020)	„I think it is not the case that we are like a telco and that there are clearly some categories of content, whether it's terrorism or child exploitation, that people expect us to moderate and address, but we're also clearly not like a news publisher in that we don't create the content and we don't choose upfront what we publish, we give people a voice to be able to publish things.“ (U.S. Senate Judiciary Committee, 2020, 02:16:36)	„not like a news publisher“, „we don't create the content“, „we give people a voice“ → Behauptung der Neutralität und Objektivität, Technik/Plattform als wertfreie Vermittlungsinstanz.

Tabelle 4: Codierung Effizienz-und Sicherheitsframe

Quelle	Textausschnitt	Erste Codierung
Meta Blogpost (2018)	„We trained a new image recognition model on an unprecedented 3.5 billion publicly available photos [...] Using a 1-billion-image version of this dataset enabled our model to score the highest mark ever, 85.4 percent accuracy, on the widely used ImageNet image recognition benchmark.“ (Pesenti, 2018, Abs. 7)	„3.5 billion photos“, „85.4 percent accuracy“ → Effizienz durch Skalierung und Geschwindigkeit.
Instagram Statement (2025)	„Even if they accidentally censor just 1% of posts, that’s millions of people. And we’ve reached a point where it’s just too many mistakes and too much censorship.“ (Zuckerberg, 2025, Abs. 5)	„1%“, „millions of people“ → Effizienz- und Skalierungsargument (Zahlen verdeutlichen Reichweite und Fehlerausmaß).
Senatsanhörung (2020)	„So Senator, what we try to do is identify content that violates our policy before anyone in the community has to see it or even report it to us. And for some categories like terrorism, which I've cited before, about 98% or 99% of the content that we take down our AI and human systems find before anyone even has to report it to us. On hate speech, we're up to 94% of the content that we take down our AI systems and content reviewers find it before people have to report it to us. What we try to drive on more effectiveness is basically finding more and more of that harmful content earlier, before it is seen broadly across our system.“ (U.S. Senate Judiciary Committee, 2020, 01:59:18)	„98–99%“, „94%“, „before anyone has to report it“, „find earlier“ → Effizienz durch Technik, Sicherheit für die Community, Skalierbarkeit.

Tabelle 5: Codierung Fortschritts- und Innovationsframe

Quelle	Textausschnitt	Erste Codierung
Meta Blogpost (2018)	„Artificial intelligence has the potential to transform lives, creating new digital experiences while safeguarding our existing online interactions. Balancing rapid and responsible progress requires foundational investment, as well as a strategic vision for how to build and use this evolving technology.“ (Pesenti, 2018, Abs. 1)	„potential“, „transform lives“, „responsible progress“ → Fortschritt als positive Zukunftsaussichten.
Instagram Statement (2025)	„The recent elections also feel like a cultural tipping point towards once again prioritizing speech. So we’re going to get back to our roots and focus on reducing mistakes, simplifying our policies, and restoring free expression on our platforms.“ (Zuckerberg, 2025, Abs. 6)	„cultural tipping point“, „restoring free expression“ → Fortschritt als Neuanfang / Innovation in Governance inszeniert.
Senatsanhörung (2020)	„We also ran an unprecedented civic engagement program to encourage people to take part in our democracy. We ran the largest voter information campaign in history. 140 million people visited our voting information center, including more than 33 million on election day alone. We estimate that we helped more than 4.5 million people register to vote and helped States recruit 100,000 poll workers.“ ((U.S. Senate Judiciary Committee, 2020,27:32)	„unprecedented“, „largest in history“, „helped millions register to vote“ → Innovation & Fortschritt als gesellschaftlicher Mehrwert, Meta als Wegbereiter demokratischer Zukunft.

Tabelle 6: Codierung Responsibility Shifting

Quelle	Textausschnitt	Erste Codierung
Meta Blogpost (2018)	„All of these open-source frameworks and models will be available on Facebook.ai. This new site ... is intended to help researchers and engineers outside of Facebook take their AI work from research to production more quickly, employing the exact same tools that we use to serve 2 billion people.“ (Pesenti, 2018, Abs. 6)	„researchers and engineers outside of Facebook“ → Verantwortung an externe Community verlagert.
Instagram Statement (2025)	„First, we’re going to get rid of fact checkers and replace them with community notes, similar to X, starting in the U.S. [...] So, over the next couple of months, we’re going to phase in a more comprehensive community note system.“ (Zuckerberg, 2025, Abs. 8)	„community notes“, „replace fact checkers“ → Verantwortung wird auf die Nutzer-Community verschoben.
Senatsanhörung (2020)	„[...] people do not want Facebook to be the arbiter of truth and deciding everything that is true and false, and for what it’s worth, I strongly agree with that and I do not think it is the right thing for us to assume that role.“ (U.S. Senate Judiciary Committee, 2020, 01:21:09)	„people do not want Facebook to be the arbiter of truth“, „community has said...“, „I do not think it is the right thing for us“ → Verantwortung wird auf die Gesellschaft/Community verschoben, nicht bei Meta.

Anhang C

Anhang C enthält die Transkripte des Blogposts von Pesenti (2018) sowie des Instagram-Statements von Zuckerberg (2025), die zur besseren Referenzierbarkeit mit nummerierten Absätzen versehen wurden. Die Senatsanhörung von 2020 ist nicht im Anhang enthalten, da das offizielle Transkript öffentlich online mit Zeitstempeln zugänglich ist. Alle drei Dokumente bilden jedoch die Materialgrundlage für die in Kapitel 5 durchgeführte Frame-Analyse.

Transkripte der analysierten Quellen mit Absatznummerierung

Dokument 1: Blogpost Pesenti 2018 – *AI at F8 2018: Open frameworks and responsible development*

(1) Artificial intelligence has the potential to transform lives, creating new digital experiences while safeguarding our existing online interactions. Balancing rapid and responsible progress requires foundational investment, as well as a strategic vision for how to build and use this evolving technology.

(2) On Day 2 of F8 — Facebook’s annual developer’s conference — we detailed the company’s approach to advancing AI through our open development frameworks and collaborative relationship with the wider AI community. Our keynote speakers, including Facebook’s Chief Technology Officer Mike Schroepfer (pictured above), also unveiled cutting-edge research in computer vision, natural language understanding, and reinforcement learning, and shared our commitment to the ethical development and deployment of AI.

Advancing AI through open development

(3) The path for taking AI from research to production has historically involved paid licenses, multiple development tools, and other hurdles. Closed and scattered development frameworks can make it complicated and time-intensive to test new approaches, deploy them, and iterate to improve accuracy and performance. To help accelerate and optimize this process, we are announcing our plan for the next version of our open source AI framework, *PyTorch 1.0*, which will be available in beta in the coming months. With PyTorch 1.0, developers will be able to experiment rapidly in a flexible, immediate execution mode, and then seamlessly transition to a highly optimizable, graph-based mode for deployment. The PyTorch framework has quickly become one of the most popular

frameworks for AI researchers. Since we released the original version (0.1.6) just over a year ago, it has taken off on GitHub, and it was the second-most cited framework in papers at ICLR.

(4) We also announced the expansion of *ONNX* (Open Neural Network Exchange), an open format for representing deep learning models, to allow AI engineers to more easily move models between frameworks without having to do resource-intensive custom engineering. ONNX has been open source since it was released in 2017, and it now supports additional tools, including a production-ready Core ML converter, Baidu’s PaddlePaddle platform, and Qualcomm SNPE in addition to Amazon Web Services’ Apache MXNet; Facebook’s *Caffe2*, *PyTorch*, and PyTorch 1.0; Google’s TensorFlow, and Microsoft’s Cognitive Toolkit. ONNX is tightly woven into PyTorch 1.0.

(5) In addition to these broadly applicable development frameworks, Facebook is releasing resources for more specific uses of AI. Our ResNext3D model, with state-of-the-art accuracy and efficiency for video understanding, will be available in June (a collection of other video models, including *Res 2+1 D*, is being released today). We’re also open-sourcing *Translate* — a PyTorch language library — for fast, flexible machine translations that scale, and sharing our early work on a project called *Multilingual Unsupervised and Supervised Embeddings* (MUSE), to help expand the number of languages available for translation.

(6) All of these open-source frameworks and models will be available on *Facebook.ai*. This new site — launched today — is intended to help researchers and engineers outside of Facebook take their AI work from research to production more quickly, employing the exact same tools that we use to serve 2 billion people. By sharing this growing suite of open source tools, we’re underscoring our commitment to work collaboratively with the developer community to build meaningful new products and services.

Research in vision, language, and learning

(7) Facebook’s AI researchers pursue a wide range of topics and questions, but at F8 we showcased recent work in computer vision, natural language understanding, and reinforcement learning. In a collaborative effort by the Applied Machine Learning (AML) and Facebook AI Research (FAIR) teams, we trained a new image recognition model on an unprecedented 3.5 billion publicly available photos, using hashtags to classify the images. The researchers had to create a new technique that makes hashtags — which are

noisy and imprecise labels — useful for AI training. Using a 1-billion-image version of this dataset enabled our model to score the highest mark ever, 85.4 percent accuracy, on the widely used ImageNet image recognition benchmark. This performance reveals the long-term potential of not only improving image recognition by training on larger datasets, but also using existing labels (rather than annotations applied specifically for the purposes of AI training). We plan to open-source the embeddings of these models in the near future.

(8) In the field of 3D image mapping, Facebook researchers used the PyTorch toolkit to generate full 3D surfaces that can be applied, in real time, to footage of human bodies in motion. The new tool, which we call *DensePose*, can enable compelling new augmented reality effects, such as transferring a texture onto a moving body. More important, this foundational research points to greater understanding of scenes for computer vision systems. We hope to publish this work as an open source library in the coming weeks.

(9) To promote the development of more useful autonomous agents — whether that means virtual assistants or robotic systems — FAIR has collaborated with researchers at Georgia Tech to develop a new, multistage AI task, called *EmbodiedQA*, that pushes the limits of reinforcement learning and natural language understanding. The team created virtual agents that must learn how to navigate computer-generated indoor spaces and how to understand and use natural language in order to answer questions about their environment. FAIR also built the collection of virtual environments — called *House3D* — used in this research, allowing agents to train dramatically faster than a physical robot would in physical spaces. We’ve open-sourced *House3D*, and will be releasing the data and model related to *EmbodiedQA* in the near future, to help the entire AI research community to make faster progress toward creating smarter, more genuinely autonomous intelligent assistants.

(10) FAIR is exploring other learning environments for agents, with two projects that focus on systems that play games. Our *StarCraft* bot uses reinforcement learning to operate in the cluttered, complex, and partially observable environments of the popular real-time strategy video game *StarCraft*. We’ve developed a state-of-the-art rule set as a baseline for the bot, which we plan to open-source for use in other *StarCraft*-related AI research. And while DeepMind’s *AlphaGo* system has shown impressive performance in beating humans at the game Go and teaching itself without human supervision, the model

itself remains a closely held secret. To make this work both reproducible and available to AI researchers around the world, we created an open source Go bot, called *ELF OpenGo*, that performs well enough to answer some of the key questions unanswered by AlphaGo. Despite a relatively limited amount of training and no supervision, ELF OpenGo has achieved professional status and a 14-0 win record against four different top 30 human Go players.

(11) Our primary goal with these bots, and with most of our research, is to create a foundation for future work by the larger AI community. That's why we open-source the datasets and models for as much of our research as possible, including the tools and datasets used to create our StarCraft bot (which will be available soon) and both the code and a pretrained model for ELF OpenGo (available now).

Using AI responsibly

(12) As we work to shape the future of AI with our research and our open development tools, it's important to apply questions of ethics, fairness, and trust to every effort in this field. That includes using existing AI techniques in a responsible way, such as in the blood donations feature that we released last year. More than 8 million people in India, Bangladesh, and Pakistan have chosen to use the feature to identify themselves as blood donors, in order to better connect with people and organizations requesting blood.

(13) But this issue is bigger than a single project, so we shared Facebook's perspective on the ethical use of AI. Our efforts to eliminate algorithmic bias and irresponsible AI deployment include careful consideration of the code we generate as well as the people we hire to write, manage, and approve those AI systems. Promoting a diverse workforce might prevent some problems in the planning stages of a project by bringing more points of view to the table. Other problems could be caught during development, by our research and privacy reviews. And even after production, we can search for bias with internal tools, such as Fairness Flow, which measures how an algorithm interacts with specific groups of people.

(14) As AI evolves, so will the ethical concerns it raises. So while Facebook continues to push for the open development and research of this technology, we are also investing in people, processes, and tools to help ensure that we are maximizing the positive impacts of these AI systems and minimizing the negative ones.

Dokument 2: Transkript Instagram-Statement Zuckerberg (2025)

(1) Hey, everyone.

(2) I want to talk about something important today, because it's time to get back to our roots around free expression on Facebook and Instagram.

(3) I started building social media to give people a voice. I gave a speech at Georgetown five years ago about the importance of protecting free expression, and I still believe this today. But a lot has happened over the last several years. There's been widespread debate about potential harms from online content. Governments and legacy media have pushed to censor more and more.

(4) A lot of this is clearly political, but there's also a lot of legitimately bad stuff out there: drugs, terrorism, child exploitation. These are things that we take very seriously, and I want to make sure that we handle responsibly.

(5) So we built a lot of complex systems to moderate content. But the problem with complex systems is they make mistakes. Even if they accidentally censor just 1% of posts, that's millions of people. And we've reached a point where it's just too many mistakes and too much censorship.

(6) The recent elections also feel like a cultural tipping point towards once again prioritizing speech. So we're going to get back to our roots and focus on reducing mistakes, simplifying our policies, and restoring free expression on our platforms.

(7) More specifically, here's what we're going to do:

(8) First, we're going to get rid of fact checkers and replace them with community notes, similar to X, starting in the U.S. After Trump first got elected in 2016, the legacy media wrote nonstop about how misinformation was a threat to democracy. We tried in good faith to address those concerns without becoming the arbiters of truth. But the fact checkers have just been too politically biased and have destroyed more trust than they've created, especially in the U.S. So, over the next couple of months, we're going to phase in a more comprehensive community note system.

(9) Second, we're going to simplify our content policies and get rid of a bunch of restrictions on topics like immigration and gender that are just out of touch with mainstream discourse. What started as a movement to be more inclusive has increasingly been used

to shut down opinions and shut out people with different ideas — and it's gone too far. So I want to make sure that people can share their beliefs and experiences on our platforms.

(10) Third, we're changing how we enforce our policies to reduce the mistakes that account for the vast majority of censorship on our platforms. We used to have filters that scanned for any policy violation. Now, we're going to focus those filters on tackling illegal and high severity violations; and for lower severity violations, we're going to rely on someone reporting an issue before we take action. The problem is that the filters make mistakes, and they take down a lot of content that they shouldn't. So, by dialing them back, we're going to dramatically reduce the amount of censorship on our platforms. We're also going to tune our content filters to require much higher confidence before taking down content. The reality is that this is a tradeoff. It means we're going to catch less bad stuff, but it will also reduce the number of innocent people's posts and accounts that we accidentally take down.

(11) Fourth, we're bringing back civic content. For awhile the community asked to see less politics because it was making people stressed. So, we stopped recommending these posts. But it feels like we're in a new era now, and we're starting to get feedback that people want to see this content again. So we're going to start phasing this back into Facebook, Instagram, and Threads, while working to keep the communities friendly and positive.

(12) Fifth, we're going to move our trust and safety and content moderation teams out of California, and our U.S.-based content review is going to be based in Texas. As we work to promote free expression, I think that it will help us build trust to do this work in places where there is less concern about the bias of our teams.

(13) Finally, we're going to work with President Trump to push back on governments around the world that are going after American companies and pushing to censor more. The U.S. has the strongest constitutional protections for free expression in the world. Europe has an ever-increasing number of laws institutionalizing censorship and making it difficult to build anything innovative there. Latin American countries have secret courts that can order companies to quietly take things down. China has censored our apps from even working in the country. The only way that we can push back on this global trend is with the support of the U.S. government.

(14) And that's why it's been so difficult over the past four years — when even the U.S. government has pushed for censorship by going after us and other American companies. It has emboldened other governments to go even further. But now we have the opportunity to restore free expression and I am excited to take it. It'll take time to get this right. And these are complex systems. They're never going to be perfect. There's also a lot of illegal stuff that we still need to work very hard to remove.

(15) But the bottom line is that after years of having our content moderation work focused primarily on removing content, it is time to focus on reducing mistakes, simplifying our systems, and getting back to our roots about giving people a voice.

(16) I'm looking forward to this next chapter.

(17) Stay good out there.

(18) And more to come soon.