

Précis [pʁe 'si:]

// Mai 2026

KI und Deliberation: Normative Ideale im Lichte aktueller KI-Forschung – Eine Literaturübersicht

Dennis Friess, Carina Weinmann und Mira Warné
Kontakt: dennis.friess@hhu.de

ABSTRACT

Dieser Beitrag bietet einen systematischen Überblick über die aktuelle Forschung zum Einsatz künstlicher Intelligenz (KI) in Online-Diskussionen und ordnet diese mithilfe deliberativer Normen ein. Die Prinzipien Egalität, Rationalität, Reziprozität und Zivilität dienen dabei als theoretische Heuristik zur Analyse von Studien, die zwischen 2015 und 2022 erschienen sind.

Auf Grundlage einer Auswertung von insgesamt 171 wissenschaftlichen Artikeln zeigt die Studie, dass sich die bisherige KI-Forschung im Kontext von Online-Diskussionen vornehmlich auf die Förderung von Rationalität und Zivilität konzentriert hat. Insbesondere Techniken wie Argument Mining oder die Erkennung von Hassrede werden eingesetzt, um Diskussionen strukturierter, sachlicher und respektvoller zu gestalten. Ansätze, die mittels KI Egalität und Reziprozität fördern wollen und damit zentrale Voraussetzungen demokratischer Öffentlichkeit adressieren, werden bislang deutlich weniger beachtet.

Die Befunde schärfen den Blick für bestehende Forschungslücken und verdeutlichen den Bedarf an stärker normativ fundierten KI-Ansätzen, die auf eine inklusivere und dialogorientierte Gestaltung digitaler Öffentlichkeiten abzielen.



Dr. Dennis Friess ist wissenschaftlicher Koordinator des DIID.
Dr. Carina Weinmann ist Kommunikationswissenschaftlerin am Institut für Sozialwissenschaften der Heinrich-Heine-Universität Düsseldorf.
Mira Warné M.A. war bis 2025 wissenschaftliche Mitarbeiterin am Institut für Sozialwissenschaften der Heinrich-Heine-Universität Düsseldorf.

Referenzen zum Thema:

Friess, D., Weinmann, C. & Warné, M., (2025) "AI and Deliberation: Normative Ideals in the Light of Current AI Research - A Review", *Journal of Deliberative Democracy*, 21(1), 1-11, <https://doi.org/10.16997/jdd.1805>

Friess, D., & Weinmann, C. (2025). Making Online Participation More Intelligent: Considerations on How Artificial Intelligence Can Support Public Online Participatory Processes. In E. Paulis, R. Kies & A. Östling (Hrsg.). *Public Deliberation in the Digital Age. Platforms, Participation, and Legitimacy* (S. 143-157). Routledge.

Carstens, J. A., & Friess, D. (2024). AI Within Online Discussions: Rational, Civil, Privileged?. *Minds & Machines*, 34(10), 1-25. <https://doi.org/10.1007/s11023-024-09658-0>.

Friess, D., & Eilders, C. (2015). A Systematic Review of Online Deliberation Research. *Policy & Internet*, 7(3), 319-339. <https://doi.org/10.1002/poi3.95>

Einleitung

Das Forschungsfeld der künstlichen Intelligenz (KI) hat in den letzten Jahren ein bemerkenswertes Wachstum erlebt und eine rasch zunehmende Zahl von Forschungsarbeiten zu KI-Anwendungen in Online-Diskussionen hervorgebracht. Im weitesten Sinne umfasst KI „Computersysteme, die Algorithmen, Methoden des maschinellen Lernens, natürliche Sprachverarbeitung und andere Techniken umfassen, die im Auftrag einer Person eingesetzt werden, um das Kommunikationsergebnis zu verbessern“ (Hancock et al., 2020, S. 90; eigene Übersetzung).

In digitalen Diskurskontexten bietet KI leistungsfähige Werkzeuge zur Analyse und Gestaltung von Online-Interaktionen. Dies ist besonders relevant hinsichtlich der gut dokumentierten Probleme im Zusammenhang mit der geringen Qualität und manchmal toxischen Natur von Online-Diskussionen (Coe et al., 2014). Während KI-gestützte Algorithmen häufig dafür kritisiert werden, Echokammern und Polarisierung zu verstärken (Coleman, 2018), wird KI zunehmend auch als Instrument betrachtet, das Online-Diskussionen verbessern und eine demokratische Debatte fördern könnte (Stoll et al., 2020). Trotz des wachsenden Interesses bleibt ein Großteil der bestehenden Forschung deskriptiv oder technisch ausgerichtet und konzentriert sich darauf, wie KI in solchen Umgebungen funktioniert oder wie sie das Nutzer:innenverhalten beeinflusst. Häufig fehlt eine systematische Bewertung von KI-Anwendungen anhand demokratischer Normen.

Die deliberative Demokratietheorie, eine der einflussreichsten Perspektiven in der (digitalen) Demokratieforschung, geht davon aus, dass der öffentliche Diskurs bestimmte normative Standards, wie Zivilität, Egalität, Rationalität und Reziprozität, erfüllen sollte, um demokratische Ideale zu unterstützen (Friess & Eilders, 2015). Dennoch bewerten nur wenige Studien die Rolle von KI in Online-Diskussionen explizit aus dieser normativen Perspektive. Diese Studie schließt diese Lücke, indem ein Überblick über interdisziplinäre KI-Forschung aus einer deliberativen Perspektive gegeben wird. Die zentrale Forschungsfrage lautet: Welche deliberativen Normen werden in der Forschung zur Anwendung von KI in Online-Diskussionen adressiert?

Um diese Frage zu beantworten, wurde zwischen 2015 und 2022 publizierte KI-Forschung anhand der deliberativen Prinzipien Zivilität, Egalität, Rationalität und Reziprozität systematisiert. Die Übersicht zeigt auf, in welchen Bereichen KI zur Stärkung deliberativer Prinzipien beizutragen versucht und macht zugleich deutlich, wo Forschungsbedarfe existieren. Ziel der Analyse ist es, sowohl Forschenden als auch Praktiker:innen relevante Ansatzpunkte für die Weiterentwicklung demokratischer Online-Diskurse durch KI zu bieten.

Theoretischer Rahmen

Durch ihren paradigmatischen Status innerhalb der (digitalen) Demokratieforschung und ihren Fokus auf den Diskurs sind deliberative Demokratietheorien (Habermas, 1992) besonders gut geeignet, um KI-Tools im Kontext von Online-Diskussionen normativ zu verorten und Gestaltungsziele zu markieren.

Mit Blick auf das Deliberationskonzept gibt es konkurrierende Interpretationen der theoretischen Literatur und daher auch Vorstellungen davon, was genau mit Deliberation gemeint ist. Jedoch teilen die meisten Autor:innen die Auffassung, dass es sich dabei um einen konstruktiven und respektvollen Austausch von Argumenten zwischen Gleichberechtigten handelt (Friess & Eilders, 2015). In Anlehnung an diesen minimalen Konsens wird Deliberation in dieser Studie als eine anspruchsvolle Form der Kommunikation definiert, die durch die Normen der Zivilität, Egalität, Rationalität und Reziprozität gekennzeichnet ist.

Zivilität bezieht sich auf den respektvollen Umgang unter den Teilnehmenden. Diese Dimension ist von entscheidender Bedeutung, da gegenseitiger Respekt auch bei Meinungsverschiedenheiten ein konstruktives und integratives Diskussionsklima schaffen kann (Papacharissi, 2004). Studien deuten allerdings darauf hin, dass Online-Diskussionen tendenziell eine hohe Anzahl inziviler, also hasserfüllter und abfälliger, Nutzer:innenkommentare aufweisen (Coe et al., 2014), die sich negativ auf die Nutzer:innen auswirken können (Anderson et al., 2018).

Egalität bedeutet, dass alle Teilnehmenden gleiche Rechte und Teilnahmemöglichkeiten am Diskurs haben (Gutmann & Thompson, 2004). Diese Dimension stellt sicher, dass jede Stimme gehört werden kann und somit das Potenzial hat, das Ergebnis zu beeinflussen (Habermas, 1990). Durch *Egalität* kann eine Vielfalt an Meinungen berücksichtigt werden, ohne dass dominante Stimmen andere überlagern (Bächtiger & Beauvais, 2016). Untersuchungen zeigen jedoch, dass sich nicht alle gesellschaftlichen Gruppen gleichermaßen in Online-Diskussionen einbringen (Phillips et al., 2025) und dass Online-Diskussionen häufig von einer kleinen Anzahl sehr aktiver Nutzer:innen dominiert werden (Graham & Wright, 2014).

Die Norm der *Rationalität* meint, dass die Diskussion auf Argumenten und der Suche nach konstruktiven Lösungen basiert (Gutmann & Thompson, 2004). Denn erst wenn Teilnehmende ihre Ansichten begründen, andere Argumente hinterfragen und praktische Lösungen vorschlagen, entfaltet *Deliberation* ihr „Wahrheitspotenzial“ (Habermas, 2006, S. 413). Die empirische Forschung zeichnet mit Blick auf *Rationalität* kein klares Bild: So finden Studien einerseits formal eine beträchtliche Menge an Argumenten (Berg et al., 2024), während andere eine eher skeptische Sichtweise liefern, vor allem hinsichtlich der Komplexität der Argumentation in Online-Diskussionen (Santana, 2019). Weitere Studien legen nahe, dass *Rationalität* in hohem Maße vom Design und Kontext von Online-Diskussionen abhängt (Esau et al., 2017).

Reziprozität bedeutet, dass Teilnehmende aufeinander eingehen und sich in einen wechselseitigen Austausch begeben (Morrell, 2018). *Reziprozität* ist in vielfacher Hinsicht entscheidend für *Deliberation*, da sie epistemische Funktionen dieser fördert (Estlund & Landemore, 2018), Empathie und gegenseitiges Verständnis ermöglicht (Barber, 1984) sowie *Zivilität* unterstützt (Bächtiger & Beauvais, 2016). Studien deuten jedoch auf einen allgemeinen Mangel an *Reziprozität* in Online-Diskussionen hin, was bedeutet, dass diese häufiger monologische als dialogische Formen annehmen (Esau & Friess, 2022).

Zusammen bieten diese vier Dimensionen einen geeigneten theoretischen Rahmen, um die aktuelle KI-Forschung zu systematisieren.

Methode

Ziel der Studie ist es, einen Überblick darüber zu gewinnen, wie KI im Kontext von Online-Diskussionen eingesetzt bzw. beforscht wird. Um dies zu leisten, wurde methodisch auf ein *Scoping Review* zurückgegriffen (Arksey & O'Malley, 2005), da dieses sich besonders gut für die Kartierung heterogener Forschungsbereiche und die Identifizierung von Lücken sowie thematischen Mustern eignet (Paré et al., 2015). Das Review wurde in fünf Schritten durchgeführt: Definition der Ziele, Identifikation relevanter Studien, manuelle Prüfung auf Einschlägigkeit, manuelle Kategorisierung entsprechend des theoretischen Rahmens und Synthese der Ergebnisse.

Mittels einer Schlagwortsuche auf Google Scholar wurden 407 potenziell relevante Publikationen, die zwischen 2015 und 2022 erschienen sind, identifiziert. Die Suche umfasste englisch- und deutschsprachige Literatur und bezog auch graue Literatur mit ein (Berichte, White Paper etc.). Nach manueller Sichtung wurden 171 Artikel in den Literaturkorpus aufgenommen. Diese stammen überwiegend aus der Informatik, aber auch aus den Sozialwissenschaften und dem Bereich der computergestützten Sozialwissenschaften.

Jeder Artikel wurde manuell dahingehend kategorisiert, ob er explizit oder implizit eine oder mehrere der oben vorgestellten deliberativen Dimensionen adressiert. Die Ergebnisse sind entlang der vier deliberativen Dimensionen strukturiert.

Ergebnisse

Die Ergebnisse zeigen, dass sich die bisherige KI-Forschung hauptsächlich auf die Dimensionen *Rationalität* und *Zivilität* konzentriert hat, während *Egalität* und *Reziprozität* in der KI-Forschung bislang weniger Beachtung gefunden haben (Abbildung 1).

Der Großteil der Literatur befasst sich mit Problemen in Online-Diskussionen, die der Dimension *Rationalität* zugeordnet werden können (115 Artikel). Der größte Teil der Studien in dieser Kategorie (66 Artikel) befasst sich mit der Erhöhung der Transparenz und Verständlichkeit von Diskussionen und untersucht die Beziehungen zwischen Argumenten.

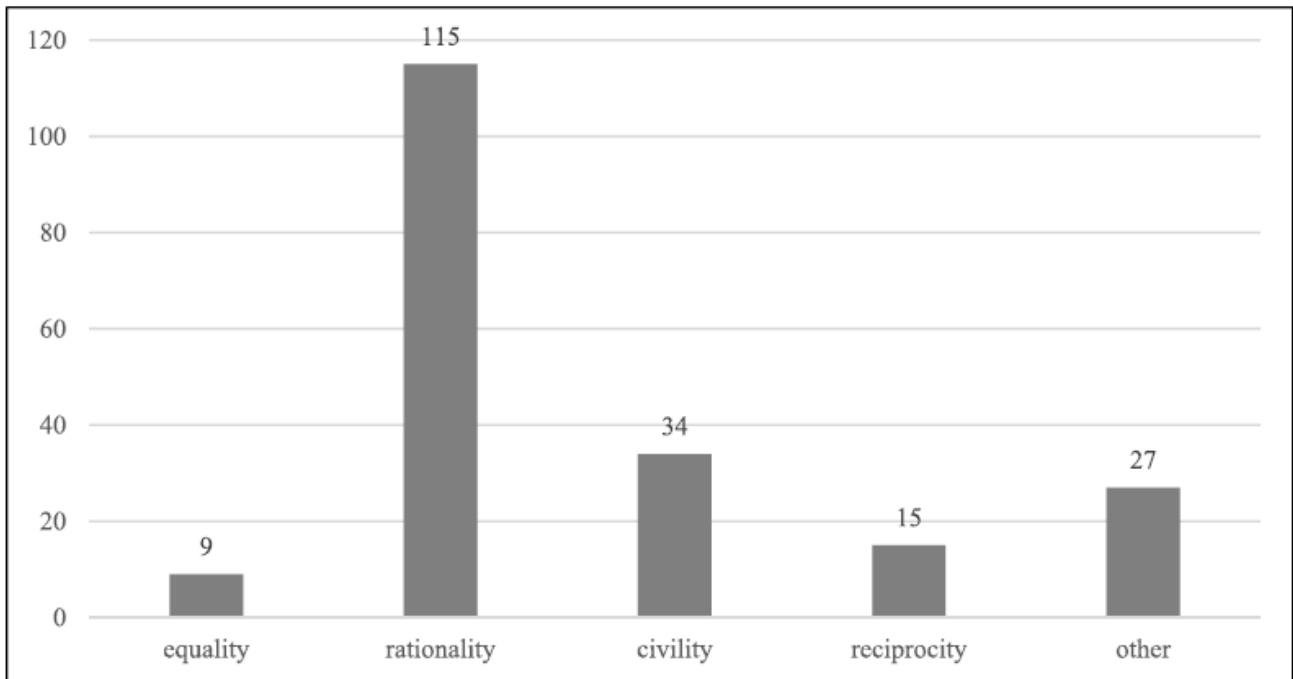


Abb. 1: Verteilung der Artikel auf die vier Dimensionen der Deliberation.

Argument-Mining-Verfahren stehen dabei im Zentrum (Chakrabarty et al., 2020). Eine weitere Methode zur Modellierung von Diskussionsstrukturen sind themenbasierte Informationssysteme (IBIS), die im Vergleich zum Argument Mining ein zusätzliches Themenelement integrieren und als Grundlage für automatisierte Diskussionsmoderation dienen (Gu et al., 2021). Darüber hinaus wird Topic Modelling eingesetzt, um Themen zu identifizieren sowie Beiträge thematisch zu strukturieren und Veränderungen im Diskussionsverlauf nachzuverfolgen (Mutanga & Abayomi, 2022).

Die zweitgrößte Gruppe (46 Artikel) umfasst Studien, die sich auf die Ebene der individuellen Beiträge konzentrieren. Die vorgeschlagenen KI-Techniken erkennen die Haltung eines Beitrags (Sobhani et al., 2015), die Polarität der Haltung (Sirrianni et al., 2021) und der emotionalen Stimmung (Pandey, 2021). Außerdem extrahieren sie automatisch Schlüsselwörter (Ito, 2018), erkennen Fakten und identifizieren Quellen (Dusmanu et al., 2017).

Weitaus weniger Studien (14 Artikel) untersuchen den Zusammenhang zwischen Nutzer:innen und Beiträgen, etwa durch Clustering bestimmter Typen (Arana-Catania et al., 2021), personalisierte Zusammenfassungen (Tepper et al., 2018) oder auch Polarisierungsberechnung (Althuniyan et al., 2019).

Die Forschung zur Zivilität (34 Artikel) wird dominiert von der automatisierten Erkennung von Inzivilität, wobei Ansätze zur Erkennung von Hassrede (Mozafari et al., 2019), Sarkasmus (Verma et al., 2021) und Cybermobbing (Al-Garadi et al., 2019) oder auch einer Kombination dieser Aspekte (Stoll et al., 2020) vorgeschlagen werden. Einige Studien gehen über eine reine Erkennung hinaus, indem sie inzivile Ereignisse prognostizieren (Zhang et al., 2018) oder Counter-Speech-Bots zur gezielten Bekämpfung einsetzen (Clever et al., 2022).

Reziprozität wird selten in den analysierten Studien adressiert (15 Artikel). Die dabei hervorgebrachten Ansätze reichen von der Erkennung von Bezugnahmen zwischen Beiträgen (Morio & Fujita, 2018), über Bots, welche die Bezugnahme auf andere Teilnehmende fördern sollen (Wyss & Beste, 2017), bis hin zu automatisierter Konfrontation der Teilnehmenden mit gegenteiligen Positionen (Krauthoff et al., 2018).

Egalität ist die am wenigsten untersuchte Dimension (neun Artikel). Hier finden sich Ansätze zur Förderung ausgewogener Beteiligung oder proaktiver Teilnahme durch Moderations-Bots (Kim et al., 2021). Andere Studien identifizieren automatisiert Rollen von Teilnehmenden und unterscheiden etwa zwischen aktiven und passiven Nutzer:innen (Wijenayake et al., 2020).

Diskussion

Ziel der Studie war es, einen systematischen Überblick darüber zu gewinnen, wie KI im Kontext von Online-Diskussionen beforscht wird, und diese Forschung entlang zentraler deliberativer Normen einzuordnen. Die Ergebnisse zeigen deutlich, dass sich die bisherige KI-Forschung vor allem auf die Dimensionen Rationalität und Zivilität konzentriert, während Egalität und Reziprozität vergleichsweise selten adressiert werden.

Dieses Ungleichgewicht lässt sich zum Teil durch methodische und konzeptionelle Herausforderungen erklären. Aspekte wie Rationalität und Zivilität sind häufig auf der Ebene einzelner Beiträge operationalisierbar und damit für aktuelle KI-Systeme vergleichsweise gut zugänglich. Demgegenüber entfalten sich Egalität und Reziprozität entweder im Vorfeld der Diskussionen oder in Interaktionsmustern über längere Diskussionsverläufe hinweg. Ihre Analyse ist daher schwer umsetzbar oder erfordert komplexe, kontext-sensitive Modelle sowie aufwendige Annotationen auf Diskursebene.

Gleichzeitig ist dieses Ungleichgewicht normativ problematisch. Eine Fokussierung auf Rationalität und Zivilität allein greift zu kurz, da deliberative Qualität wesentlich durch den Aufbau und die Struktur von Interaktionen bestimmt wird. Ohne die systematische Berücksichtigung von Egalität und Reziprozität besteht die Gefahr, dass KI-gestützte Systeme zwar zu sachlicheren und höflicheren Diskussionen beitragen, dabei jedoch bestehende Ungleichheiten reproduzieren und dialogische Defizite fortschreiben. Mit anderen Worten: KI kann unter diesen Bedingungen die Form, nicht aber umfassend die Qualität deliberativer Prozesse verbessern.

Vor diesem Hintergrund erscheint es notwendig, den Fokus zukünftiger Forschung stärker auf interaktive und relationale Dimensionen von Deliberation zu richten. Dies umfasst insbesondere die Entwicklung von KI-gestützten Ansätzen, die eine ausgewogenere Beteiligung fördern, etwa durch adaptive Moderationsmechanismen oder gezielte Aktivierung weniger sichtbarer Teilnehmender. Ebenso relevant sind Systeme, die dialogische Bezugnahmen unterstützen, beispielsweise durch das Sichtbarmachen von Anschlusskommunikation oder die gezielte Förderung von Perspektivübernahmen.

Darüber hinaus sollten verstärkt Modelle entwickelt werden, die Diskursverläufe über die Zeit hinweg analysieren und so die Dynamik deliberativer Prozesse besser erfassen können. Dies impliziert auch die Notwendigkeit einer engeren Verzahnung von technischer KI-Forschung und normativer Demokratietheorie. Nur wenn deliberative Prinzipien nicht lediglich als nachträgliche Bewertungsmaßstäbe, sondern als integrale Designkriterien verstanden werden, kann KI ihr Potenzial zur Förderung demokratischer Öffentlichkeit entfalten. Entsprechend kommt der interdisziplinären Zusammenarbeit zwischen Informatik, Kommunikationswissenschaft und politischer Theorie eine zentrale Rolle zu.

Insgesamt verdeutlichen die Ergebnisse, dass die jüngere KI-Forschung wichtige Fortschritte bei der Verbesserung einzelner demokratischer Qualitätsdimensionen erzielt hat, zugleich jedoch zentrale Voraussetzungen deliberativer Kommunikation unzureichend berücksichtigt. Die Identifikation dieser Lücken bietet einen wichtigen Ausgangspunkt für zukünftige Forschung, die darauf abzielt, KI nicht nur zur Optimierung bestehender Diskurse, sondern zur Förderung inklusiverer und dialogisch strukturierter Öffentlichkeiten einzusetzen.

Literatur

- Al-Garadi, M. A., Hussain, M. R., Khan, N., Murtaza, G., Nweke, H. F., Ali, I., Mujtaba, G., Chiroma, H., Khattak, H. A., & Gani, A. (2019). Predicting cyberbullying on social media in the big data era using machine learning algorithms: Review of literature and open challenges. *IEEE Access*, 7, 70701–70718. <https://doi.org/10.1109/ACCESS.2019.2918354>
- Althuniyan, N., Sirrianni, J. W., Rahman, M. M., & Liu, X. (2019). Design of mobile service of intelligent large-scale cyber argumentation for analysis and prediction of collective opinions. In D. Wang & L.-J. Zhang (Hrsg.), *Artificial intelligence and mobile services – AIMS 2019* (pp. 135–149). Springer. https://doi.org/10.1007/978-3-030-23367-9_10
- Anderson, A. A., Yeo, S. K., Brossard, D., Scheufele, D. A., & Xenos, M. A. (2018). Toxic talk: How online incivility can undermine perceptions of media. *International Journal of Public Opinion Research*, 30(1), 156–168. <https://doi.org/10.1093/ijpor/edw022>

- Arana-Catania, M., van Lier, F.-A., Procter, R., Tkachenko, N., He, Y., Zubiaga, A., & Liakata, M. (2021). Citizen participation and machine learning for a better democracy. *Digital Government: Research and Practice*, 2(3), 1–22. <https://doi.org/10.1145/3452118>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Bächtiger, A., & Beauvais, E. (2016). Taking the goals of deliberation seriously: A differentiated view on equality and equity in deliberative designs and processes. *Journal of Public Deliberation*, 12(2). <https://doi.org/10.16997/jdd.254>
- Barber, B. R. (1984). *Strong democracy*. University of California Press.
- Berg, J., Fagerholm, A., & Strandberg, K. (2024). Quality user-generated content? A case study of the quality of online news comments on the site of Finnish public service broadcaster Yle. *Journalism Practice*, 20(1) 338–359. <https://doi.org/10.1080/17512786.2024.2329759>
- Chakrabarty, T., Hidey, C., Muresan, S., McKeown, K., & Hwang, A. (2020). AMPERSAND: Argument mining for persuasive online discussions. <https://arxiv.org/abs/2004.14677>
- Clever, L., Klapproth, J., & Frischlich, L. (2022). Automatisierte (Gegen-)Rede? Social bots als digitales Sprachrohr ihrer Nutzer*innen. In J. Ernst, M. Trompeta, & H.-J. Roth (Hrsg.), *Gegenrede digital* (S. 11–26). Springer. https://doi.org/10.1007/978-3-658-36540-0_2
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper website comments. *Journal of Communication*, 64(4), 658–679. <https://doi.org/10.1111/jcom.12104>
- Coleman, M. C. (2018). Bots, social capital, and the need for civility. *Journal of Media Ethics*, 33(3), 120–132. <https://doi.org/10.1080/23736992.2018.1476149>
- Dusmanu, M., Cabrio, E., & Villata, S. (2017). Argument mining on Twitter: Arguments, facts and sources. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (S. 2317–2322). <https://doi.org/10.18653/v1/D17-1245>
- Esau, K., & Friess, D. (2022). What creates listening online? Exploring reciprocity in online political discussions with relational content analysis. *Journal of Deliberative Democracy*, 18(1), 1–16. <https://doi.org/10.16997/jdd.1021>
- Esau, K., Friess, D., & Eilders, C. (2017). Design matters! An empirical analysis of online deliberation on different news platforms. *Policy & Internet*, 9(3), 321–342. <https://doi.org/10.1002/poi3.154>
- Estlund, D., & Landemore, H. (2018). The epistemic value of democratic deliberation. In A. Bächtiger et al. (Hrsg.), *The Oxford handbook of deliberative democracy* (S. 112–131). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198747369.013.26>
- Friess, D., & Eilders, C. (2015). A systematic review of online deliberation research. *Policy & Internet*, 7(3), 319–339. <https://doi.org/10.1002/poi3.95>
- Graham, T., & Wright, S. (2014). Discursive equality and everyday talk online: The impact of “super-participants.” *Journal of Computer-Mediated Communication*, 19(3), 625–642. <https://doi.org/10.1111/jcc4.12016>
- Gu, W., Moustafa, A., Ito, T., Zhang, M., & Yang, C. (2021). A case-based reasoning approach for supporting facilitation in online discussions. *Group Decision and Negotiation*, 30(3), 719–742. <https://doi.org/10.1007/s10726-021-09731-4>
- Gutmann, A., & Thompson, D. F. (2004). *Why deliberate democracy?* Princeton University Press.
- Habermas, J. (1990). *Moral consciousness and communicative action*. MIT Press.
- Habermas, J. (1992). *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats*. Suhrkamp.
- Habermas, J. (2006). Political communication in media society. *Communication Theory*, 16(4), 411–426. <https://doi.org/10.1111/j.1468-2885.2006.00280.x>
- Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication. *Journal of Computer-Mediated Communication*, 25(1), 89–100. <https://doi.org/10.1093/jcmc/zmz022>
- Ito, T. (2018). Toward agent-based large-scale decision support system. *Proceedings of the Hawaii International Conference on System Sciences*,

- 351–360.
<https://doi.org/10.24251/HICSS.2018.047>
- Kim, S., Eun, J., Seering, J., & Lee, J. (2021). Moderator chatbot for deliberative discussion. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–26.
<https://doi.org/10.1145/3449161>
- Krauthoff, T., Meter, C., Baurmann, M., Betz, G., & Mauve, M. (2018). D-BAS: A dialog-based online argumentation system. In *Computational models of argument* (S. 325–336). IOS Press.
- Morio, G., & Fujita, K. (2018). Annotating online civic discussion threads for argument mining. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* (S. 546–553).
<https://doi.org/10.1109/WI.2018.00-39>
- Morrell, M. (2018). Listening and deliberation. In A. Bächtiger et al. (Hrsg.), *The Oxford handbook of deliberative democracy* (S. 236–250). Oxford University Press.
<https://doi.org/10.1093/oxfordhb/9780198747369.013.55>
- Mozafari, M., Farahbakhsh, R., & Crespi, N. (2019). A BERT-based transfer learning approach for hate speech detection. <https://arxiv.org/abs/1910.12574>
- Mutanga, M. B., & Abayomi, A. (2022). Tweeting on COVID-19 pandemic in South Africa. *African Journal of Science, Technology, Innovation and Development*, 14(1), 163–172.
<https://doi.org/10.1080/20421338.2020.1817262>
- Pandey, C. (2021). redBERT. *International Journal of Open Source Software and Processes*, 12(3), 32–47.
<https://doi.org/10.4018/IJOSSP.2021070103>
- Papacharissi, Z. (2004). Democracy online. *New Media & Society*, 6(2), 259–283.
<https://doi.org/10.1177/1461444804041444>
- Paré, G., Trudel, M. C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge. *Information & Management*, 52(2), 183–199. <https://doi.org/10.1016/j.im.2014.08.008>
- Philippis, G., Kelm, O., Friess, D., & Ziegele, M. (2025). From Unpolitical Sporadics to Political Activists: Identifying political social media user types. *European Journal of Communication*, 40(2), 188–205
<https://doi.org/10.1177/02673231251327860>
- Santana, A. D. (2019). Toward quality discourse. *Newspaper Research Journal*, 40(4), 467–486.
<https://doi.org/10.1177/0739532919873089>
- Sirrianni, J. W., Liu, X., & Adams, D. (2021). Predicting stance polarity. *IEEE Transactions on Computational Social Systems*, 8(3), 655–667.
<https://doi.org/10.1109/TCSS.2021.3056596>
- Sobhani, P., Inkpen, D., & Matwin, S. (2015). From argumentation mining to stance classification. In *Proceedings of the workshop on argumentation mining* (S. 67–77).
<https://doi.org/10.3115/v1/W15-0509>
- Stoll, A., Ziegele, M., & Qiring, O. (2020). Detecting impoliteness and incivility. *Computational Communication Research*, 2(1), 109–134.
<https://doi.org/10.5117/CCR2020.1.005>
- Tepper, N., Hashavit, A., Barnea, M., Ronen, I., & Leiba, L. (2018). Collabot. In *Proceedings of the ACM International Conference on Web Search and Data Mining* (S. 771–774).
<https://doi.org/10.1145/3159652.3160588>
- Verma, P., Shukla, N., & Shukla, A. P. (2021). Techniques of sarcasm detection. In *Proceedings of the ICACITE conference* (S. 968–972). IEEE.
<https://doi.org/10.1109/ICACITE51222.2021.9404585>
- Wijenayake, P., Silva, D. de, Alahakoon, D., & Kirigeegamage, S. (2020). Automated detection of social roles in online communities. In *Proceedings of the International Conference on Software Engineering and Information Management* (S. 63–68).
<https://doi.org/10.1145/3378936.3378973>
- Wyss, D., & Beste, S. (2017). Artificial facilitation. *Journal of Information Technology & Politics*, 14(3), 214–231.
<https://doi.org/10.1080/19331681.2017.1338175>
- Zhang, J., Danescu-Niculescu-Mizil, C., Sauper, C., & Taylor, S. J. (2018). Characterizing online public discussions. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 1–27.
<https://doi.org/10.1145/3274467>