

Interdisziplinäres Streitgespräch – Nutzerkommentar- analysen aus ethisch-rechtlicher Perspektive

Brokering, Annalena

A.Brokering@sms.ed.ac.uk
The University of Edinburgh, Law School, Vereinigtes
Königreich

Guhr, Svenja

Guhr@linglit.tu-darmstadt.de
TU Darmstadt, Deutschland

Einleitung

Täglich werden auf der ganzen Welt Onlineartikel, Blogbeiträge etc. veröffentlicht, zu denen Leserinnen und Leser (i.F. generisches Femininum) Kommentare verfassen. Aufgrund der hohen Anzahl an Partizipierenden gelten Nutzerbeiträge als besonders authentische Echtzeitrückmeldungen und erlauben einen Zugang zu heterogenen Meinungsäußerungen (Busch, 2017). Auch durch Partizipation auf Social Media Plattformen, in Onlineforen und öffentlichen Chats werden Daten generiert, die wertvolle Informationen über Nutzerverhalten und menschliches Denken beinhalten. Dies gilt umso mehr, als diese Plattformen Orte sind, an denen Menschen miteinander in Verbindung treten, in verschiedenen Formen und Dimensionen Gemeinschaft pflegen, Informationen verbreiten und ihre Meinungen austauschen. Dabei generierte Daten zeichnen sich durch ihren interaktiven, kontemporären und personenbezogenen Charakter aus und ermöglichen folglich Rückschlüsse auf Meinungen, Interessen und Stimmungen in der Bevölkerung. Entsprechend sind sie von besonderem Interesse für private Unternehmen oder öffentliche Institutionen (Schoen, 2002; Holtz-Bacha, 2019: 276). Zunehmend wird auch im akademischen Kontext auf Nutzerdaten zurückgegriffen (u.a. Mohammad, 2016; Aker et al., 2016).

Forschungsgegenstand

Gegenstand des vorzustellenden Projektes ist eine Pilotstudie, in der zwei Masterthesisprojekte in Beziehung zueinander gesetzt wurden. Bei dem ersten Thesisprojekt (Guhr, 2019) handelt es sich um eine

im Bereich der Digital Humanities durchgeführte computergestützte Analyse von Leserkomentaren in französischen Onlinemedien. Das zweite Thesisprojekt aus dem Bereich des IT-/Datenschutzrechts betrachtet ethische und rechtliche Erwägungen bei der Analyse von nutzergenerierten Inhalten auf Social Media Plattformen und die Frage, wie deren Berücksichtigung in wissenschaftlichen Datenanalyseprojekten unterstützt werden kann. Infolge der Auseinandersetzung mit dem jeweils anderen Masterthesisprojekt entstand ein interdisziplinäres Streitgespräch zwischen den Autorinnen.

Die Masterthesis (Guhr, 2019) umfasst u.a. verschiedene gemischt qualitativ-quantitative Analysen von Onlinezeitungsartikeln und zugehörigen Leserkomentaren zur französischen Präsidentschaftswahl 2017. Die Daten sind über den Onlineauftritt einer großen französischen Tageszeitung öffentlich zugänglich. 40 ausgewählte Onlineartikel mit den dazugehörigen 3.127 Leserkomentaren wurden zu einem Korpus zusammengestellt. Die extrahierten Leserkommentardaten beinhaltenen zusätzlich zu den nutzergenerierten Beitragstexten auch Datum und Uhrzeit der Beitragserstellung sowie die Nicknames und teilweise bürgerliche Namen der Nutzerinnen. Mithilfe von Distant Reading Methoden wurden im Korpus behandelte Themen identifiziert. Anschließend wurde eine automatisierte Sentimentanalyse der Kommentare durchgeführt, um Informationen über die emotionale Einstellung der Nutzerinnen zu Wahlkampfthemen und zum/zur Präsidentschaftskandidat/in herausstellen zu können.

Der zweiten Masterthesis (Brokering, 2019) lag die Frage zugrunde, wie Datenanalyseprojekte im akademischen Kontext rechtskonform und ethischer gestaltet werden können. Am Beispiel von Forschung mit Social Media Daten wurde herausgestellt, wie durch Analysen von nutzergenerierten Inhalten Interessen und Rechte der Nutzerinnen berührt werden. Für die hierdurch aufgeworfenen, neuen ethischen und besonders datenschutzrechtlichen Fragestellungen fehlt es bestehenden inhaltlichen und institutionellen Ansätzen der Forschungsethik noch an befriedigenden Antworten, die eine ethische Praxis von Social Media Datenanalysen gewährleisten. Daraufhin wurde evaluiert, inwiefern das IT-rechtliche Konzept des *Regulation by Design* eine effektivere Implementierung ethischer und rechtlicher Erwägungen in Social Media Datenanalysen unterstützen kann. *Regulation by Design* zielt auf eine proaktive Berücksichtigung regulatorischer Erwägungen bereits im Zeitpunkt des Designs, d.h. der Planung und Entwicklung, von Produkten und Aktivitäten wie auch Forschung. Es findet seine bekannteste Ausprägung im datenschutzrechtlichen Prinzip des *Privacy by Design*.

Interdisziplinäres Streitgespräch

Im Dialog der Autorinnen trafen die Perspektiven der praxisorientierten und der juristischen Forschung aufeinander. Aus letzterer wurde Kritik am Umgang mit persönlichen Informationen geäußert und für eine höhere Sensibilität gegenüber den Interessen der Nutzerinnen und insbesondere datenschutzrechtlichen Erwägungen plädiert. Sobald Social Media Daten eine Identifizierbarkeit der postenden Personen auch nur ermöglichen, z.B. weil sie die Nutzernamen oder auch die IP-Adresse der Nutzerin enthalten, handelt es sich um persönliche Daten und damit finden datenschutzrechtliche Vorgaben wie die europäische DSGVO Anwendung. Diese erfordert typischerweise die Information der betroffenen Nutzerin über die konkrete Verwertung ihrer Daten und die Einwilligung in diese. Die Wirksamkeit einer mittels der AGB des jeweiligen Social Media Anbieters erteilten Einwilligung ist als zweifelhaft zu bewerten, da sie nicht projektspezifisch ist. Angesichts des Umstands, dass die verwendeten Nutzerdaten zu Forschungszwecken umgewidmet werden und originär im Rahmen der privaten Nutzung von Social Media Diensten entstanden sind, ist in Erwägung zu ziehen, ob über das datenschutzrechtlich erforderliche Mindestmaß hinausgehende Maßnahmen zum Schutz der Interessen der Nutzerinnen ethisch geboten sind. Auch eine Anonymisierung der Nutzerdaten, z.B. durch Entfernen des Nutzernamens kann das Re-Identifikationsrisiko angesichts fortschrittlicher De-Anonymisierungstechniken nur reduzieren. Hier sind weitergehende u.a. auch im Verhältnis zum Grad an Sensibilität der betroffenen Nutzerinhalte angemessene Anonymisierungsmaßnahmen in Erwägung zu ziehen. Gleichzeitig ist zu berücksichtigen, dass Nutzerinnen an ihren originell und kreativ gestalteten Beiträgen auch urheberrechtliche Interessen und Rechte haben, sodass andererseits eine Erkennbarkeit der Autorin durch die Forschenden sicherzustellen sein kann. Die praxisorientierte Forscherin wies demgegenüber auf die schwierige Umsetzbarkeit aufwendiger Maßnahmen zum Schutz der Nutzerinnen angesichts begrenzter finanzieller, technischer und zeitlicher Spielräume in der Forschungspraxis hin sowie auf die Gefahr, dass Forschungsdaten durch Datenschutzmaßnahmen an Wert/Aussagekraft verlieren würden. Als Beispiele nannte sie den Wert von Nutzernamen als potenzielle Informationsquelle hinsichtlich Gender und Nationalität sowie für die kumulative Betrachtung verschiedener Beiträge einer Person. Auch die Erhebung von Datum und Uhrzeit der Beitragserstellung ermögliche eine chronologische Ordnung von Beiträgen. Dabei kritisierte die Juristin, dass bereits aus derartigen Informationen umfangreiche Aktivitätsprofile über einzelne Nutzerinnen erstellt werden könnten, die ggf. in Verbindung mit Nutzungsdaten derselben Nutzerinnen auf weiteren Social Media Plattformen Rückschlüsse auf Tagesabläufe, Vorlieben und Social Media Verhalten einzelner Nutzerinnen erlauben. Im daraus resultierenden

Streitgespräch wurde erkennbar, wie schwierig es ist, die jeweiligen Positionen in einer der anderen Forschenden verständlichen Weise zu kommunizieren.

Weiteres Vorgehen im Projekt war es, die rechtlich-ethischen Herausforderungen gemeinsam zu definieren, wobei die Perspektiven beider Forschungsrichtungen Beachtung finden sollten. Auf dieser gemeinsamen Grundlage und unter Berücksichtigung verschiedener Ansätze des *Regulation by Design*-Konzeptes wurden Methoden und Herangehensweisen diskutiert, die eine effektive Berücksichtigung der Herausforderungen in der Forschungspraxis erreichen sollen.

Das Projekt verfolgt damit das Ziel, den Dialog zwischen datenbasierter Forschung und IT-Recht anzuregen und insbesondere das Bewusstsein für Nutzerinteressen und Datenschutzerwägungen unter Forschenden zu erhöhen. Es soll reflektiert werden, wie die Kommunikation zwischen IT-Rechtlerinnen und Datenforschenden unter Berücksichtigung der unterschiedlichen Perspektiven, Interessen und Limitierungen verbessert werden kann. Die gemeinsamen Definitionen der Herausforderungen und die diskutierten Lösungsvorschläge sollen Datenforschenden ermöglichen, ihre Forschungsarbeit ohne größeren Mehraufwand bereits im Stadium der Vorbereitung und Durchführung von Datenanalysen rechtskonform und ethisch sensibel zu gestalten.

Bibliographie

Aker, Ahmet / Paramita, Monica / Kurtic, Emina / Funk, Adam / Barker, Emma (2016): „Automatic label generation for news comment clusters“, in: *Proceedings of the 9th International Natural Language Generation Conference*, Edinburgh, UK: 61–69 <https://pdfs.semanticscholar.org/4da7/ac02c56d43312425a854d63e71f89dd288ec.pdf> [letzter Zugriff 26. Juni 2019].

Brokering, Annalena (2019): *Drawing from approaches in regulatory theory for the regulation of new technologies and design theory, how can ethical considerations be effectively incorporated into data science activities?*. Masterthesis, The University of Edinburgh, Edinburgh Law School.

Buchanan, Elizabeth / Zimmer, Michael (2016): „Internet Research Ethics“, in: *Stanford Encyclopedia of Philosophy* <https://plato.stanford.edu/entries/ethics-internet-research/> [letzter Zugriff 19. August 2019].

Busch, Andreas (2017): Informationsinflation: Herausforderungen an die politische Willensbildung in der digitalen Gesellschaft, in: Gapski, Harald / Oberle, Monika / Stauffer, Walter (eds.): *Medienkompetenz. Herausforderungen für Politik, politische Bildung und Medienbildung*, Schriftenreihe der Bundeszentrale für Politische Bildung. Bonn: Bundeszentrale für Politische Bildung 53-62 <http://www.bpb.de/lernen/digitale-bildung/medienpaedagogik/medienkompetenz->

schriftenreihe/257594/informationsinflation [letzter Zugriff 11. Juni 2019].

Dashtipour, Kia / Poria, Soujanya / Hussain, Amir / Cambria, Erik / Hawalah, Ahmad Y. A. / Gelbukh, Alexander / Zhou, Qiang (2016): „Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques“, in: *Cognitive Computation* (2016) 8: 757-771 <https://link.springer.com/article/10.1007/s12559-016-9415-7> [letzter Zugriff 26. Juni 2019].

Golder, Susan P. / Ahmed, Shahd / Norman, Gill / Booth, Andrew (2017): „Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review“, in: *Journal of Medical Internet Research* 19 <http://eprints.whiterose.ac.uk/117721/> [letzter Zugriff 26. September 2019].

Guhr, Svenja (2019): *Computergestützte Analyse von französischen Onlinemedien zur Präsidentschaftswahl 2017*, Masterthesis, Georg-August-Universität Göttingen.

Holtz-Bacha, Christina (2019): „Demoskopie - Medien - Politik. Umfragen im Bundestagswahlkampf 2017“, in: Holtz-Bacha, Christina: *Die (Massen-)Medien im Wahlkampf. Die Bundestagswahl 2017*. Wiesbaden: Springer Fachmedien Wiesbaden GmbH 263-280.

Locatelli, Elisabetta (2018): „Ethics of Social Media Research: State of the Debate and Future Challenges“, in: Hunsinger, Jeremy / Klastrup, Lisbeth / Allen, Matthew M. (eds.): *Second International Handbook of Internet Research*. Dordrecht: Springer 1-22.

McKee, Heidi / Porter, James E. (2008): „The Ethics of Digital Writing Research: A Rhetorical Approach“, *College Composition and Communication* 59: 711 http://wrconf08.writing.ucsb.edu/Pdf_Articles/McKee_Article.pdf [letzter Zugriff 26. September 2019].

Mohammad, Saif (2016): „A Practical Guide to Sentiment Annotation: Challenges and Solutions“, in: *Proceedings of the NAACL 2016 Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media (WASSA)* 174-179 <http://www.aclweb.org/anthology/W16-0429> [letzter Zugriff 16. Mai 2019].

Moreno, Megan A. / Goniou, Natalie / Moreno, Peter S. / Diekema, Douglas (2013): „Ethics of Social Media Research: Common Concerns and Practical Considerations“, in: *Cyberpsychology, Behavior, and Social Networking* 16: 708-713 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3942703/> [letzter Zugriff 16. Mai 2019].

Perez Vallejos, Elvira / Koene, Ansgar / Carter, Christopher J. / Hunt, Daniel / Woodard, Christopher / Urquhart, Lachlan / Bergin, Aislinn / Statche, Ramona (2019): „Accessing Online Data for Youth Mental Health Research: Meeting the Ethical Challenges“, in: *Philosophy & Technology* 32: 87-110 <https://link.springer.com/article/10.1007/s13347-017-0286-y> [letzter Zugriff 26. September 2019].

Schoen, Harald (2002): „Wirkung von Wahlprognosen auf Wahlen“ in: Berg, Thomas (ed.) (2002): *Moderner*

Wahlkampf. Blick hinter die Kulissen. Opladen: Leske und Budrich 171-191.

Williams, Matthew L. / Burnap, Pete / Sloan, Luke (2017): „Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation“, in: *Sociology* 51: 1149-1168 <https://journals.sagepub.com/doi/pdf/10.1177/0038038517708140> [letzter Zugriff 26. September 2019].