

INTERDISZIPLINÄRES STREITGESPRÄCH - NUTZERKOMMENTARANALYSEN

AUS ETHISCH-RECHTLICHER PERSPEKTIVE

Annalena Brokering (University of Edinburgh, Law School, UK) und Svenja Guhr (TU Darmstadt)

annalena.brokering@law-school.de, svenja.guhr@tu-darmstadt.de

Masterthesis Brokering (2019)

Ethisch-rechtliche Analyse der Sekundärnutzung von Social Media Daten aus Perspektive von NutzerInnen & Unterstützung der Einbeziehung rechtlicher und ethischer Erwägungen in datenbasierte Forschung durch 'Regulation by Design'-Ansatz.

Masterthesis Guhr (2019)

Gemischt qualitativ-quantitative Analysen von Onlinenewsartikeln & Lesercommentaren zur franz. Präsidentschaftswahl 2017. Topic Modeling Analyse & Sentimentanalyse der Kommentare zur Detektion emotionaler Einstellung zu Wahlkampfthemen & Kandidaten.

Einwilligung

B: Wurden die KommentatorInnen über die Nutzung ihrer Kommentare in Kenntnis gesetzt und um Zustimmung gebeten?

G: Nein. Bei der automatisierten Extraktion von tausenden Nutzercommentaren konnte nicht jede KommentatorIn um Zustimmung gebeten werden. Wie hätte die Kontaktaufnahme erfolgen sollen? Bei Datenerhebung war noch unklar, welche Daten später verwendet würden.

B: Diese Argumentation ignoriert das informationelle Selbstbestimmungsrecht der NutzerInnen und das datenschutzrechtliche Erforderlichkeitsgebot.

Datenanalyse



Urheberrechte

B: Wurde berücksichtigt, dass bei der Weiterverwendung von originellen und kreativen Beiträgen das Urheberrecht der KommentatorInnen betroffen sein könnte?

G: Kreative Beiträge sind besonders für qualitative Nutzerdatenanalysen relevant und werden oft in Publikationen aufgeführt. Unter welchen Voraussetzungen greift Urheberrechtsschutz? Welche Konsequenzen sind damit verbunden?

B: Urheberrechtlicher Schutz von Nutzerbeiträgen ist jeweils im Einzelfall anhand der Originalität und des Umfangs festzustellen und z.B. durch das Einholen einer Datennutzungseinwilligung zu lösen.

Anonymität

B: Wird die Identität der KommentatorInnen durch Anonymisierung der Daten geschützt?

G: Viele OnlinenutzerInnen verwenden nicht ihre Klarnamen, sondern Pseudonyme - in Publikationen ersetzt durch LeserIn(+ n).

B: Die Verwendung von Pseudonymen schützt NutzerInnen nicht unbedingt vor Re-Identifizierung. Durch die Suchmaschineneingabe von Kommentarbeispielen könnten Nutzernamen wiedergefunden werden. Bei Verwendung desselben Pseudonyms auf mehreren Onlineplattformen ergäben sich weitere Rückschlüsse auf die NutzerInnen.

G: Aber sind es nicht gerade die NutzerInnen selbst, die Ihre Daten, d.h. Nutzernamen und -beiträge online veröffentlichen? - Der hohe, mit einer vollständigen Anonymisierung verbundene Aufwand sollte im Verhältnis zur Sensibilität des Themas bzw. der betroffenen Nutzergruppe stehen (z.B. Jugendschutz).

Output *S*Tagger

```
LeserIn 1      05/05/2017 - 08h01
Aucune théorie ne mentionne que Le Pen
avait peut-être tout simplement picolé
ce soir là. Son comportement était
vraiment celui d'une ivrogne,
agressive, grimaces, ricanements,
confusion mentale entre SFR et Alstom,
paroles mécaniques, propos sans queue
ni tête, surexcitation suivi d'un
effondrement en fin de débat un peu
comme un ballon de baudruche qui se
dégonfle. Ah les ravages de la vodka
russe...
```

```
LeserIn 2      05/05/2017 - 06h15
Que pouvait-on espérer de plus dans un
débat avec l'éTron National ? Au moins
une chose est certaine: EM, étant
donné son avance dans les sondages,
avait beaucoup à perdre contre Ramène
Le Pire, qui, par son comportement,
nous a ramené à l'époque où son père
était à la tête de FN, et n'a pas su
en profiter.
```

Anonymisierte Lesercommentare

Repräsentativität

B: Wie wird sichergestellt, dass Forschungsergebnisse inhaltlich zutreffend sind, wenn Daten losgelöst von ihrem ursprünglichen Kontext, der Social Media Kommunikation analysiert werden? Wie kann die Repräsentativität der Analyseergebnisse gewährleistet werden, obwohl nur ein begrenzter Personenkreis Onlineartikel kommentiert?

G: Repräsentativität von Forschungsdaten ist Utopie, weil jede Datenauswahl bereits eine Art Interpretation verkörpert bzgl. erwarteter Aussagekraft und Wichtigkeit der Daten. Wichtig ist, die Grenzen ihrer Aussagekraft offenzulegen sowie die Gefahr, die von unpräzisen und unvollständigen Forschungsergebnissen ausgeht, weil Onlinenutzercommentare nicht immer tatsächliche Aussagen widerspiegeln.

B: Könnten durch die Betrachtung einer Gruppe Rückschlüsse auf Unbeteiligte getroffen werden?

Interessen/Grenzen der Wissenschaften

- Fehlende Kenntnis von Grenzen des Datenschutz- und Urheberrechts
- Begrenzte finanzielle, technische und zeitliche Spielräume
- Hoher Aufwand bei Einwilligungseinholung für jeden Nutzerbeitrag, ohne Sicherheit ob Daten verwendbar
- Verwendbarkeit und Relevanz von Daten wird erst nach Datenerhebung ermittelt
- Rechtskonforme Forschung unattraktiv wegen Gefahr, dass Daten an Wert/Aussagekraft verlieren
- Designkonzepte, Methoden und technische Tools zur rechtskonformen Umsetzung oft unbekannt oder erschwert zugänglich

Interessen/NutzerInnenrechte

- Recht auf informationelle Selbstbestimmung, d.h. Kenntnis und Kontrolle über Verwendung persönlicher Daten
- Bei personenbezogenen Daten, d.h. wenn Bezugsperson identifizierbar, greift Datenschutzrecht, z.B. DSGVO
- Nutzung personenbezogener Daten bedarf eines Rechtgrundes, typw. der Einwilligung der Bezugsperson
- Nutzereinwilligungen über AGB von Onlineplattformen zu unspezifisch für Forschungszweck
- Recht auf Privatsphäre, d.h. Respektierung des 'privaten' Kommunikationsraums von sozialer Medien
- Gefahr von 'chilling effects', d.h. Einschränkung von Onlineverhalten
- Recht auf Vergessen/Korrektur

Lösungsvorschläge/Herangehensweisen

■ **Ermittlung relevanter ethisch-rechtlicher Erwägungen und deren Realisierbarkeit für jedes Forschungsprojekt, von Datenakquise bis Publikation.**

■ **Bereits in Konzeptionsphase Evaluierung betroffener NutzerInneninteressen und -rechte:**

→ 'Data Protection Impact Assessments' zur Ermittlung von Datenschutzrisiken; dabei z.B. Rückgriff auf Design Thinking Ansätze, wie Roadmaps, um gesamten Forschungsprozess von Konzeption bis Publikation/Verbreitung zu erfassen

■ **Datenschutzkonformes Forschungsdesign: technische Umsetzung mit Hilfe von Privacy by Design-Ansätzen**

→ z.B. anonymisiertes Data Scraping; Einsatz von massentauglichen Consent-Mechanismen

■ **Proaktive Einbindung von NutzerInnen in Forschungskonzeption, zur Ermittlung konkret betroffener Nutzerinteressen u. Ermöglichung der Teilhabe.**