

Forschungsethische Herausforderungen der automatisierten Analyse digitaler Datenspuren

Jahrestagung der DG PuK
Bremen 19. Mai 2023

Automatisierung von Kommunikation

- Neue technische Möglichkeiten große Mengen digitaler Datenspuren zu erheben, auszuwerten und aufzubereiten
- Neue Erkenntnisse für soziale Kommunikationsphänomene (Geise & Waldherr, 2021)
- Risiken wie digitale Überwachung oder Verletzungen der Privatsphäre (Kennedy et al., 2020)
- Bekannte forschungsethische Herausforderungen sowie neue ethische Fragen (Hollingshead et al., 2021)

Angewandte Forschungsethik

- Reflexion der eigenen Handlungsentscheidungen und Abwägung der Interessen und Rechte aller an empirischer Forschung beteiligten Personen (Schlütz & Möring, 2016)
- Prozess der Abwägung zwischen
 - Makro-Ebene: Rechtlichen Rahmenbedingungen
 - Meso-Ebene: Wissenschaftlichen Standesregeln und institutionell etablierten Praktiken
 - Mikro-Ebene: Individuellen ethischen Handlungsentscheidungen

(McKee & Porter, 2008, 2009; Schlütz & Möring, 2016)

Angewandte Forschungsethik

- **Selbstbestimmung / Autonomie**
 - Freie, informierte und widerrufbare Zustimmung zur Teilnahme an der Studie oder deren Ablehnung
- **Gerechtigkeit**
 - Teilhabe an wissenschaftlicher Erkenntnis / Nutzen
- **Schadensvermeidung**
 - Erwarteter Nutzen muss in einem ausgewogenen Verhältnis zu möglichen Schädigungen der beteiligten Personen stehen

(Beauchamp & Childress, 2013; Heinrichs, 2010; McKee & Porter, 2009; Schlütz & Möhring, 2016)

Ethisch-methodische Herausforderungen

- Methodische und ethische Anforderungen können in Widerspruch zueinander stehen (Schlütz & Möhring, 2016)
- Forschungsethik als handlungseröffnender und kreativitätsfördernder Denkraum (Coenen, 2022)
- Ziel: Forschungsprojekt insgesamt verbessern (Ess, 2020)

Welche forschungsethischen Herausforderungen nehmen Forschende bei der automatisierten Analyse von digitalen Datenspuren wahr und wie gehen sie mit ihnen um?

Methode

- Halbstandardisierte Leitfadengespräche mit (quantitativ) forschenden Kommunikationswissenschaftler*innen unterschiedlicher Qualifikationsstufen
- Feldphase: April – Juli 2021
- Erhebungssituation: Zoom-Gespräch
- n = 15
 - Forscher*innen, die automatisierte Analyse digitaler Datenspuren explizit thematisierten
 - 6 Frauen, 9 Männer; 7 Professor*innen, 7 Post-Docs, 1 Prae-Doc
 - ø Dauer: 58 Min.; Range: 41 bis 82 Min.
- Deduktiv-induktive Auswertung mittels qualitativer Inhaltsanalyse in MAXQDA (Kuckartz & Rädiker, 2020)

Ergebnisse

Herausforderung I: Art und Umfang der Daten

- **Balance zwischen „so viele Daten wie möglich, so wenige Daten wie nötig“**

Lösungsansätze:

- Datensparsamkeit
- Keine Erhebung von personenbezogenen Daten
- Verwendung von Datenspenden
- Verwendung von sog. White-Lists

Ich meine so rein aus Forschendenperspektive ist es natürlich schon immer schön so viele Daten wie möglich zu haben, auch gerade als Kontrollvariablen. Demographische Daten, die dann aber ja vielleicht dann doch sensibel sein können. Ich denke, das muss man gut abwägen. Da kann es Konflikte geben.

#107

Ergebnisse

Herausforderung I: Art und Umfang der Daten

- Balance zwischen „so viele Daten wie möglich, so wenige Daten wie nötig“

Lösungsansätze:

- **Datensparsamkeit**
- Keine Erhebung von personenbezogenen Daten
- Verwendung von Datenspenden
- Verwendung von sog. White-Lists

Also ich glaube, das ist fast der wichtigste Schritt, wenn man das Untersuchungsdesign sich überlegt, schon mal zu überlegen, "Welche Daten brauche ich überhaupt?" Also, dass ich auch wirklich nicht zu viel erhebe, also nicht jetzt alles sammle und dann liegt das irgendwo. #107

Ergebnisse

Herausforderung I: Art und Umfang der Daten

- Balance zwischen „so viele Daten wie möglich, so wenige Daten wie nötig“

Lösungsansätze:

- Datensparsamkeit
- **Keine Erhebung von personenbezogenen Daten**
- Verwendung von Datenspenden
- Verwendung von sog. White-Lists

Wir haben damals eine ganze Menge Vorsichtsmaßnahmen getroffen. Die haben schon bei der Erhebung sowas wie Klarnamen gar nicht mit erhoben. Das ist jetzt einer der Vorteile gewesen, wenn man es automatisch macht. Man kann die Datensparsamkeit dort gleich machen.

#75

Ergebnisse

Herausforderung I: Art und Umfang der Daten

- Balance zwischen „so viele Daten wie möglich, so wenige Daten wie nötig“

Lösungsansätze:

- Datensparsamkeit
- Keine Erhebung von personenbezogenen Daten
- **Verwendung von Datenspenden**
- **Verwendung von sog. White-Lists**

Aber z.B. bei Datenspenden kann man natürlich keinem normalen Teilnehmenden zumuten, durch die Daten zu gehen und dann zu sagen: „Ja, geh nochmal durch, ob da in Zeile 28.400 eine sensible URL ist“. Das heißt, da muss man dann immer anfangen irgendwie Lösungen zu entwickeln, die es den Leuten irgendwie transparent machen. Mit White-Listing arbeiten z.B., dass man sagt: „Wir sammeln nur die Daten über diese Domain oder über diese Aktivitäten. #75

Ergebnisse

Herausforderung II: Privatsphäre

- **Mustererkennung in aggregierten digitalen Datenspuren**
- **Möglichkeit der Einzelfallanalyse bleibt bestehen**
- (Zukünftige) Kombination von Datensätzen
- Re-Identifizierung von Nutzenden

Es reicht ja auch schon das Potenzial, dass ich, also einzelne Personen, identifizieren könnte, dass ich, wenn ich irgendwo auf eine Sache stoße, dass ich mir angucke, wer hat denn das gesagt, was weiß ich denn noch über diese Person? Das, ja, das muss reflektiert werden. #79

Ergebnisse

Herausforderung II: Privatsphäre

- Mustererkennung in aggregierten digitalen Datenspuren
- Möglichkeit der Einzelfallanalyse bleibt bestehen
- **(Zukünftige) Kombination von Datensätzen**
- **Re-Identifizierung von Nutzenden**

Das heißt, auch der technische Fortschritt erleichtert es, Daten zu kombinieren und daraus noch weitere Schlüsse zu ziehen.

#28

Ergebnisse

Lösungsansätze

- **Verschlüsselung der Geräte und DSGVO-konforme Dienste**
- **Verschlüsselung und Anonymisierung der Daten**
- Aggregierte Auswertung und Bereitstellung von Daten
- Pseudonymisierung / Verfremdung für Publikationen

Wir verschlüsseln quasi auch vor uns selbst. Also auch, um uns selbst zu schützen, dass wir gar nicht erst irgendwie in die Versuchung kommen, irgendwelche Dinge zu tun. #75

[...] und wir haben diese Nummern zusätzlich kryptografisch verschlüsselt. Das heißt, wir hatten einen Einmal-Schlüssel [...] erzeugt, wir haben die mit dem gehasht und dann diesen Schlüssel später weggeworfen. #38

Ergebnisse

Lösungsansätze

- **Verschlüsselung der Geräte und DSGVO-konforme Dienste**
- **Verschlüsselung und Anonymisierung der Daten**
- Aggregierte Auswertung und Bereitstellung von Daten
- Pseudonymisierung / Verfremdung für Publikationen

Was ich da mache, ist K-Anonymisierung zum Beispiel. [...] K-Anonymisierung heißt, ich versuche halt die Daten mit Rauschen zu [versehen], dass es immer ausreichend genug statistische Zwillinge da drin gibt, sodass man nie wissen kann, wen betrifft es denn eigentlich. #38

Ergebnisse

Herausforderung III: Sensible Daten

- **Verletzung der Privatsphäre bei Erhebung von Klarnamen, sexuellen Präferenzen etc.**
- **Möglichkeit der digitalen Überwachung bei Netzwerkdaten, Geodaten etc.**
- Vulnerable Zielgruppen

Lösungsansätze:

- Transparente Aufklärung über zu erhebende Daten

Smartphones, die auch Geodaten erfassen, Zeiten, Orte, Muster, genau mitschreiben können und auch durch die Sensoren sehr präzise sein können. Es ist eine heikle Kombination, also das, was die Technologie kann plus den Kontext, für den man das dann nutzt. #40

Ergebnisse

Lösungsansätze:

- **Beschränkung auf Personen des öffentlichen Lebens**
- Keine Kombination verschiedener Datensätze

Bei öffentlichen Sprechern gehen wir davon aus, die sind qua Rolle, qua Beruf [...] stehen die in der Öffentlichkeit, so dass die indirekt damit auch eingewilligt haben, dass sie beobachtbar sind und das schließt eben auch wissenschaftliche Beobachtungen ein.
#79

Ergebnisse

Lösungsansätze:

- Beschränkung auf Personen des öffentlichen Lebens
- **Keine Kombination verschiedener Datensätze**

Das hätte man kombinieren können. Und dass es niemals den Informed Consent dieser Leute gab und ich das dann abgelehnt habe, weil ich das unethisch fand, also diese Idee, dass man Befragungsdaten mit Social-Media-Profilen kombiniert, ohne dass die Personen wissen, dass das passieren kann. Das ist natürlich krass. Das fand ich krass und das werde ich nicht machen. #16

Diskussion

- Forscher*innen haben erste lösungsorientierte Strategien entwickelt
- Strategien basieren meist auf Forschungspraxis, nicht auf systematischen Informationen
- Handlungsempfehlungen und Best-Practice-Beispiele als Entscheidungshilfe
 - Empfehlungen der Association of Internet Researchers (Markham & Buchanan, 2012; franzke et al., 2020)
 - Data Ethics Decision Aid (DEDA) (franzke et al., 2021)
 - Webseite des FeKoM-Projektes: <https://forschungsethik-kmw.de>

Diskussion

- **Integration in der Hochschullehre**
- DGPuK als übergeordneter Ansprechpartner

Jedes Mal, wenn du auf einer DGPuK-Jahrestagung in der Sitzung sagst „Hier Forschungsethik“, dann rollt die Hälfte mit den Augen und denkt sich: „Eh, kommen die schon wieder mit ihrer Forschungsethik.“ Also man sieht ja, so Bewusstsein auf der älteren Ebene schaffen ist wichtig, aber gleichzeitig ist es [wichtig], einfach den Jüngeren selbstverständlich beizubringen [...]. Deswegen finde ich das eigentlich mit am zentralsten, dass man es in der Lehre direkt beibringt. #60

Diskussion

- Integration in der Hochschullehre
- **DGPuK als übergeordneter Ansprechpartner**

Im Prinzip ist das eine Aufgabe, die wir dann immer an die DGPuK herantragen müssen: „Was finden wir wichtig genug, um ggf. auch unvermeidbare Kompromisse hinsichtlich ethischer Fragen einzugehen?“. #75

Forschungsethische Herausforderungen der automatisierten Analyse digitaler Datenspuren

Kontakt: a.zillich@filmuniversitaet.de; eva-maria.roehse@tu-dortmund.de

Webseite: <https://forschungsethik-kmw.de>

Projektteam



Prof. Dr. Daniela Schlütz
Projektleitung
Filmuniversität Babelsberg
KONRAD WOLF



Prof. Dr. Wiebke Möhring
Projektleitung
TU Dortmund



Dr. Arne Freya Zillich
Akademische Mitarbeiterin
Filmuniversität Babelsberg
KONRAD WOLF



Eva-Maria Roehse
Wissenschaftliche Mitarbeiterin
TU Dortmund



Junior-Prof. Dr. Elena Link
Kooperationspartnerin
Johannes Gutenberg-Universität
Mainz

Literatur

Beauchamp, T. L., & Childress, J. F. (2019). *Principles of biomedical ethics*. Oxford: Oxford University Press.

Coenen, E. (2022). Forschungsethik und Datenvielfalt. Einige Gedanken zu einem ambivalenten Verhältnis am Beispiel einer Bestattungsethnografie. In C. Lohmeier & T. Wiedemann (Hrsg.), *Datenvielfalt in kommunikationswissenschaftlichen Forschungskontexten: Potenziale und Herausforderungen*, (S. 255–272). Wiesbaden: Springer.

Ess, C. (2002). *Ethical decision-making and Internet research: Recommendations from the AoIR ethics working committee*. <https://aoir.org/reports/ethics2.pdf>.

franzke, a. s., Bechmann, A., Zimmer, M., & Ess, C. M. (2020). *Internet Research: Ethical Guidelines 3.0*. <https://aoir.org/reports/ethics3.pdf>.

franzke, a. s., Muis, I., & Schäfer, M. T. (2021). Data Ethics Decision Aid (DEDA): a dialogical framework for ethical inquiry of AI and data projects in the Netherlands. *Ethics and Information Technology*, 23(3), 551-567. <https://doi.org/10.1007/s10676-020-09577-5>.

Geise, S., & Waldherr, A. (2021). Computational communication science: Lessons from working group sessions with experts of an emerging research field. In U. Engel, A. Quan-Haase, S. X. Liu, Su & L. Lyberg (Hrsg.), *Handbook of computational social science*, Bd. 1 (S. 66–82). London: Routledge.

Heinrichs, B. (2010). Medizinische Forschung am Menschen. In M. Fuchs, T. Heinemann, B. Heinrichs, D. Hübner, J. Kipper, K. Rottländer, T. Runkel, T. M. Spranger, V. Vermeulen, & M. Völker-Albert (Hrsg.), *Forschungsethik: Eine Einführung* (S. 56–81). Stuttgart: Verlag J. B. Metzler.

Literatur

Hollingshead, W., Quan-Haase, A., & Chen, W. (2021). Ethics and privacy in computational social science. In U. Engel, A. Quan-Haase, S. X. Liu, & L. Lyberg (Hrsg.), *Handbook of computational social science*, Bd. 1 (S. 171–185). London: Routledge.

Kennedy, H., Oman, S., Taylor, M., Bates, J., & Steedman, R. (2020). *Public understanding and perceptions of data practices: A review of existing research*. University of Sheffield.

McKee, H., & Porter, J. (2008). The ethics of digital writing research: A rhetorical approach. *College Composition and Communication*, 59(4), 711–749.

Markham, A., & Buchanan, E. (2012). *Ethical decision-making and Internet research 2.0: Recommendations from the AoIR ethics working committee*. <http://aoir.org/reports/ethics2.pdf>.

McKee, H. A., & Porter, J. E. (2009). *The ethics of internet research: A rhetorical, casebased Process*. New York: Peter Lang.

Kuckartz, U., & Rädiker, S. (2020). *Fokussierte Interviewanalyse mit MAXQDA: Schritt für Schritt*. Wiesbaden: Springer.

Schlütz, D. & Möhring, W. (2016). Kommunikationswissenschaftliche Forschungsethik – Sonntagsworte, Selbstzweck, Notwendigkeit? *Medien & Kommunikationswissenschaft*, 64(4), 483-496.