

Big Social Data

Ass. jur. *Tristan Julian Tillmann*, Institut für Informations-, Telekommunikations- und Medienrecht (ITM), Westfälische Wilhelms-Universität Münster

1 Einleitung

„Liberæ sunt [...] nostræ cogitationes.“

Cicero, Rede für Milo, XXIX, 79.

Unsere Gedanken sind frei. Dem überwiegenden Teil der deutschen Bevölkerung kommt bei diesem Leitmotiv ein bekanntes deutsches Volkslied (Breuer 1920: 118) in den Sinn. Zuletzt sang Jan Böhmermann dieses bei einer Solidaritätslesung zugunsten des in der Türkei inhaftierten Journalisten Deniz Yücel im Schauspiel Frankfurt (Dangel 2017). Der Gedanke dahinter taucht in der Geschichte jedoch häufiger auf. Schon Marcus Tullius Cicero hielt im Jahre 52 v. Chr. eine Apologie für den des Mordes angeklagten Titus Annius Milo, in der er das obige Zitat ausspricht. Der Wert dieser Aussage war und ist den freien, denkenden Menschen seit jeher bewusst. Es ist ein nicht zu unterschätzendes, hohes Gut.

Mit der stattfindenden Digitalisierung und dem verstärkten Einfluss von Big Data stellt sich die Frage, ob dieser Glaube an die Freiheit der Gedanken noch Gültigkeit hat. Auf den ersten Blick mag es frevelhaft erscheinen, eine solche fundamentale Säule einer freien Gesellschaft anzutasten. Denn der Bevölkerung ist es mittlerweile durchaus bewusst, dass den Betreibern von sozialen Netzwerken und anderen Plattformen bis hin zu zukünftigen Arbeitgebern eine Menge Daten, zum Teil auch persönlichster Art, bekannt sein können. Insbesondere jüngere Menschen „teilen“ online eine Vielzahl von persönlichen Informationen. Neuste Studien zeigen, dass auf Big Data gestützte Techniken verblüffende Möglichkeiten bieten. Bei diesen geht es nicht um die Sammlung von freiwillig öffentlich oder „privat“ geteilten Informationen, sondern um die Generierung völlig neuer Daten. So können unter Umständen Informationen ermittelt werden, denen sich der Einzelne über seine eigene Persönlichkeit überhaupt nicht bewusst ist oder die er mit absolut niemandem zu teilen bereit ist. So beginnt ein Szenario, welches sich für die wohl meisten Bürger der freien Welt als dystopisch darstellt.

Abstract

Immer mehr Menschen nutzen soziale Netzwerke. Das bekannteste und meist genutzte ist Facebook. Facebook hat aktuell mehr als zwei Milliarden Nutzer, die jeden Tag unzählige Daten produzieren. Neuste Studien zeigen, welche Informationen sich durch Big Data-Anwendungen aus diesen Daten generieren lassen. Dies birgt ein nicht zu unterschätzendes Risiko für jeden einzelnen Nutzer. Es entstehen jedoch auch vielfältige Möglichkeiten für Firmen und staatliche Stellen adäquater und schneller auf Veränderungen und andere Anforderungen der Bevölkerung zu reagieren. Der vorliegende Beitrag zeigt auf, welchen tatsächlichen Informationsgehalt Facebook-Profile enthalten.

Gerade Facebook stellt ein Paradebeispiel für einen riesigen Datensammler dar. Über jeden Nutzer und auch über Nicht-Nutzer sind eine Fülle von Informationen hinterlegt. Längst werden Bekanntschaften aus dem realen Leben „gestalked“ um vermeintlich zu wissen mit wem man es zu tun hat. Wissenschaftler hingegen sehen in Facebook eine Ansammlung von Probanden, deren Daten für sozialwissenschaftliche und psychologische Studien verwendet werden können. Ein Grund dafür ist, neben den verästelten sozialen und semantischen Daten, dass diese ohne allzu großen Aufwand ausgelesen werden können (Lambiotte & Kosinski 2014: 1935).

BIG SOCIAL DATA

2 Social Data vor dem Einsatz von Big Data-Technologien

Der Umfang an verfügbaren Daten über jeden einzelnen Internetnutzer ist schon seit langem enorm. Insbesondere Nutzer von sozialen Netzwerken wie Facebook, Instagram & Co. sind für die dahinterstehenden Firmen mehr oder weniger gläsern. Aber auch wer sich von sozialen Netzwerken fernhält hinterlässt eine deutliche Spur, die beispielsweise von Webtrackern (Röttgen 2016) oder direkt beim Internetdiensteanbieter, aufgezeichnet wird. Des Weiteren werden „offline“-Aktivitäten, beispielweise durch das Smartphone (z.B. durch Location-based Services (Pieringer 2012: 560, 563)) oder durch eine Vielzahl von anderen Möglichkeiten digital registriert und überwacht (Kosinski et al. 2013: 5802; Lambiotte & Kosinski 2014: 1937). Ein Auslesen und Analysieren der so gesammelten Daten ist mitunter sehr einfach und wird auch praktiziert. Es scheint ein vollumfänglich privates Leben im 21. Jahrhundert nicht mehr zu geben (Nussbaum 2007; Solberg 2010: 327).

Bisher ist es so, dass diese Daten lediglich eine Sammlung von Informationen aus der Vergangenheit und Gegenwart darstellen. Das so entstehende Bild ist kongruent mit dem Wissen der Person über sich selbst (Hermstrüwer 2016: 94). Auch bei Zugrundelegung dieser Prämisse gilt es vorsichtig mit persönlichen Daten umzugehen. Denn die gesammelten und von Internetnutzern öffentlich bereitgestellten Daten erlauben es Individuen online zu durchleuchten. Es entstehen umfassende Nutzerprofile, die vom Einzelnen nicht mehr kontrolliert und überblickt werden können (Dienlin 2015: 118). Privatpersonen und Unternehmen nutzen diese in großem Umfang zum Beispiel zur Optimierung ihres Angebots oder um gezielt Werbung zu adressieren. Dennoch bewerten insbesondere jüngere Teile der Bevölkerung das bestehende Risiko anders als es Nicht-„Digital Native“-Generationen tun. Vor allem auf Social Media Plattformen stellen sie sich, zum Teil hemmungslos, dar (Nussbaum 2007).

3 Big Data unterstützte Generierung von Informationen

Big Data-Technologien ermöglichen nun ganz neue Wege. Big Data umfasst eine Datenansammlung, die so groß, komplex, schnell und schwach strukturiert ist, dass sie mit normalen Datenbankmethoden nicht mehr verarbeitet werden kann (Roßnagel 2013: 562; Simo 2015: 14). Diese Datenmassen können verschiedenen Analyse-Algorithmen ausgesetzt werden um Muster und Zusammenhänge zu ermitteln (Roßnagel 2013: 562; Hermstrüwer 2016: 94/95; Wolf 2015: 25). Es geht somit nicht nur um das bloße Abrufen von gesammelten Informationen (Hermstrüwer 2016: 94), sondern um das Generieren von neuen Daten. Google soll durch Beobachtung des Suchverhaltens so zum Beispiel in der Lage sein den Beginn einer Grippewelle bestimmen zu können (Weichert 2013: 254; Flohr 2013). Den einzelnen Nutzer betreffend können Informationen ermittelt werden, die dieser womöglich selbst nicht kennt oder die er bisher niemals mit einem anderen Menschen geteilt hat (Hermstrüwer 2016: 94/95, 366). Es besteht sogar die Möglichkeit Entscheidungen von Personen und Personengruppen vorherzusagen (Hermstrüwer 2016: 94/95; Nebel 2015: 89; Roßnagel 2013: 562).

1. „Schattenprofile“

Über die vom User in sein Profil eingepflegten Daten hinaus erstellt Facebook ein sogenanntes „Schattenprofil“. Dieses speist sich aus einer Vielzahl von Informationen, die der Betroffene dem sozialen Netzwerk nie bereitgestellt hat. Jedes Mal, wenn die Kontaktdaten eines Smartphones Facebook zur Verfügung gestellt werden, versucht Facebook Querverbindungen zu finden (Hill 2017). Tun dies auch nur wenige „Freunde“ hat Facebook Zugriff auf eine Unmenge von Informationen, die der einzelne User Facebook gegebenenfalls nicht überlassen wollte (z.B. verschiedene E-Mail-Adressen, Kontaktfotos, Handynummer(n)). Somit kennt Facebook wahrscheinlich jede Adresse, an der ein bestimmter User jemals gelebt hat, jede E-Mail-Adresse und Telefonnummer die er jemals besessen hat und vieles mehr (Hill 2017). Diese „Schattenprofile“ erlauben es Facebook die sozialen Interaktionen genauestens zu verfolgen. Facebook kann darüber hinaus jedoch auch ermitteln, welche Personen sich wahrscheinlich kennen und diesen „Freunde“ vorschlagen. Besonders brisant wird

BIG SOCIAL DATA

dies, wenn Facebook einem User, der eine vermeintlich anonyme Samenspende abgegeben hat, Jahre später seine biologische Tochter als „Freund“ vorschlägt (Hill 2017).

2. Profilbildanalyse

Täglich laden die User des populärsten (Constine 2017) sozialen Netzwerks Facebook eine Vielzahl von Bildern hoch. Das „wichtigste“ ist natürlich das Profilbild. Ist der User auf diesem Profilbild zu sehen, kann es mit Hilfe eines Algorithmus analysiert werden. Mit Hilfe von künstlichen neuronalen Netzwerken werden riesige Datenmengen, wie zum Beispiel Bilddateien, analysiert und so Muster ermittelt (Kosinski & Wang 2017: 5). Diese übertreffen schon heute menschliche Fähigkeiten, beispielsweise bei der Erkennung von Hauptkrebs (Esteva et al. 2017: 117). Einer neusten Studie zufolge, ist es jedoch auch möglich, die sexuelle Orientierung des Users zu ermitteln. Dabei wurde zunächst ein künstliches neuronales Netzwerk an Bildern von hetero- und homosexuellen Menschen trainiert, damit es Muster ermitteln konnte. Sodann wurden diesem zwei Bilder vorgelegt, bei denen es zu entscheiden hatte, welche Person hetero- und welche homosexuell ist (Kosinski & Wang 2017: 16). Die Erfolgsquote lag, bezogen auf Facebook-Bilder, bei 74% (Kosinski & Wang 2017: 29). Im Gegensatz dazu liegen menschliche Einschätzungen lediglich bei einer Trefferquote zwischen 55 und 65% (Rule et al. 2009: 1246/1247; Kosinski & Wang 2017: 27). Die Risiken die hinter dieser Technik stecken, sind nicht zu unterschätzen. Der digitale Fußabdruck, den die Nutzer von sozialen Netzwerken hinterlassen, kann zwar minimiert werden aber das eigene Gesicht lässt sich nicht einfach verändern. Auf Grund der Omnipräsenz von Smartphones lassen sich von Personen ohne große Hindernisse Bilder erstellen und ggf. analysieren. Besondere Brisanz besteht in Ländern, in denen Homosexualität nicht nur verboten, sondern teilweise sogar mit der Todesstrafe bedroht wird (Bearak & Cameron 2016).

3. Aussagekraft von „Likes“ und der Sprache

Auf Facebook teilen die User eine Vielzahl von mitunter sensiblen Daten. Facebook erlaubt den Usern aber vor allem mit Hilfe eines „Likes“ ihre Zustimmung einfach und vermeintlich risikolos auszudrücken. Tatsächlich zeigen Studien, dass diese „Likes“ mit Hilfe von Big Data-Technologien enormes Aussagepotential enthalten.

Wissenschaftler sind in der Lage einige hoch sensible Eigenschaften der User zu bestimmen. Dazu gehören die sexuelle Orientierung, die Ethnie, die religiösen und politischen Ansichten, Alter, Geschlecht, Charakterzüge, Zufriedenheit und ob die Eltern getrennt leben (Kosinski et al. 2013: 5802). Diese Bestimmung erfolgt automatisiert und mit einer hohen Treffsicherheit. Nur über die „Likes“ ist es mit einer Wahrscheinlichkeit von 93% möglich das Geschlecht, von 85% die politische Ansicht („Democrat“ oder „Republican“) und von 88% zu bestimmen, ob ein männlicher User homosexuell ist (Kosinski et al. 2013: 5803). Die Wissenschaftler haben Verbindungen zwischen der zu bestimmenden Eigenschaft und dem „gelikedten“ Inhalt erstellt. So ist zum Beispiel ein guter Indikator für hohe Intelligenz das „liken“ von Seiten wie „Science“ oder „Thunderstorms“, wohingegen eine geringe Intelligenz mit „Harley Davidson“ oder „I Love Being A Mom“ korreliert. Besonders brisant ist, dass zum Teil bereits ein einziger „Like“ ausreicht um eine nicht zu vernachlässigende Aussage zu treffen. Wenn man bedenkt, dass der gemeine User deutlich mehr „Likes“ verteilt, steigt die Aussagefähigkeit deutlich an (Kosinski et al. 2013: 5804). Zudem ist zu berücksichtigen, dass statt „Likes“ auch andere digitale Aufzeichnungen in dieser Weise verwendet werden können. Beispielhaft sind der Browser-Verlauf und Sucheinträge bei Suchmaschinen zu nennen (Kosinski et al. 2013: 5805).

Nimmt man nun noch das klassische Hilfsmittel des Big Five-Modells (auch OCEAN-Modell) zur Hand, lassen sich aus den Facebook-„Likes“ noch weitere Informationen erlangen. Das Big Five-Modell ist international als Standardmodell der Persönlichkeitserforschung anerkannt (Lambiotte & Kosinski 2014: 1934; De Raad & Perugini 2002: 3, 5). Demnach bestehen fünf Hauptdimensionen der Persönlichkeit: Offenheit, Gewissenhaftigkeit, Extraversion, Verträglichkeit und Neurotizismus (Collins et al. 2015: 25; De Raad & Perugini 2002: 6). Je nachdem wie ausgeprägt die einzelnen Faktoren sind, lassen sich Aussagen über die Persönlichkeit eines Menschen machen. Diese können dazu verwendet werden, um zu bestimmen, wie sich ein Mensch tendenziell verhalten wird. Beispielsweise konnte ein Zusammenhang zwischen Extraversion und dem unentschuldigsten Fernbleiben von Arbeitnehmern nachgewiesen werden (De Raad & Perugini 2002: 6).

Die Wissenschaft ist durch die Kombination von „Likes“ und dem Big Five-Modell in der Lage, zu bestimmen wie

eine Testperson bestimmte Fragen beantworten wird (Grassegger & Krogerus 2016). Eine Computer gestützte Persönlichkeitseinschätzung, die auf Facebook-„Likes“ basiert, ist in der Lage sogar die Selbsteinschätzung des Users zu übertreffen (Youyou et al. 2015: 1036). Es werden lediglich 10, 70, 150 bzw. 300 „Likes“ benötigt, um einen durchschnittlichen Arbeitskollegen, einen Freund, ein Familienmitglied und den Partner in den Schatten zu stellen (Youyou et al. 2015: 1037). Die zugrunde liegenden Studien zeigen, dass soziale Netzwerke einen ausreichenden Umfang an Informationen enthalten, um akkurat die Persönlichkeit eines Menschen zu bestimmen (Kern et al. 2014: 166). Die Universität Cambridge betreibt eine Web-Anwendung, die die Persönlichkeit bestimmt. Die Anwendung analysiert, je nachdem welcher digitale Fingerabdruck zur Verfügung gestellt wird, die Facebook „Likes“ oder Twitter Posts (<https://applymagicsauce.com/>). Eine Anwendung die ebenso verführerisch wie angsteinflößend ist.

Doch Facebook enthält neben den „Likes“ einige weitere Informationsquellen (Lambiotte & Kosinski 2014: 1935). Ebenfalls mit Hilfe des Big Five-Modells konnten Wissenschaftler die allgemeine Zufriedenheit mit dem Leben der User sicher bestimmen. Dazu verwendeten sie neben den „Likes“ auch die Anzahl an „Freunden“, die Anzahl von „Tags“ auf Fotos, das Alter, den Beziehungsstatus und die in Posts verwendeten Wörter (Collins et al. 2015: 31). Gerade letztere bieten eine große Angriffsfläche. Persönliche Charakteristiken werden auch durch unsere Sprache, das heißt durch verwendete Wörter und Stilistik, ausgedrückt (Kern et al. 2014: 158). Eine Analyse von Facebook Posts von mehr als 69.000 Facebook Usern hat ergeben, dass zum Beispiel „fucking“ oder „sick_of“ auf eine niedrige emotionale Stabilität hinweist und „lakersbasketball“ und „beach“ eher auf eine hohe (Kern et al. 2014: 162). Jeder Post bei Facebook hinterlässt eine Spur zur eigenen Persönlichkeit (Kern et al. 2014: 166).

Aus diesem Wissen kann Macht generiert werden. Gerade Firmen dürften an diesen Möglichkeiten größtes Interesse haben. Zunächst könnte beispielweise ermittelt werden, welche Bedürfnisse entstehen werden (BITKOM 2012: 36/37; Columbus 2016). Kunden und mögliche Neukunden könnten einem „Targeting“ ausgesetzt werden, womöglich bevor diese selbst wissen, dass

ein Bedürfnis nach dem angebotenen Produkt besteht (Lambiotte & Kosinski 2014: 1934). Maßgeschneiderte Nachrichten, die automatisch, genau und, da personalisiert, kostengünstig versendet werden, sind keine Zukunftsmusik mehr. Zudem kann allgemein das Kundenverhalten studiert werden um strategische Entscheidungen besser treffen zu können (Stoicescu 2015: 29). Ein psychologisches Targeting, bei dem lediglich ein einziger „Like“ zur Bestimmung des psychologischen Profils zugrunde gelegt wird, kann die Clickraten von Facebook-Anzeigen um über 60% steigern (Matz et al. 2017: 3). Wenn man die Möglichkeiten bedenkt, die entstehen, wenn man ein vollständiges psychologisches Profil zu Grunde legt, kann einem Angst und Bange werden. Diese Technik könnte dazu verwendet werden um Schwächen auszunutzen und z.B. direkt Glücksspiel- oder Alkoholwerbung an entsprechend anfällige Personen zu adressieren (Matz et al. 2017: 4). Auch Human Resource-Abteilungen könnten zukünftige Arbeitnehmer vollumfänglich auf eine Kompatibilität mit dem Job und der Firma überprüfen (Youyou et al. 2015: 1039). Diese Modelle und Methoden bringen die Gefahr der Beeinflussung und Manipulation mit sich.

4 Fazit

Die „Freiheit der Gedanken“ ist nicht vollständig in Gefahr. Allerdings gibt es im Zeitalter von Big Data die Möglichkeit, schnell und automatisiert viele Informationen über einzelne Personen zu erlangen. Und wie gezeigt geht die Analyse schon viel weiter. Die psychologischen Eigenschaften und Einstellungen können analysiert und bewertet werden, ohne dass der Einzelne etwas davon mitbekommt (Lambiotte & Kosinski 2014: 1938). So besteht durchaus die Möglichkeit, dass in Zukunft beispielsweise Human Resource-Abteilungen auf solche und ähnliche Techniken zurückgreifen um zukünftige Mitarbeiter zu screenen. Anders als bisher könnte es nicht nur um ein Sammeln von „status quo“-Informationen gehen, also solchen, die der Einzelne „freiwillig“ zum Beispiel über soziale Netzwerke preisgegeben hat. Möglich könnten Prognosen über Entwicklungspotenziale und zusätzliche Informationen sein, die der Betroffene gar nicht preisgeben bereit und unter Umständen bisher niemals preisgegeben hat. So kann die Gefahr von Diskriminierungen und Entscheidungen, die auf nicht richtigen Fakten beruhen, entstehen (Her-

mstrüwer 2016: 97; Nebel 2015: 97). Denn Big Data-Anwendungen sind nicht unfehlbar.

Die neuen Möglichkeiten stellen einen enormen Verlust an Privatsphäre dar. Die User können nicht überblicken, welche preisgegebene Information zu welchem Wissensstand führt und wer diesen generiert bzw. nutzt. Die Rekombination- und Aggregationsmöglichkeiten mit anderen Daten sind unendlich und mit den Informationen wird auch in Zukunft immer weiter gearbeitet. Eine informierte Einwilligung in die Sammlung und Verarbeitung von Daten ist vor diesem Hintergrund schlicht nicht möglich. Jeder einzelne muss sich nicht nur fragen, ob er diese konkrete Information öffentlich machen möchte, sondern auch welche Kreuzverbindungen möglicherweise entstehen können. Diese letzte Frage lässt sich im Grunde nicht beantworten, da die in den Studien eingesetzten Verfahren und Algorithmen viel zu komplex sind. Im Grunde kann demjenigen, der seine Daten nicht zum Nulltarif unters Volk bringen will, nur angeraten werden, so wenig Daten wie möglich öffentlich zu machen. Ob dies im 21. Jahrhundert ein realistischer Rat ist, sei dahingestellt.

Vertiefungshinweise: Literatur und Links

- **Collins, Sun, Kosinski, Stillwell & Markuzon (2015).** Are You Satisfied with Life?: Predicting Satisfaction with Life from Facebook, in: Social Computing, Behavioral-Cultural Modeling, and Prediction, Agarwal et. al. Cham 2015.
- **Kosinski, Stillwell & Graepel (2013).** Private traits and attributes are predictable from digital records of human behavior, PNAS 2013, 5802-5805.



ABIDA (Assessing Big Data) Über die Dossiers

Das Projekt ABIDA, gefördert vom Bundesministerium für Bildung und Forschung, lotet gesellschaftliche Chancen und Risiken der Erzeugung, Verknüpfung und Auswertung großer Datenmengen aus und entwirft Handlungsoptionen für Politik, Forschung und Entwicklung. Dabei nähert ABIDA sich dem Thema Big Data aus einer grundlegend interdisziplinären Perspektive. Mehr Informationen finden Sie auf www.abida.de.

In den ABIDA-Dossiers werden regelmäßig ausgewählte Big Data-Themen kurz und prägnant dargestellt, um dem Leser einen Überblick zu liefern und einen Einstieg in die Thematik zu ermöglichen. Weitere Dossiers sind verfügbar unter www.abida.de/content/dossiers.

BIG SOCIAL DATA

Literaturnachweise

- Bearak & Cameron (2016).** Here are the 10 countries where homosexuality may be punished by death. Online verfügbar unter: <https://www.washingtonpost.com/news/worldviews/wp/2016/06/13/here-are-the-10-countries-where-homosexuality-may-be-punished-by-death>.
- BITKOM (2012).** Big Data im Praxiseinsatz – Szenarien, Beispiele, Effekte.
- Breuer (1920).** Der Zupfgeigenhansl, 90. Auflage Leipzig 1920.
- Giebel (1986).** Rede für Milo, Stuttgart 1986.
- Collins, Sun, Kosinski, Stillwell & Markuzon (2015).** Are You Satisfied with Life?: Predicting Satisfaction with Life from Facebook, in: Social Computing, Behavioral-Cultural Modeling, and Prediction, Agarwal et. al. Cham 2015.
- Columbus (2016).** Ten Ways Big Data Is Revolutionizing Marketing And Sales. Online verfügbar unter: <https://www.forbes.com/sites/louiscolumbus/2016/05/09/ten-ways-big-data-is-revolutionizing-marketing-and-sales/#4f4873c121cf>.
- Constine (2017).** Facebook now has 2 billion monthly users and... responsibility. Online verfügbar unter: <https://techcrunch.com/2017/06/27/facebook-2-billion-users/>.
- Dangel (2017).** Böhmermann singt für inhaftierten Journalisten Yücel. Online verfügbar unter: <http://www.hessenschau.de/kultur/jan-boehmermann-singt-in-frankfurt-fuer-inhaftierten-journalisten-yuecel.freedeniz-lesung-schauspiel frankfurt-100.html>.
- De Raad & Perugini (2002).** Big Five Assessment, Seattle 2002.
- Dienlin (2015).** Ist die politische Meinung privat oder öffentlich? Der Blick der Medienpsychologie, in: Privatheit, Öffentlichkeit und demokratische Willensbildung in Zeiten von Big Data, Roßnagel Baden-Baden 2015.
- Esteva, Kuprel, Novoa, Ko, Swetter, Blau & Thrun (2017).** Dermatologist – level classification of skin cancer with deep neural networks, Nature 2017, 115-118.
- Flohr (2013).** Gestatten, Stefan Henß, Student und Hellseher. Online verfügbar unter: <http://www.spiegel.de/lebenundlernen/job/student-der-tu-darmstadt-will-mit-computer-die-zukunft-voraussagen-a-938593.html>.
- Grassegger & Krogerus (2016).** Ich habe nur gezeigt, dass es die Bombe gibt. Online verfügbar unter: <https://www.dasmagazin.ch/2016/12/03/ich-habe-nur-gezeigt-dass-es-die-bombe-gibt/>.
- Hermstrüwer (2016).** Informationelle Selbstgefährdung, Tübingen 2016.
- Hill (2017).** How Facebook Figures Out Everyone You've Ever Met. Online verfügbar unter: <https://gizmodo.com/how-facebook-figures-out-everyone-youve-ever-met-1819822691>.
- Kern, Eichstaedt, Schwartz, Dziurzynski, Ungar, Stillwell, Kosinski, Ramones & Seligman (2014).** The Online Social Self: An Open Vocabulary Approach to Personality, Assessment 2014, 158-169.
- Kosinski, Stillwell & Graepel (2013).** Private traits and attributes are predictable from digital records of human behavior, PNAS 2013, 5802-5805.
- Kosinski & Wang (2017).** Deep neural networks can detect sexual orientation from faces, wird erscheinen im: Journal of Personality and Social Psychology. Online verfügbar unter: <https://osf.io//zn79k/>.
- Kosinski, Wang, Lakkaraju & Leskovec (2016).** Mining Big Data to Extract Patterns and Predict Real-Life Outcomes, Psychological Methods 2016, 493-506.
- Lambiotte & Kosinski (2014).** Tracking the Digital Footprints of Personality, Proceedings of the IEEE 2014, 1934-1939.
- Matz, Kosinski, Nave & Stillwell (2017).** Psychological targeting as an effective approach to digital mass persuasion, PNAS Early Print. Online verfügbar unter: <http://www.pnas.org/content/early/2017/11/07/1710966114.full>.
- Nebel (2015).** Facebook knows your vote! – Big Data und der Schutz politischer Meinung in sozialen Netzwerken, in: Privatheit, Öffentlichkeit und demokratische Willensbildung in Zeiten von Big Data, Roßnagel Baden-Baden 2015.
- Nussbaum (2007).** Say Everything, New York Magazine Februar 2007. Online verfügbar unter: <http://nymag.com/news/features/27341/index2.html>.
- Pieringer (2012).** There's No App For That: Protecting Users From Mobile Service Providers And Developers Of Location-Based Applications, Journal of Law, Technology & Policy 2012, 559-577.

- Roßnagel (2013).** Big Data – Small Privacy? Konzeptionelle Herausforderungen für das Datenschutzrecht, ZD 2013, 562-568.
- Röttgen (2016).** Gefällt mir, gefällt mir nicht – Tracking im Internet. Online verfügbar unter: <http://www.abida.de/sites/default/files/07%20Webtracking.pdf>.
- Rule, Ambady & Hallett (2009).** Female sexual orientation is perceived accurately, rapidly, and automatically from the face and its features, Journal of Experimental Social Psychology 2009, 1245-1251.
- Simo (2015).** Big Data: Opportunities and Privacy Challenges, in: Privatheit, Öffentlichkeit und demokratische Willensbildung in Zeiten von Big Data, Roßnagel Baden-Baden 2015.
- Stoicescu (2015).** Big Data, the perfect instrument to study today's consumer behavior, Database Systems Journal 2015, Vol. VI, no. 3, 28-42.
- Solberg (2010).** Data Mining on Facebook: A Free Space For Researchers Or An IRB Nightmare?, Journal of Law, Technology & Policy 2010, 311-343.
- Weichert (2013).** Big Data und Datenschutz Chancen und Risiken einer neuen Form der Datenanalyse, ZD 2013, 251-260.
- Wolf (2015).** Big Data und Innere Sicherheit, Marburg 2015.
- Youyou, Kosinski & Stillwell (2015).** Computer-based personality judgments are more accurate than those made by humans, PNAS 2015, 1036-1040.