



katharina a. ZWEIG
tobias d. KRAFFT
anita KLINGEL
enno PARK

Sozioinformatik

Ein neuer Blick auf Informatik und Gesellschaft



Zusatzmaterial zum Buch unter:
plus.hanser-fachbuch.de

HANSER



Ihr Plus – digitale Zusatzinhalte!

Auf unserem Download-Portal finden Sie zu diesem Titel kostenloses Zusatzmaterial. Geben Sie dazu einfach diesen Code ein:

plus-w19p3-92haz

plus.hanser-fachbuch.de



Bleiben Sie auf dem Laufenden!

Unser **Computerbuch-Newsletter** informiert Sie monatlich über neue Bücher und Termine. Profitieren Sie auch von Gewinnspielen und exklusiven Leseproben. Gleich anmelden unter:

www.hanser-fachbuch.de/newsletter



Katharina A. Zweig

Tobias D. Krafft

Anita Klingel

Enno Park

Sozioinformatik

Ein neuer Blick auf Informatik und Gesellschaft

HANSER

Autoren:

Prof. Dr. Katharina A. Zweig, TU Kaiserslautern

Tobias D. Krafft, TU Kaiserslautern

Anita Klingel, Stuttgart

Enno Park, Berlin



Alle in diesem Buch enthaltenen Informationen wurden nach bestem Wissen zusammengestellt und mit Sorgfalt geprüft und getestet. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grund sind die im vorliegenden Buch enthaltenen Informationen mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Autor(en, Herausgeber) und Verlag übernehmen infolgedessen keine Verantwortung und werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Weise aus der Benutzung dieser Informationen – oder Teilen davon – entsteht.

Ebenso wenig übernehmen Autor(en, Herausgeber) und Verlag die Gewähr dafür, dass die beschriebenen Verfahren usw. frei von Schutzrechten Dritter sind. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Dieses Werk ist urheberrechtlich geschützt.

Alle Rechte, auch die der Übersetzung, des Nachdruckes und der Vervielfältigung des Buches, oder Teilen daraus, vorbehalten. Kein Teil des Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder ein anderes Verfahren) – auch nicht für Zwecke der Unterrichtsgestaltung – reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

© 2021 Carl Hanser Verlag München

Internet: www.hanser-fachbuch.de

Lektorat: Dipl.-Ing. Natalia Silakova-Herzberg

Herstellung: Anne Kurth

Covergestaltung: Max Kostopoulos

Coverkonzept: Marc Müller-Bremer, www.rebranding.de, München

Titelbild: © shutterstock.com/emojoez

Satz: Frederik Maximilian Stegner

Druck und Bindung: Eberl & Koesel, Altusried-Krugzell

Printed in Germany

Print-ISBN 978-3-446-45213-8

E-Book-ISBN 978-3-446-46803-0

Inhalt

Vorwort	XI
1 Warum Sozioinformatik – und warum jetzt?	1
1.1 Automatische Essaybewertung – Die Zukunft der Benotung von Prüfungsleistungen?	1
1.2 Zusammenfassung	5
Teil I Grundlegende Definitionen	7
2 Grundlagen aus anderen Disziplinen	9
2.1 Technikfolgenabschätzung	9
2.1.1 Definition und Ziele	10
2.1.2 Formen	11
2.1.3 Phasen und Methoden	13
2.1.4 Grenzen	15
2.1.5 Das Collingridge-Dilemma	18
2.2 Modelle für menschliche Entscheidungen	19
2.2.1 Das Modell des Homo Oeconomicus und die Spieltheorie	19
2.2.2 Maslowsche Bedürfnispyramide und weitere menschliche Nebenbedingungen in der Entscheidungsfindung	20
2.2.3 Verzerrungen, die die Sachgrundlage für die Entscheidung verändern ..	23
2.2.4 Verzerrungen, die unabhängig von Sachgründen die Entscheidung beeinflussen	26
2.2.5 Nudging	28
2.2.6 Systemische Ansätze	30
2.3 Aufbau von Studien in der Verhaltensökonomie	32
2.3.1 Randomisierte Studien in der Verhaltensökonomie	32
2.3.2 A/B-Testing	34

3	Technikfolgenabschätzung und Sozioinformatik	35
3.1	Grundlegende Thesen der Sozioinformatik.....	38
3.1.1	These 1: Die digitale Transformation erzeugt eine Vielzahl neuer, emer- genter Phänomene.....	38
3.1.2	These 2: Technikfolgenabschätzung rund um die Digitalisierung benö- tigt neue Methoden der Modellierung, Analyse und zur Vorhersage der von ihr ausgelösten, emergenten (Neben-)Wirkungen.	41
3.1.3	These 3: Die Technikfolgenabschätzung von sozioinformatischen Sys- temen muss in der Informatik verankert sein.....	41
3.2	Definition des Forschungsfeldes.....	46
3.3	Lehrziele des vorliegenden Buches.....	47
4	Sozioinformatische Systeme	49
4.1	Definition von „System“.....	50
4.2	Systembeschreibungen als Modelle.....	55
4.3	Komplexe Systeme.....	56
4.3.1	Zur Definition des Begriffs „komplex“.....	57
4.3.2	Komplexität in der Psychologie.....	57
4.3.3	Komplexität in der Informatik.....	60
4.3.4	Komplexität in der Physik.....	61
4.3.5	Überblick „Komplexe Systeme“.....	64
4.3.6	Exkurs: Verändert maschinelles Lernen die Lösbarkeit komplexer Pro- bleme?.....	65
4.4	Eine erste Definition soziotechnischer Systeme.....	67
4.4.1	Zur Historie des Begriffs eines soziotechnischen Systems.....	68
4.4.2	Definition soziotechnischer Systeme nach Kienle & Kunau.....	68
4.4.2.1	Technische Systeme nach Kienle & Kunau.....	68
4.4.2.2	Soziale Systeme nach Kienle & Kunau.....	69
4.4.2.3	Soziotechnische Systeme.....	70
4.5	Ein neuer Ansatz für die Definition von sozioinformatischen Systemen.....	71
4.5.1	Technikfolgenabschätzung und die Definition soziotechnischer Syste- me nach Kienle & Kunau.....	72
4.5.2	Definition von sozioinformatischen Systemen im Rahmen der Technik- folgenabschätzung.....	73
4.6	Perverse Anreizstrukturen.....	75
4.7	Kybernetik.....	78
Teil II	Ein Ansatz für die Technikfolgenabschätzung von sozioinformatischen Systemen	83
5	Phänomeninduzierte sozioinformatische Analyse	85

5.1	Phasen einer phänomeninduzierten sozioinformatischen Analyse	91
5.1.1	Phase 1: Beschreibung der technischen Grundlagen	92
5.1.2	Phase 2: Identifikation der relevanten sozialen Akteure	95
5.1.3	Phase 3: Erstellen von Wirkungsgefügen.....	98
5.1.4	Phase 4: Analyse des Wirkungsgefüges.....	108
5.1.5	Phase 5: Identifikation möglicher Gegenmaßnahmen.....	110
6	Technikinduzierte sozioinformatische Analyse	113
6.1	Phasen einer technikinduzierten sozioinformatischen Analyse	114
6.1.1	Phase 1: Beschreibung der technischen Grundlagen	115
6.1.2	Phase 2: Identifikation der relevanten sozialen Akteure	117
6.1.3	Phase 3: Erstellen des Wirkungsgefüges	118
6.1.4	Phase 4: Analyse des Wirkungsgefüges.....	119
6.1.5	Phase 5: „Zweite-Gedanken-Denken“	121
6.2	Verwendung eines sozioinformatischen Systems für einen neuen Zweck.....	125
Teil III Typische Muster von Technikfolgen in sozioinformatischen Systemen		129
7	Emergente Phänomene der Einführung digital berechneter Rankings	131
7.1	Ranking von Webseiten durch Googles PageRank	131
7.2	Generisches Wirkungsgefüge beim Einsatz von verhaltensbasierten Rankings ..	140
7.3	Ranking von wissenschaftlicher Leistung durch den h-Index	144
7.4	Ranking von Produkten in Onlineshops durch Nutzerbewertungen	150
7.5	Selbstverstärkende Effekte von Rankings	153
8	Emergente Phänomene der Einführung personalisierter Dienstleistungen	157
8.1	Personalisierung durch kollaboratives Filtern	157
8.2	Emergente Phänomene beim Einsatz von personalisierten Inhaltsempfehlungen	166
8.3	Personalisierung von Preisen im Onlinehandel	169
9	Emergente Phänomene der Aufmerksamkeitsökonomie	179
9.1	Aufmerksamkeitsökonomie	180
9.2	Aufmerksamkeitsheischende und aufmerksamkeitsbindende Techniken	181
9.2.1	Notifications und E-Mails als aufmerksamkeitsheischende Techniken ..	181
9.2.2	Aufmerksamkeitsbindung innerhalb von Apps und Webseiten	182
9.2.3	Manipulation von kuratierenden und sortierenden Empfehlungssystemen	183

9.3	Monetarisierungsmöglichkeiten von Aufmerksamkeit	184
9.3.1	Direkte Monetarisierung von Aufmerksamkeit	184
9.3.2	Indirekte Monetarisierung: Aufmerksamkeit als Ware	184
9.4	Generisches Wirkungsgefüge: Aufmerksamkeitsökonomie	185
9.5	Der Rabbit-Hole-Effekt bei YouTube	187
9.6	Aufmerksamkeitsheischende und -bindende Techniken in Spiele-Apps	191
9.6.1	Aufmerksamkeitsbindende Spielmechanik	192
9.6.2	Modellierung als Wirkungsgefüge	194
9.6.3	Fazit des dritten Buchteils	198
Teil IV Zukünftige Forschungsgebiete der Sozioinformatik		199
10	Prothesen und Implantate: Der Mensch als Cyborg	201
10.1	Das Cochlea-Implantat	203
10.1.1	Funktionsweise des Cochlea-Implantates	203
10.1.2	Soziales und medizinisches Modell von Gehörlosigkeit	205
10.1.3	Wann ein Cochlea-Implantat sinnvoll ist	205
10.1.4	Die Gehörlosenkultur	206
10.1.5	Das Cochlea-Implantat als soziotechnisches System	207
10.1.6	Das Cochlea-Implantat ist nur bedingt ein sozioinformatisches System	208
10.2	Die künstliche Bauchspeicheldrüse	209
10.3	Das Brain-Computer-Interface	212
10.4	Projektion 2035	215
11	Forschungsbedarfe zu Modellierung, Analyse und Steuerung sozioinformatischer Systeme	233
11.1	Forschungsfragen zur sozioinformatischen Analyse	233
11.1.1	Phase 1: Technische Grundlagen und Phase 2: Identifikation der relevanten sozialen Akteure und ihrer Motivationen	233
11.1.2	Phase 3: Erstellen des Wirkungsgefüges	234
11.1.3	Phase 4: Analyse des Wirkungsgefüges	235
11.1.4	Phase 5: Identifikation von Gegenmaßnahmen und „Zweite-Gedanken-Denken“	236
11.2	Forschungsfragen zu generischen Wirkungsgefügen	238
11.3	Forschungsfragen zur Ungleichbehandlung	238
Literatur		245
Definitionsverzeichnis		263
Stichwortverzeichnis		264

Auf plus.hanser-fachbuch.de finden Sie zu diesem Titel kostenlose digitale Zusatzinhalte, unter anderem:

- Links zu weiterführenden Informationen;
- die von den Autoren erstellten Abbildungen zur Verwendung unter CC-BY-Lizenz; sowie
- eine Liste aller im Buch aufgeführten Internetseiten, als anklickbarer Link.

Vorwort

Seit dem Wintersemester 2013/14 gibt es an der TU Kaiserslautern einen neuen Studiengang namens „**Sozioinformatik**“. Die Entscheidung, einen neuen Studiengang einzurichten, macht man sich nicht leicht – aber es zeichnete sich ab, dass es eine neue Art von Software gibt, deren Technikfolgen weit über das bisher Bekannte hinausgehen und für die es zu diesem Zeitpunkt wenig Beschreibungs- und Analysemöglichkeiten gab. Ein gutes Beispiel für solche Softwaresysteme mit komplexen Technikfolgen sind Suchmaschinen, die es erst ermöglichen, dass Menschen in Milliarden von Dokumenten im Internet das finden, was sie gerade benötigen. Zu Beginn versuchte Yahoo noch, durch Einrichtung eines manuell kuratierten Kataloges Ordnung in die von Tag zu Tag wachsende Zahl von Webseiten zu bringen. Damit wurde eine analoge Idee zur Kategorisierung von Büchern auf die neue digitale Wirklichkeit angewendet – und scheiterte. In der Informatik sagt man von solchen scheiternden Transfers, dass sie nicht „skalieren“. Erst **PageRank**, der Algorithmus hinter Googles Erfolg, schaffte es, Nutzern zu zeigen, was sie suchten – ohne eine Ordnung dahinter zu finden oder überhaupt erst zu suchen. Algorithmen wie der Google-Suchalgorithmus gehören zur großen Gruppe der Empfehlungsalgorithmen, die versuchen, Nutzern und Nutzerinnen Inhalte oder Produkte zu empfehlen, indem sie analysieren, wie andere Menschen bisher auf diese Inhalte reagiert haben. Die dabei gefundenen Muster werden zur Vorhersage genutzt, was Nutzer:innen gut gefallen könnte. Nach dieser Art sortieren sich der News Feed bei Facebook, die Timeline bei Twitter, die automatische Playlist bei YouTube und die Reihenfolge der Nachrichten im Nachrichtenfeed. Die Muster beeinflussen sogar, wer wem als möglicher neuer Kontakt empfohlen wird – und all das könnte, zusammen mit der begrenzten menschlichen Aufmerksamkeit sowie unserer Vorliebe für Themen, zu denen wir schon eine Meinung haben, dazu führen, dass wir manipulierbar werden.

Zu diesem Themenkomplex gab es in den letzten Jahren viele Diskussionen, erste Erklärungsversuche und wenig Fakten. Klar ist aber, dass die diskutierten Phänomene von „Filterblase“ über „Fake News“ zu „digitaler Wahlmanipulation“ nur verständlich sind, wenn man sowohl die sozialen Akteure als auch die Software an sich in den Blick nimmt. Ein typisches Beispiel für solche Interaktionen zwischen Mensch und Maschine sind die Vorschläge bei der Suchvervollständigung: Kaum fängt man an zu tippen, schon schlägt einem die Suchmaschine ein paar mögliche Suchanfragen vor. Dabei ist nicht völlig transparent, nach welchen Kriterien die möglichen Vervollständigungen gewählt werden – die Popularität der Anfrage in letzter Zeit spielt aber auf jeden Fall eine Rolle. Und so wird die Suchanfrage nach „Merkel ist“ am 18.5.2019 unter anderem vervollständigt zu „schuld am Bre-

xit“ und „politisch am Ende“. Und selbst wenn man eigentlich etwas anderes gesucht hat, macht diese Information etwas mit dem Suchenden: „Denken das wirklich so viele? Ist das wahr?“ Hier kommunizieren also Suchende vermittelt über den Algorithmus mit anderen Suchenden – wobei unklar ist, wie genau das algorithmische Ergebnis interpretiert werden darf. Wenn sich nun ein Suchender durch die Vervollständigung dazu verleiten lässt, etwas zu suchen, was er eigentlich gar nicht suchen wollte, wird der Algorithmus das als weitere Interessensbekundung zählen. Damit wird ein positiver Feedbackzyklus initiiert, den der Algorithmus unter Umständen selbst erzeugt hat.

Die Sozioinformatik versteht die Interaktionen zwischen Mensch und Software und die durch Software vermittelten Interaktionen zwischen Menschen als komplexes System. Der Begriff **komplexes System** ist in der statistischen Physik entstanden und führte in den 1980er-Jahren zu einiger Aktivität als **Complex Systems Science**. Diese „Komplexitätswissenschaft“ beschäftigt sich insbesondere mit sogenannten **emergenten Phänomenen**, das sind solche, die nur durch die Interaktion der Systemteile verständlich sind – genau wie im obigen Beispiel der Suchvervollständigung. Die in der Complex Systems Science entwickelten Konzepte und Modelle werden in Deutschland nur wenig diskutiert und werden nicht zur Erklärung von gesellschaftlichen Phänomenen verwendet. In der Informatik wird es aber nun völlig unverzichtbar, diese Perspektive einzunehmen. Es bedarf daher aus unserer Sicht einer neuen Generation von Büchern zum Thema „Informatik und Gesellschaft“, da durch das Internet plötzlich deutlich größere soziale Gruppen miteinander interagieren und es dadurch zu komplexen Phänomenen kommt. Die dabei entstehenden sozio-technischen Systeme – oder eben genauer: Die dabei entstehenden **sozioinformatischen Systeme** – müssen in ihrer Gesamtheit modelliert und analysiert werden, um ihr zukünftiges Verhalten vorherzusagen und – wo möglich – zu steuern. Dies ist insbesondere dort wichtig, wo Software von menschlichem Verhalten lernen soll.

Wir sind nicht die Ersten, die eine solche sozioinformatische Gesamtanalyse fordern. 2016 schrieben Kate Crawford und Ryan Calo in ihrem Artikel „There is a blind spot in AI Research“: „Alongside such efforts, designers and researchers from a range of disciplines need to conduct what we call social-systems analyses of AI. They need to assess the impact of technologies on their social, cultural and political settings“ [Crawford and Calo, 2016].

Warum ist das wirklich notwendig? Ein gutes Beispiel dafür ist die Reaktion von manchen Uber-Fahrern und -Fahrerinnen auf die dynamische, algorithmengesteuerte Preisgestaltung ihres Arbeitgebers: Sie verabreden sich beispielsweise dazu, sich gleichzeitig abzumelden – das verringert künstlich das Angebot und der Algorithmus erhöht den Fahrpreis. Das technische System – der implementierte Algorithmus und die Daten, die in ihn eingehen – setzt dabei einen Anreiz für die Fahrer:innen. Die Idee des Konzerns ist, dass damit in den Gegenden, wo gerade eine größere Zahl an Fahrer:innen benötigt wird, auch mehr von ihnen dazu motiviert werden, sich in ihr Auto zu setzen. Gleichzeitig setzt der Algorithmus aber eben auch den Anreiz dafür, das System auszutricksen und den Fahrpreis zu erhöhen, obwohl kein „echter“ Mehrbedarf da ist. Es handelt sich hier um einen sogenannten **perversen Anreiz**, der dazu führt, dass Menschen ihn für andere Zwecke nutzen als den gedachten. Man sieht schon an diesem kleinen Beispiel, dass ein Teil der heutigen algorithmischen Entscheidungssysteme das Zusammenleben im sozialen System beeinflussen. Für die Entwicklung und Evaluation dieser Software ist es deshalb auch notwendig, zu verstehen, wie Menschen auf Anreize reagieren, wie sie mit dem algorithmischen System umgehen – sowohl bei der Eingabe als auch bei der Interpretation seiner Ausgabe.

In diesem Buch wollen wir aufzeigen, wie eine sozioinformatische Perspektive dazu führen kann, die digitalen Phänomene besser zu modellieren und zu analysieren. Das Lehrbuch ist entstanden aus der Vorlesung „Einführung in die Sozioinformatik“, die Professorin Zweig seit Einrichtung des Studiengangs „Sozioinformatik“ im Wintersemester 2013/14 hält. Tobias D. Krafft schrieb 2015 die erste Forschungsarbeit zur Methodenentwicklung in der Sozioinformatik im Rahmen seiner Bachelorarbeit mit dem Titel: „Vorstellung eines sozioinformatischen Analyseansatzes zur Technikfolgenabschätzung in Anlehnung an Vesters Sensitivitätsmodell am Beispiel des Unternehmens ‚Uber‘ als sozio-technisches System“ [Krafft, 2015]. Basierend auf diesen Überlegungen haben wir die Methodik gemeinsam weiterentwickelt. Die Verankerung in der Technikfolgenabschätzung und die Zukunftsanalyse von Cyborgs haben sich in Diskussionen mit Enno Park ergeben, während das Bild an vielen Stellen gestrafft, ergänzt und abgerundet wurde durch Anita Klingel. Beide haben vor allen Dingen eine interdisziplinäre Perspektive hinzugefügt. Unschätzbar sind auch die Dienste von Frederik M. Stegner, der in mühsamer Kleinarbeit all unsere Beiträge vereinheitlicht und unzählige typografische Details korrigiert hat.

Das Buch behandelt den Stoff von ca. 10 Doppelstunden. Die Vorlesung selbst besteht aus Theorieteilen, die sich mit interaktiven, diskursiven Anteilen abwechseln. Diese Teile haben wir in Form von Übungsaufgaben im Buch abgebildet:



Im Buch finden sich immer wieder Textboxen, die eine oder mehrere Fragen zur Diskussion stellen. Diese Fragen entstammen unserem didaktischen Konzept, dass die Teilnehmer einer Vorlesung immer wieder Zeit benötigen, um sich selbst ein Urteil zu bilden oder ihr Gehirn auf die folgenden Informationen vorzubereiten. Es lohnt sich daher, die Fragen auch dann zu beantworten, wenn Sie das Buch für sich alleine lesen. Wir empfehlen in einer Vorlesung, die Diskussion in kleinen Gruppe von nebeneinander sitzenden Studierenden führen zu lassen. Je nach Umfang der Fragen lassen wir den Studierenden dafür 3–10 Minuten Zeit. Die Ergebnisse werden im Plenum zusammengetragen und an der Tafel dokumentiert. In jeder 90-minütigen Doppelstunde planen wir insgesamt etwa 30 Minuten Zeit für solche Diskussionen und die Sammlung der Ergebnisse ein. ■

Wir würden uns freuen, von Ihren Erfahrungen mit diesem didaktischen Konzept zu hören (zweig@cs.uni-kl.de).

Wir beginnen das Buch mit einem kleinen Beispiel, das aufweist, inwiefern es beim Einsatz von Software zu etwas kommen kann, das wir als **emergentes Phänomen** bezeichnen, nämlich solche Verhaltensweisen, die erst durch Interaktion verschiedener Personen und Hard- oder Software zustande kommen. Diese Phänomene nennen wir **sozioinformatische Phänomene**. Buchteil I dient dann vor allen Dingen der Begriffsklärung und stellt verschiedene Konzepte vor, deren Kenntnis im weiteren Verlauf des Buches notwendig ist. Buchteil II stellt zwei verschiedene Methoden der sozioinformatischen Analyse vor, die einmal der Technikfolgenanalyse und einmal der Technikfolgenabschätzung beim Einsatz von Hard- und Software dient. Buchteil III stellt eine Reihe von sozioinformatischen Phänomenen vor, die sich in verschiedenen Varianten darstellen. Daraus lassen sich generische Phänomene ableiten, deren Kenntnis dabei helfen kann, unerwünschte Technikfolgen möglichst schon während der Entwicklung von Software vorherzusehen und somit abzuschwächen oder abzuwenden. Im letzten Buchteil, Buchteil IV, schauen wir in die Zu-

kunft: Wie wird es uns ergehen, als zukünftige Cyborgs? Dieser Buchteil wird durch weitere Forschungsfragen im Bereich der Sozioinformatik ergänzt.



Nicht nur Technik wirkt sich unmittelbar auf Gesellschaft aus – auch Sprache hat diesen Effekt. Das lässt sich auch in Studien nachweisen [Stahlberg and Sczesny, 2001]. Daher haben wir uns in diesem Buch bewusst darum bemüht, Männer* und Frauen* gleichermaßen anzusprechen.

Alle Themen im Buch betreffen regelmäßig die verschiedensten Wissenschaftsfelder: Von der Soziologie über die Rechts-, Politik- und Wirtschaftswissenschaften bis hin zur Psychologie und Philosophie. Wir konnten diese nur streifen und hoffen daher, dass die vorgelegten Methoden und Überlegungen in allen beteiligten Disziplinen auf fruchtbaren Boden stoßen. Die von uns beschriebenen Beispiele zeigen, wie oft Entwickler:innen bei der Gestaltung von Hard- und Software Entscheidungen treffen, die besser geworden wären, wenn sie gemeinsam mit Vertretern und Vertreterinnen der oben genannten Wissenschaftsfelder und mit Bürgerinnen und Bürgern getroffen worden wären. Zu einer solchen gemeinsamen Gestaltung der digitalen Transformation möchten wir mit dem vorliegenden Buch beitragen.

Kaiserslautern, Stuttgart & Berlin, 10.03.2021

Katharina A. Zweig, Tobias D. Krafft, Anita Klingel und Enno Park

1

Warum Sozioinformatik – und warum jetzt?

Dass die Digitalisierung immer stärker in unsere Leben eingreift, ist schon kaum mehr als eine Platitude. Wenn es aber um die Frage geht, welche Auswirkungen bei dem Einsatz eines neuen Software- oder Hardwareprodukts zu erwarten sind, fällt es den meisten doch schwer, eine detaillierte Technikfolgenabschätzung abzugeben, die klar macht, wo und wie sehr die Digitalisierung in unseren Leben Veränderungen auslösen wird. Die letzten Jahre haben stattdessen gezeigt, dass die Nutzer:innen, Regierungen, die Gesellschaft als Ganzes immer wieder überrascht werden von positiven wie negativen Nebenwirkungen von, beispielsweise, neuartigen digitalen Geschäftsmodellen, den Möglichkeiten von 3D-Druck z. B. für die Herstellung günstiger Prothesen, aber auch illegaler Waffen oder den neuen Manipulationsmöglichkeiten in der virtuellen Welt. In diesem Buch geht es um die Frage danach, ob und wie eine solche Technikfolgenabschätzung strukturiert durchgeführt werden kann. Auf den ersten Blick scheint es kaum möglich zu sein, bei solch breiten Anwendungsszenarien der Digitalisierung überhaupt einen Blick in die Glaskugel werfen zu können. Wir präsentieren hier eine Methode, die zeigt, dass eine strukturierte Analyse der Anreize für menschliches Verhalten, die durch den Einsatz einer Software verändert werden, es ermöglicht, einige Reaktionen von Menschen auf diesen Einsatz zu modellieren, zu analysieren und teilweise auch vorherzusagen. Und dazu betrachten wir als erstes Beispiel Software, die insbesondere in den USA und anderen englischsprachigen Ländern verwendet werden, um Essays von Schülerinnen und Schülern oder Studierenden zu bewerten.

■ 1.1 Automatische Essaybewertung – die Zukunft objektiver und effizienter Benotung von Prüfungsleistungen?

Insbesondere im amerikanischen Bildungssystem spielen Essays als Prüfungsleistung eine große Rolle. Kaum ein Aufnahmetest an einer Hochschule kommt ohne sie aus, und auch in Studiengängen an deutschsprachigen Universitäten sind Aufsätze und Referate gängige Prüfungsformate. Das Problem daran: Solche Texte inhaltlich zu bewerten ist arbeitsintensiv und letzten Endes gibt es stets eine subjektive Komponente: Anders als beispielsweise Multiple-Choice-Tests gibt es bei Essays und akademischen Texten neben den faktischen Anteilen immer Komponenten, bei denen sich die Geschmäcker unterscheiden. Entspre-

chend müssen Anbieter solcher Prüfungen viel Zeit und Geld in qualifizierte Kräfte investieren, die Tausende von Essays korrigieren und bewerten. Insbesondere bei Tests wie dem *Graduate Record Examination* (GRE) oder dem *Test of English as a Foreign Language* (TOEFL), die in den USA häufig auch über Universitätszulassungen entscheiden, müssen die Noten zudem so objektiv und vergleichbar wie möglich vergeben werden, um Klagen zu vermeiden.

Ohne Frage hätte also eine automatische Bewertung von Essays, so sie denn möglich wäre, mindestens zwei Vorteile:

1. **Effizienz:** Sie wäre schneller und auf Dauer deutlich günstiger als menschliche Bewertungen.
2. **Fairness:** Sie würde alle Texte auf exakt dieselbe Art und Weise behandeln, wäre somit konsistent und würde daher – so die gängige Schlussfolgerung – niemanden bevorzugen.

Basierend auf dieser Argumentation ist es nachvollziehbar, dass der Bildungssektor seit Jahrzehnten darauf hofft, dass Algorithmen diese Arbeit eines Tages übernehmen können. Seit mehr als zwanzig Jahren werden immer wieder Patente angemeldet und wissenschaftliche Studien veröffentlicht, die genau dies versprechen (so bspw. [Burstein et al., 1998, Cahill et al., 2018]). Zusammengefasst wird die Diskussion um Potenzial und Grenzen solcher Systeme unter dem Schlagwort „Automated Essay Scoring“. Sie alle nutzen einen Bewertungsalgorithmus, der den digitalisierten Text auf Basis vorher festgelegter Kriterien überprüft und daraus eine Bewertung ableitet.



Übungsaufgabe 1:

Diskutieren Sie, welche Auswirkungen es haben könnte, wenn überall dort, wo Essays verlangt werden, diese von Computern bewertet würden. Welche Chancen und Vorteile würden sich ergeben? Sehen Sie Nachteile oder Risiken? Was könnte auf lange Sicht geschehen?

Um diese Fragen beantworten zu können, ist es sinnvoll, sich den Gesamtprozess anzusehen, in dem automatische Essaybewertungssysteme eingesetzt werden: Zunächst erhalten Lernende die Aufgabe, einen Essay zu einem vorgegebenen Thema mit einer vorgegebenen Wortzahl zu schreiben und einzureichen. Insbesondere in Tests wie den beiden oben genannten oder bei den Aufnahmeprozessen der großen Elite-Universitäten sind dabei die Aufgaben stark standardisiert und die erwartete Textstruktur klar vorgegeben. Bei einer Aufgabe, in der ein kontroverses Argument diskutiert werden sollen, würde eine Einleitung erwartet werden, Argumente für Pro und Contra, eine dialektische Synthese der Argumente und eine zusammenfassende Schlussfolgerung. Zudem sind meistens klare Grenzen für die minimale und maximale Wortzahl angegeben. Wird der Essay von menschlichen Gutachter:innen bewertet, dann sind dies in den meisten Fällen nicht die Lehrpersonen, sondern extra für die Bewertung geschultes Fachpersonal, das die Kandidat:innen nicht kennt: Die Gutachter:innen werden insbesondere angewiesen, neben der sprachlichen Korrektheit die Argumentationsstruktur und die Form des Textes zu bewerten. Fakten werden auf ihre inhaltliche Korrektheit dagegen nicht überprüft [Anson and Perelman, 2017]. Die starke Strukturierung der Aufgabe und des Essays kommt dabei auch der Bewertung durch eine

Software entgegen. Das Problem bei den automatische Essaybewertungssystemen ist aber, dass sie an die Grenzen dessen stoßen, wie Maschinen Informationen verarbeiten können. Wie alle Informatiker:innen wissen, müssen Informationen hierzu digitalisiert werden, also in Zahlenform erfassbar sein. Einige Aspekte eines Essays lassen sich dabei leichter digitalisieren als andere: So kann die Textlänge leicht als Anzahl der Zeichen oder Wörter bestimmt werden. Ob der Text allerdings auch überzeugend ist, ja sogar, ob er einfach nur faktisch korrekt ist, kann ein Computer nicht bewerten, auch nicht mit Methoden der sogenannten künstlichen Intelligenz.

Aber vielleicht muss der Computer das auch gar nicht können, solange er nur möglichst viele Texte mit möglichst genau derselben Note bewertet, wie es menschliche Gutachter:innen täten. Stattdessen nutzen daher automatische Essaybewertungssysteme leicht quantifizierbare Eigenschaften des Textes, sogenannte „features“, deren Werte mit den Bewertungen menschlicher Prüfer:innen korrelieren. Dazu gehört beispielsweise die Anzahl ungewöhnlicher Substantive oder die Länge einzelner Sätze. Beide korrelieren positiv mit der Höhe der von Menschen vergebenen Noten. Daneben suchen die Bewertungssysteme auch nach Stichwörtern, die anzeigen, dass hier Argumente gegeneinander abgewogen werden, wie „similarly“, „additionally“, oder „in contrast to this“. Ob diese Abwägung inhaltlich konsistent ist, kann aber nicht automatisch bewertet werden. Tatsächlich ist eine solch mechanische Bewertung immer noch der Stand der Dinge. Das folgende Zitat stammt aus einem Review von Januar 2020:

Another major characteristic of (automatic essay grading, AES) algorithms is that length matters. [...] Mark D. Shermis, a strong advocate of AES, reported that he has run the Gettysburg Address, what Gary Wills (1992) has called “the words that remade America”, through several early AES machines, and the 271 word document received only 2s and 3s (Bloom, 2012).

The focus on length is a central element of e-rater’s algorithm for calculating a holistic score. Development and organization are calculated simply by counting the number of discourse elements (the ETS term for paragraphs) in an essay and their average length (Attali & Burstein, 2006; Attali & Powers, 2008; Quinlan, Higgins, & Wolff, 2009). Criterion, the classroom adaptation of e-rater, flags any paragraph with fewer than four sentences. [Perelman, 2020]

Ein solcher Ansatz funktioniert als Annäherung nur so lange gut, wie die Schreibenden die Art und Weise, wie sie einen solchen Text verfassen, nicht an die neuen Bedingungen anpassen. An zwei Enden werden die Defizite dieser Quantifizierung menschlichen Schreibens deutlich: Zum Einen würde ein solches automatische Essaybewertungssystem Schreibende schlecht bewerten, die es schaffen, originelle Gedanken und Argumente in kurze und einfache Sätze zu verpacken – obwohl das eine von uns Menschen hochgeschätzte und seltene Fähigkeit ist. Das oben zitierte Beispiel der Ansprache von Gettysburg zeigt das Problem deutlich. Zum anderen können Prüflinge und Dritte, die die Kriterien des angewendeten Systems einmal durchschaut haben, dieses leicht ausspielen: So erschuf beispielsweise der MIT-Wissenschaftler Les Perelman einen Algorithmus namens *Basic Automatic BS Essay Language Generator* (BABEL)¹, der zwar die vom System erwarteten Überprüfungs-kriterien berücksichtigt, inhaltlich aber völlig sinnbefreite Texte generiert. Die Abkürzung BS im Akronym BABEL steht daher zweifelsfrei für *bullshit*, für „völligen Unsinn“.

¹ Der Generator ist hier online verfügbar: <https://babel-generator.herokuapp.com/>.

Die von BABEL generierten Texte erzielten anschließend beste Bewertungen in verschiedenen automatischen Essaybewertungssystemen, die heute flächendeckend in den USA verwendet werden. Obwohl diese Tests für die Aufnahme an begehrten Universitäten so ausschlaggebend sind, zeigen diese Ergebnisse klar, wie einfach es ist, die automatischen Essaybewertungssysteme auszutricksen. In seinem Forbes-Artikel „No, Software Still Can't Grade Student Essays“ bringt Peter Greene genau diese Veränderung der Anreizstruktur für Prüflinge auf den Punkt:

The ultimate argument about Perelman's work with BABEL is that his submissions are "bad faith writing." That may be, but the use of robo-scoring is bad faith assessment. What does it even mean to tell a student, "You must make a good faith attempt to communicate ideas and arguments to a piece of software that will not understand any of them." [Greene, 2020]

In der Folge entstand ein fast schon bizarres Wettrennen, indem die verantwortliche Organisation hinter dem GRE-Test, die *Educational Testing Services* (ETS), wiederum einen Algorithmus konzipierte, um BABEL-Texte zu erkennen [Perelman, 2020]. Ob diese Mechanismen aber auch die auf die Software zielenden Manipulationen menschlicher Kandidat:innen erkennen können, ist unklar.

Solche Reaktionen und Gegenreaktionen sind typisch für **sozioinformatische Systeme**, also solche Systeme, die aus einem oder mehreren sozialen Systemen bestehen, die über eine Software miteinander interagieren. Software wie die beschriebenen automatischen Essaybewertungssysteme verändert dabei die Anreizstrukturen der beteiligten Personen und Institutionen: Für die Institutionen werden die Gutachten auf lange Sicht günstiger, die Kandidat:innen bemerken aber gleichzeitig, dass ein paar einfache Tricks ihnen zu besseren Noten verhelfen können. Damit wird die Bewertung weniger aussagekräftig und unter Umständen müssen die Universitäten als Nutzer der Bewertung dann wiederum mehr Aufwand in die Bewertung von Bewerbungsunterlagen stecken, um eine ertrickste Bestleistung zu identifizieren. Dieser gesamtgesellschaftliche Effekt kann auf den ersten Blick als überraschend eingestuft werden.

Tatsächlich sind die Reaktionen und Gegenreaktionen des sozialen Systems auf die Einführung eines Softwaresystems **aus der Perspektive der durch diesen Einsatz veränderten Anreizstrukturen** weniger überraschend. Eine Definition des Begriffs der Anreizstruktur findet sich in Abschnitt 4.6, grob verstehen wir darunter die Menge der Regeln und Prozesse in einem sozialen oder sozioinformatischen System, die bestimmen, welches menschliche Verhalten zu welchem Ergebnis beiträgt. In diesem Fall bietet der Einsatz der Software für die Verwender:innen unter anderem den Anreiz, Kosten für menschliche Gutachtachter:innen einzusparen, und setzt gleichzeitig den Kandidat:innen Anreize, möglichst lange Texte zu schreiben – mit möglichst vielen selten genutzten Wörtern.

■ 1.2 Zusammenfassung

Dieses Beispiel illustriert anschaulich, was Gegenstand und Ziel der Sozioinformatik ist: Der Einsatz automatischer Essaybewertungssysteme verändert Anreizstrukturen für Lehrende und Lernende – diese Veränderung von Strukturen innerhalb eines gegebenen, sozioinformatischen Systems ist das Thema dieses Buches. Dabei funktioniert die angewendete Technik für einen spezifischen Teilbereich und unter festgelegten Bedingungen – außerhalb dieses Bereiches wird sie allerdings schnell dysfunktional und kann dann sogar gesamtgesellschaftliche Folgen haben. Das Beispiel zeigt auch, warum die Digitalisierung im Allgemeinen das Potenzial hat, so tief in unsere Leben einzugreifen: Software verändert durch eine Änderung der Anreizstrukturen oftmals die sozialen Strukturen der Systeme, in der sie eingesetzt wird. Diese Technikfolgen gilt es, möglichst früh in der Entwicklung und im Einsatz von Software zu erkennen und, wenn es sich um ungewollte Konsequenzen handelt, möglichst auch vorausschauend zu vermeiden. Das Ziel der Sozioinformatik liegt daher darin, die für die Technikfolgenabschätzung notwendigen Kausalketten zu modellieren, sie zu analysieren und dann dort, wo es möglich ist, eine fundierte Vorhersage zu machen, zu welchen Reaktionen und Gegenreaktionen es kommen könnte. In diesem Buch entwickeln wir hierfür eine strukturierte Methode der Technikfolgenabschätzung, in der psychologische, wirtschaftliche, informatische, ethische und soziologische Aspekte in eine Kausalketten- und Netzwerkanalyse eingebettet werden.

Dafür schaffen wir im folgenden Teil I die definitorischen und konzeptionellen Grundlagen, bevor wir uns in Teil II ausführlicher mit Zielen und Methoden der sozioinformatischen Analyse beschäftigen.

TEIL I

Grundlegende Definitionen

Die letzten Jahre haben gezeigt, dass durch die Interaktion von Mensch und digitalen Artefakten bereits bekannte soziale Phänomene wie beispielsweise die Verbreitung von Verschwörungstheorien beschleunigt und andere Phänomene, z. B. eine gemeinschaftlich geschriebene Enzyklopädie, überhaupt erst ermöglicht wurden. Ist es denkbar, die aus dieser Interaktion basierenden **Technikfolgen** abzuschätzen? In diesem Teil werden die dafür notwendigen definitorischen und konzeptionellen Grundlagen gelegt: Leitfragen sind dabei, was ein (sozioinformatisches) System ist, wie **Sozioinformatik** und **Technikfolgenabschätzung** zusammenhängen sowie welche psychologischen Erkenntnisse darüber hinaus hilfreich sein können.

2

Grundlagen aus anderen Disziplinen

Ziel der Sozioinformatik ist es, sozioinformatische Systeme möglichst früh auf potenziell aus ihr entstehende **emergente Phänomene** und deren Auswirkung hin zu analysieren. Um dieses Ziel zu erreichen, bedient sie sich einer Reihe von Ansätzen aus anderen Disziplinen. Wie wir in diesem Buch darlegen, ist ein vielversprechender Ansatz für eine solche Technikfolgenabschätzung eine Analyse der Verhaltensänderung von Personen, die durch die Software ausgelöst wird. Wir gehen dabei davon aus, dass der Einsatz der Software einige Verhaltensweisen belohnt und andere schwieriger macht oder bestraft. Ein Beispiel dafür ist eine digitale Kreditwürdigkeitsbewertung, die bestimmte Informationen über Bankkundinnen und Bankkunden positiv oder negativ berücksichtigt. Um die daraus erwachsenden Technikfolgen abzuschätzen, muss man daher verstehen, wie sich Menschen für eine bestimmte Handlung entscheiden.

Das folgende Kapitel gibt daher zum Einen einen kurzen Überblick über Geschichte und Methoden der Technikfolgenabschätzung, zum Anderen stellt es bisherige Modellansätze zu menschlichen Entscheidungsprozessen vor.

■ 2.1 Technikfolgenabschätzung

Wir beginnen mit einer Reihe von Fragen. Wie schon in der Einleitung erwähnt, lohnt es sich auch für einzelne Leser:innen, diese Fragen für sich selbst zu beantworten, falls sie keine Lerngruppe für eine Diskussion haben.



Übungsaufgabe 2:

Diskutieren Sie die folgenden Fragen oder beantworten Sie diese für sich selbst:

1. Wie würden Sie Technik definieren?
2. Welches ist die älteste verwendete Technik, die Ihnen einfällt, und aus welcher Epoche stammt diese?
3. Nennen Sie fünf Beispiele von Technik, die gesellschaftlich unerwünschte Folgen hatte.
4. Kennen Sie Beispiele von Software, deren Verwendung gesellschaftlich unerwünschte Folgen hatte?

5. Sehen Sie für diese Beispiele einen Teil der Verantwortung bei den Softwareentwickler:innen?
6. Wer ist an Softwareentwicklung noch alles beteiligt?

Im nächsten Abschnitt legen wir zuerst die im Buch verwendete Definition von *Technik* und *Technikfolgenabschätzung* fest.

2.1.1 Definition und Ziele

Es gibt keine allgemeingültige Festlegung des Begriffs der Technik oder der Technikfolgenabschätzung [Grunwald, 2010, S. 14]. Wie bei den meisten komplexen Konzepten gibt es unstrittige Beispiele: So ist ein Computer sicher ein Stück „Technik“, aber sind es die Programmiersprachen auch, die den Computer erst nutzbar machen?

Wir folgen in diesem Buch den Definitionen des *Verein Deutscher Ingenieure* (VDI) als Arbeitsgrundlage.



Definition 1: Technik

Technik [...] umfasst:

- [...] die Menge der nutzenorientierten, künstlichen, gegenständlichen Gebilde (Artefakte oder Sachsysteme),
- die Menge menschlicher Handlungen und Einrichtungen, in denen Sachsysteme entstehen, und
- die Menge menschlicher Handlungen, in denen Sachsysteme verwendet werden.

[VDI, 2000, S. 2]

Der erste Stein, den Urmenschen verwendet haben, um andere Steine zu bearbeiten, wäre demnach noch nicht Teil der Technik gewesen, weil er so benutzt worden ist, wie er natürlich aufgefunden wurde. Als Teil der „Technik“ würde aber der bearbeitete Stein gelten. Zudem würde die Definition von Technik auch die Prozesse umfassen, mit denen Menschen Steine als Werkzeuge benutzten, um Produkte herzustellen, und alle Handlungen, in denen Menschen die bearbeiteten Steine dann weiterverwendeten. Ebenso wäre der Computer als klar erkennbares, künstliches und gegenständliches Gebilde Teil der Technik. Die Programmiersprachen wiederum sind ein Teil der menschlichen Handlungen, in denen Computer verwendet werden, und damit ebenfalls Teil der Technik nach dieser Definition.

Wir verwenden im Folgenden den Begriff *Technikfolgenabschätzung* synonym zum Begriff der *Technikbewertung*, wie ihn der Verein deutscher Ingenieure definiert hat:



Definition 2: Technikfolgenabschätzung

Technikbewertung bedeutet hier das planmäßige, systematische, organisierte Vorgehen, das

- *den Stand einer Technik und ihre Entwicklungsmöglichkeiten analysiert,*
- *unmittelbare und mittelbare technische, wirtschaftliche, gesundheitliche, ökologische, humane, soziale und andere Folgen dieser Technik und möglicher Alternativen abschätzt,*
- *aufgrund definierter Ziele und Werte diese Folgen beurteilt oder auch weitere wünschenswerte Entwicklungen fordert, [und]*
- *Handlungs- und Gestaltungsmöglichkeiten daraus herleitet und ausarbeitet, [...].*

[VDI, 2000, S. 2f]

In dieser Definition sind bereits verschiedene Ziele und Funktionen einer Technikfolgenabschätzung (TA) enthalten: So orientiert sich die Technikfolgenabschätzung stets an **extern** vorgegebenen Zielen und Werten, denen die untersuchte Technik entsprechen soll. Die Analyse der Folgen selbst kann sich dabei auf ein breites Spektrum von Aspekten beziehen – entsprechend bezieht die Technikfolgenabschätzung auch ihre Methoden und Expertise aus einer ganzen Reihe wissenschaftlicher Disziplinen, die im Folgenden beschrieben werden.

2.1.2 Formen

Wie die eigentliche Technikfolgenabschätzung konkret ausgestaltet wird, hängt von vielen Faktoren ab. Drei wichtige Dimensionen, die wir hier diskutieren wollen, können dazu genutzt werden, die diversen Ansätze zu kategorisieren:

1. Was ist die **Fragestellung**?
2. Wer ist der **Adressat** der Analyse?
3. Welche **Akteure** werden in den Prozess einbezogen?

Mit Blick auf die **Fragestellung** lassen sich zwei Formen unterscheiden:

Die *probleminduzierte Vorgehensweise*: Ein gegebenes Problem soll mit technischen Mitteln gelöst werden. Ein Beispiel hierfür wäre die Energieversorgung: Mit welcher Technik lässt sich der Energiebedarf ökonomisch und ökologisch bei geringstmöglichen Schäden decken? Eng damit verzahnt ist das Beispiel des menschengemachten Klimawandels: Wenn soziale und ökonomische Veränderungen herbeigeführt werden müssen, um die Erderwärmung zu stoppen oder wenigstens abzuschwächen, stellt sich die Frage, in welchem Ausmaß neue und vorhandene Techniken dabei helfen könnten – und zu welchem Preis.

Die Informatik geht sehr oft, vielleicht sogar meistens probleminduziert vor: Tatsächlich sind viele algorithmische Fragestellungen als Probleme definiert, wie beispielsweise das

„Kürzeste-Wege-Problem“ [Garey and Johnson, 1979]. Bisher weniger fundamental für die Informatik war dagegen eine Modellierung und Analyse der von ihr induzierten Technikfolgen.

Eine solche *technikinduzierte Vorgehensweise* stellt die zweite Form der Technikfolgenabschätzung dar: Hier wird eine neu erfundene Technik auf ihre Folgen hin untersucht. Diese Herangehensweise wird vor allem für Technikfolgenabschätzungen in Bereichen wie der Kernkraft, Gentechnik und Nanotechnologie verwendet. Die technikinduzierte Technikfolgenabschätzung für informatische Artefakte ist ein wichtiger Teil dieses Buches, um Technikfolgen des Einsatzes von Algorithmen, Big Data, künstlicher Intelligenz und Machine Learning zu antizipieren oder wenigstens im Nachhinein erklärbar zu machen.

Mit Blick auf den **Adressaten** einer Technikfolgenabschätzung gibt es grundsätzlich drei unterschiedliche Formen:

Die häufigste Form der Technikfolgenabschätzung ist die der *parlamentarischen Technikfolgenabschätzung*. Hierbei sollen politische Entscheider:innen dazu befähigt werden, Chancen und Risiken einer bestimmten Technik zu bewerten und letztere entsprechend regulativ einzuhegen.

Bei der *ökonomischen Technikfolgenabschätzung* will eine Organisation – häufig diejenige, die die Technik entwickelt – deren Potenzial abschätzen. Hier liegt der Schwerpunkt oft eher auf wirtschaftlichen Kriterien wie Markt reife, Wachstumspotenzial oder einer Kosten-Nutzen-Analyse.

Die *wissenschaftliche Technikfolgenabschätzung* schließlich zielt darauf ab, Experten und Expertinnen zu einer adäquaten Einschätzung neuer Techniken zu befähigen, sie auf einen gemeinsamen Stand zu bringen oder den wissenschaftlichen Diskurs über eine bestimmte Technik voranzutreiben.

Solche Technikfolgenabschätzungen können auch eine gesellschaftliche Debatte anstoßen oder vorantreiben, was sich am Beispiel der aktuellen Diskussion um **Künstliche Intelligenz** in allen drei Formen gut beobachten lässt: Die parlamentarische Technikfolgenabschätzung wurde durch den Einsatz verschiedener Expertenbefragungen beispielsweise für die Erstellung des Strategiepapiers „Künstliche Intelligenz“ der Bundesregierung [Bundesregierung, 2018], des Einsatz der Datenethikkommission [DEK, 2019] und der Enquete-Kommission umgesetzt [Deutscher Bundestag, 2020a]. Eine ökonomische Technikfolgenabschätzung erfolgt im Rahmen der Forschung und Entwicklung bei an der Entwicklung beteiligten, großen Unternehmen wie Google [Google, 2020]. Wissenschaftliche Technikfolgenabschätzung schließlich erfolgt an Universitäten und Instituten wie dem Deutschen Forschungszentrum für Künstliche Intelligenz (DFKI). Alle diese Formen von Technikfolgenabschätzung treiben die gesamtgesellschaftliche Debatte an, welcher Einsatz von Künstlicher Intelligenz notwendig und geboten ist, und welcher Einsatz als gesellschaftlich nicht wünschenswert gilt.

Zuletzt können auch mit Blick auf die Einbindung unterschiedlicher **Akteure** verschiedene Formen der Technikfolgenabschätzung unterschieden werden (siehe Abbildung 2.1):

Am größten ist dieser Kreis bei der *diskursiven* oder auch *argumentativen Technikfolgenabschätzung*. Hier wird eine breite öffentliche Debatte unter Einbeziehung von z. B. Befürworter:innen, Kritiker:innen, Stakeholdern und Betroffenen analysiert. Geringer fällt die Zahl der Beteiligten bei der *partizipativen Technikfolgenabschätzung* aus. In diesem Fall werden lediglich Repräsentantinnen und Repräsentanten verschiedener Bevölke-

rungsgruppen hinzugezogen, deren Perspektiven als relevant für die zu untersuchende Fragestellung angesehen wird. Dies können Betroffenenverbände, Nicht-Regierungs-Organisationen (NGOs), Bürgerinitiativen und ähnliche Gruppen sein. Am kleinsten ist der Kreis bei einer *Experten-Technikfolgenabschätzung*: Hier äußern sich Wissenschaftler:innen und Fachleute in Form von Gutachten, Stellungnahmen und Interviews zu einer Fragestellung. Ist die Entwicklung einer technischen Neuerung fortgeschritten oder befindet sich die Technik bereits auf dem Markt, bietet sich eine *konstruktive Technikfolgenabschätzung* an. Dazu werden die Entwickler:innen, die konkret an der Umsetzung der zu untersuchenden Technik arbeiten, in den Kreis der Befragten aufgenommen. Diese Vorgehensweise eröffnet auch die Chance, dass die Ergebnisse der Technikfolgenabschätzung direkt in die weitere Entwicklung der Technik einfließen können [Grunwald, 2010, S. 99].

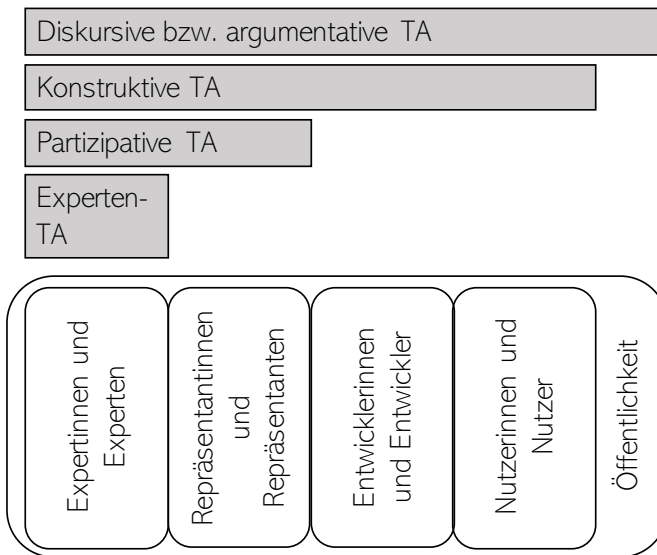


Abbildung 2.1 Teilnehmende Gruppen in den verschiedenen Technikfolgenabschätzungen

All diese Formen schließen sich keineswegs gegenseitig aus, sondern können vielmehr passend zur Fragestellung in den verschiedenen Phasen einer Technikfolgenabschätzung miteinander kombiniert werden.

2.1.3 Phasen und Methoden

Der Prozess der Technikfolgenabschätzung muss sich grundsätzlich flexibel an das Objekt seiner Untersuchung anpassen können. Dennoch haben sich in der Praxis bestimmte Phasen und Methoden herauskristallisiert, die im Folgenden vorgestellt werden.

Phasen Der Prozess der Technikfolgenabschätzung durchläuft nach Grunwald [Grunwald, 2010, S. 121] drei Phasen: Am Anfang stehen die Zieldefinition, die Beschreibung des konkreten Problems bzw. der zu untersuchenden Technik sowie methodische Festlegungen. An sie schließt sich die eigentliche Abschätzung an, die aus einer Beschreibung der

zu erwartenden Nebeneffekte, der von diesen betroffenen Gruppen oder Personen sowie einer Prognose der weiteren Entwicklung der Technik besteht. In der folgenden Bewertung sind auf Basis der getroffenen Abschätzung Chancen und Risiken der Technik zu ermitteln und zu bewerten, sowie Handlungsempfehlungen abzuleiten. In Abbildung 2.2 ist dieser Prozess visualisiert und die jeweils beteiligten Personengruppen aufgeführt.

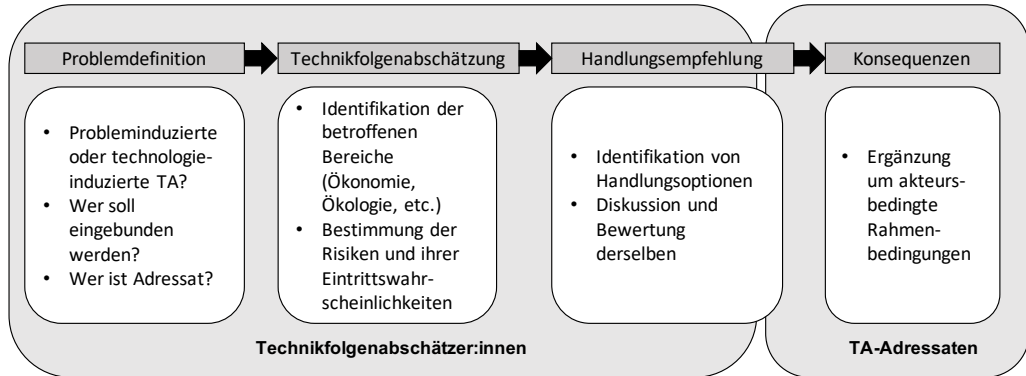


Abbildung 2.2 Phasen der Technikfolgenabschätzung mit jeweils verantwortlicher Personengruppe

Methoden In jeder dieser Phasen können unterschiedliche Methoden zum Einsatz kommen: Um *Prognosen* über die Folgen einer Technik aufstellen zu können, bedient sich die Technikfolgenabschätzung sowohl qualitativer als auch quantitativer prognostischer Verfahren. Bei der *Trendextrapolation* werden beispielsweise bestehende Entwicklungen in die Zukunft fortgeschrieben [Grunwald, 2010, S. 178]. Insbesondere identifiziert man dafür Messgrößen, die sogenannten *Systemvariablen*, und versucht deren kausale Abhängigkeiten voneinander in sogenannten *Wirkungsgefügen* zu beschreiben.



Definition 3: Wirkungsgefüge

Ein Wirkungsgefüge ist ein allgemeiner Begriff für einfache visuelle Modelle, die darstellen, wie bestimmte Messgrößen (die Systemvariablen) voneinander abhängen. Man kann sie nutzen, um abzuschätzen, welche Konsequenzen die Veränderung des Wertes einer bestimmten Variable haben könnte.

Wir folgen hier der informellen Definition von Kausalität, die Pearl, Glymour, and Jewell in ihrem Lehrbuch „Causal Inference in Statistics“ anbieten: „A variable X is a *cause* of a variable Y if Y in any way relies on X for its value“ [Pearl et al., 2016].¹

¹ Formaler definieren Pearl, Glymour und Jewell eine Menge \mathcal{F} von Funktionen, in der eine Funktion für jede Variable X enthalten ist. Die Funktion $f(X)$ der Variable X weist ihr ihren Wert in Abhängigkeit von den Werten anderer Variablen zu. Wenn der Wert von Variable Y in irgendeiner Form in die Funktion von Variable X einfließt, ist Y nach dieser Definition kausal für den Wert von X .