

75
NFP

Big Data
Nationales Forschungsprogramm

Big Data

Anwendungen,
Technologien und
gesellschaftliche
Aspekte

Résumé
des Nationalen Forschungsprogramms
«Big Data» (NFP 75)

Vorwort → 4

Executive Summary → 6

1. Einführung: Big Data, grosse Veränderungen → 9

1.1 Die zunehmende Bedeutung von Daten in der Gesellschaft → 10

1.2 Warum es eine starke Forschung braucht → 14

1.3 Das Nationale Forschungsprogramm «Big Data» → 18

1.4 Die Struktur des NFP-75-Résumés → 20

2. Big-Data-Anwendungen → 23

2.1 Verbesserung und Personalisierung der Gesundheitsversorgung → 25

2.2 Förderung der Nachhaltigkeit → 28

2.3 Besseres Verständnis sozioökonomischer Wechselwirkungen → 31

2.4 Beschleunigung der Forschung → 32

2.5 Kernaussagen zu Big-Data-Anwendungen → 33

2.6 Die Forschungsprojekte zu Big-Data-Anwendungen → 36

3. Big-Data-Technologien → 45

3.1 Effizientere Big-Data-Infrastrukturen → 46

3.2 Neuartige Ansätze für Big-Data-Analytik → 47

3.3 Herausforderungen für die Forschung zu Big-Data-Technologien → 49

3.4 Kernaussagen zu Big-Data-Technologien → 50

3.5 Die Forschungsprojekte zu Big-Data-Technologien → 53

4. Gesellschaftliche, rechtliche und ethische Aspekte von Big Data → 59

4.1 Dateneigentum, -kontrolle, -zugriff und -übertragung → 61

4.2 Datenschutz und digitale Souveränität → 64

4.3 Fairness, Nicht-Diskriminierung und Inklusion → 66

4.4 Wissensproduktion und -management → 67

4.5 Herausforderungen und Kernaussagen → 68

4.6 Die Forschungsprojekte zu gesellschaftlichen, rechtlichen und ethischen Aspekten von Big Data → 70

5. Ausblicke → 75

5.1 Anwendungen wirken sich auf weitere Bereiche aus → 76

5.2 Reduktion des Fussabdrucks der Dateninfrastrukturen → 79

5.3 Abwägung des Datenschutzes → 80

5.4 Algorithmen zur Rechenschaft ziehen → 81

6. Schlussfolgerungen der Leitungsgruppe → 85

Anhang: Das Nationale Forschungsprogramm «Big Data» (NFP 75) → 96

Vorwort

Das allgegenwärtige Sammeln und Verarbeiten von Daten hat zu tiefgreifenden Veränderungen in unseren Lebensweisen geführt. Big-Data-Technologie ist mittlerweile unerlässlich für wettbewerbsfähige, technologiegetriebene Volkswirtschaften und birgt ein enormes Potenzial für wissenschaftliche Fortschritte und für die Gesellschaft insgesamt. Zugleich bringt sie zahlreiche gesellschaftliche Herausforderungen mit sich. Der Bereich ist gesamthaft eine Toppriorität für Spitzenforschung.

Das Nationale Forschungsprogramm «Big Data» (NFP 75) zielte auf technologische Grundlagenforschung, Forschung zu technologischen Anwendungen sowie auf gesellschaftlich orientierte Forschung ab. Zu diesem Zweck lancierte der Schweizerische Nationalfonds eine offene Ausschreibung. Ein internationales Gremium selektierte aus den zahlreichen eingereichten Vorschlägen 34 Forschungsprojekte. Das NFP 75 strebte nicht nach abschliessenden Antworten auf spezifische technologische oder gesellschaftliche Fragen, sondern zielte darauf ab, in der Schweiz vorhandene Kompetenzen durch verstärkte Forschung in diesem aufstrebenden Forschungsbereich weiterzuentwickeln. So haben Forschende auf der Basis ihrer Projekte in Natur-, Ingenieur- und Sozialwissenschaften Debatten angestossen zu Themen wie Datenrecht, Privatsphäre und Souveränität, der digitalen Kluft, Fairness und Rechenschaftspflicht von Algorithmen und dem angemessenen Mass an Regulierung.

Nach fünf Jahren intensiver Arbeit zeigt die Forschung im Rahmen des NFP 75 die Möglichkeiten für spezifische Big-Data-Anwendungen auf und bietet ein Paradebeispiel dafür, wie hochstehende Grundlagenforschung aus der Schweiz direkt zu Kerntechnologien beitragen kann. Das Erreichte unterstreicht, dass die Schaffung von Lösungen eine globale Perspektive erfordert, die von Anfang an solide ethische, rechtliche und sozioökonomische Dimensionen einschliesst.

Das NFP 75 komplementiert weitere Forschungsinitiativen wie Digital Lives (2018–2019) und das Nationale Forschungsprogramm «Digitale Transformation» (NFP 77, 2020–2025), das sich auf spezifische gesellschaftliche Auswirkungen der Digitalisierung konzentriert. Diese Initiativen unterstützen den Aufbau von Forschungs-, Innovations- und Bildungskompetenzen in der Schweiz, wie es die «Strategie Digitale Schweiz» des Bundesrates vorsieht.

Die im Rahmen des NFP 75 entstandenen Leistungen unterstreichen die Bedeutung interdisziplinärer Bemühungen, wenn es darum geht, neue Technologien auf verantwortungsvolle Weise zu entwickeln und zu nutzen. Zu den Herausforderungen gehören der Schutz der Privatsphäre und die Datensicherheit bei gleichzeitigem Zugang zu und gemeinsamer Nutzung von Daten sowie die Förderung der Kompetenzen und Geschicke, die Menschen in die Lage versetzen, sich das Wissen und Know-how anzueignen, zu entwickeln und in bestehende Prozesse zu integrieren.

Das NFP 75 hat wesentlich dazu beigetragen, die notwendigen Kompetenzen im Bereich Big Data in der Schweiz zu stärken, die Voraussetzungen für interdisziplinäre Innovation zu verbessern sowie Potenziale zu fördern, um geeignete soziale und rechtliche Lösungen zu finden. Dieses Résumé enthält einen Auszug aus den Ergebnissen und Schlussfolgerungen des Nationalen Forschungsprogramms «Big Data».

Bert Müller

Delegierter der Abteilung Programme des Nationalen Forschungsrats des Schweizerischen Nationalfonds ab Januar 2021

Friedrich Eisenbrand

Delegierter bis Dezember 2020

Executive Summary

Die fortschreitende Digitalisierung der Gesellschaft führt dazu, dass riesige Datenmengen gesammelt werden. Diese sogenannten Big Data bergen das Potenzial für eine umfassende gesellschaftliche, industrielle und wissenschaftliche Wertschöpfung – wenn sie effektiv genutzt werden. Zwischen 2017 und 2021 führte das Nationale Forschungsprogramm «Big Data» (NFP 75) Forschungsprojekte durch, die reale Anwendungen und neue Technologien entwickelten gesellschaftliche Aspekte im Zusammenhang mit Big Data untersuchten und so die Schweizer Forschungs- und Innovationskapazität im Bereich Big Data erhöhten.

Fünfzehn Forschungsprojekte haben konkrete Big-Data-Anwendungen entwickelt; in Kooperationen zwischen Fachleuten aus der Datenwissenschaft einerseits und solchen aus spezifischen Disziplinen andererseits (Kapitel 2). Die Ergebnisse heben die konkreten Auswirkungen hervor, die Big-Data-Innovationen in Bereichen wie der erneuerbaren Energien, der Patientenüberwachung in Spitälern, in der evidenzbasierten Politikgestaltung oder auch in der Wissenschaft selbst haben können. Die Projekte unterstreichen zudem die Bedeutung interdisziplinärer Teams, die in der Lage sind, fachübergreifend zu arbeiten und sich in den relevanten ethischen, rechtlichen und operativen Kontexten zurechtzufinden.

Elf Forschungsprojekte im Bereich Data Science untersuchten und entwickelten Technologien, die für die Nutzung aktueller und

zukünftiger Big Data benötigt werden. Sie befassten sich mit Infrastrukturaspekten wie Datenzugriff, -bereinigung, -indexierung und -vorverarbeitung sowie mit Analyseverfahren wie Abfrageverarbeitung, Data Mining und maschinellem Lernen, um die Wissensgewinnung aus Daten zu erleichtern (Kapitel 3). Diese Fortschritte sind in der Lage, die Funktionalität und die Leistung von Big-Data-Anwendungen zu erhöhen, z. B. durch die Verbesserung des Datenschutzes oder die Verringerung der für das Modelltraining beim maschinellen Lernen benötigten Rechen- und Datenressourcen.

Acht Forschungsprojekte untersuchten spezifische gesellschaftliche Aspekte des Einsatzes von Big Data, einschliesslich ethischer und rechtlicher Aspekte (Kapitel 4). Die Projekte untersuchten konkrete Beispiele für den Einsatz von Big Data, zum Beispiel im Personalwesen und im Versicherungssektor. Ihre Ergebnisse unterstreichen, wie wichtig es ist, die Gesetzgebung an die sich verändernden Möglichkeiten der Technologie anzupassen, Leitlinien für die Nutzung von Big Data zu entwickeln, das ethische Bewusstsein und die Transparenz im Zusammenhang mit der Nutzung von Daten zu erhöhen und genau zu verfolgen, wie sich Big Data auf die Demokratie auswirkt.

Drei Querschnittaktivitäten untersuchten Hindernisse für die gemeinsame Nutzung von Daten in der Forschung, förderten die Beteiligung von Frauen in der Big-Data-Wissenschaft

und erarbeiteten einen Überblick über die mit Big Data verbundenen gesellschaftlichen Probleme.

Die rasanten technologischen Fortschritte in Big Data bieten weiterhin neue Möglichkeiten in vielen Bereichen, z. B. in der industriellen Produktion, bei erneuerbaren Energien, der Cybersicherheit oder im elektronischen Handel (Kapitel 5). Diese Entwicklungen bringen auch Herausforderungen hervor, insbesondere einen steigenden Energieverbrauch bei der Datenverarbeitung, eine Gefährdung des Gleichgewichts zwischen Datenschutz und Wertschöpfung, potenzielle Risiken der Diskriminierung und die

Notwendigkeit einer Rechenschaftspflicht. Die Bewältigung dieser Herausforderungen wird für die Optimierung der Wertschöpfung aus Daten durch Unternehmen und öffentliche Einrichtungen entscheidend sein.

Die Schlussfolgerungen der Leitungsgruppe des NFP 75 enthalten Vorschläge, wie eine verantwortungsvolle Wertschöpfung aus Big Data erleichtert werden kann, und leisten einen Beitrag zu den politischen und fachlichen Debatten über diese neue Ressource. Sie sind im Folgenden zusammengefasst und werden im Bericht (Kapitel 6) ausführlicher behandelt.

Schlussfolgerungen der Leitungsgruppe

Förderung eines geeigneten Umfelds für die Entwicklung von Big Data

- (1) Die Ausbildung von Big-Data-Fachleuten ausbauen
- (2) Rechtliche und ethische Beratung für Big-Data-Forschungs- und -Entwicklungsprojekte fördern
- (3) Zertifizierungen der Eigenschaften von Big-Data-Anwendungen ermöglichen

Integrieren von Big Data in öffentliche und private Organisationen

- (4) Verstärkte Nutzung von Big-Data-Technologien im Gesundheitssektor
- (5) Die politische Entscheidungsfindung und deren Evaluation mit Big Data stärken
- (6) Geteilte Datensammlungen, Anwendungsbenchmarks und Open-Source-Software fördern

Aktualisierung und Schaffung einer angemessenen Regulierung

- (7) Proaktivere Regulierungen von Big Data anstreben
- (8) Datenschutz und digitale Souveränität in Big-Data-Anwendungen fördern
- (9) Verstärkte länderübergreifende Vereinheitlichung von Regulierungen



1.

Einführung: Big Data, grosse Veränderungen

Zunehmend elaboriertere Hardware- und Software-technologien ermöglichen die Sammlung und die Analyse von Datenmengen in nie dagewesenem Umfang. Diese bieten das Potenzial für eine weitreichende Wertschöpfung im öffentlichen und im privaten Sektor. Die verantwortungsvolle Umsetzung von Big-Data-Anwendungen erfordert Forschung zu allen Aspekten von Big Data, einschliesslich IT-Infrastrukturen, Datenanalysemethoden sowie ethischer Richtlinien und rechtlicher Rahmenbedingungen. Das NFP 75 hat entlang dieser Wertschöpfungskette bedeutende Beiträge geleistet und die Fähigkeit gestärkt, Big-Data-Technologien zu entwickeln, Anwendungen einzusetzen und die Regulierungen in der Schweiz anzupassen..

1.1

Die zunehmende Bedeutung von Daten in der Gesellschaft

Von der Digitalisierung zu Big Data

Angetrieben von öffentlicher und privater Forschung und Innovation schreitet der technologische Fortschritt immer schneller voran und er hat tiefgreifende Auswirkungen auf unsere Lebensweise. Die Fortschritte in der Informationstechnologie fördern eine weit verbreitete Digitalisierung in der gesamten Gesellschaft und führen zu einer Fülle von Daten, die einen immer grösseren Einfluss auf die Art und Weise haben, wie wir leben und arbeiten.

Die Kapazitäten für Datenverarbeitung, -speicherung und -kommunikation wachsen seit mehreren Jahrzehnten exponentiell. Die Dichte der Transistoren auf Mikrochips hat sich alle paar Jahre verdoppelt, was zu einer ähnlichen Beschleunigung der Leistung und der Datenübertragung geführt hat.

Diese spektakulären Fortschritte werden zum Teil durch die sehr grossen Skaleneffekte in der Halbleiter- und Kommunikationstechnologie-Branche und durch ein sehr schnelles Marktwachstum ermöglicht. In Verbindung mit Fortschritten in der Softwaretechnologie – insbesondere bei Betriebs- und Datenverwaltungssystemen, Programmiersprachen und Compilern sowie Analysemethoden wie maschinellem Lernen – haben sie beispiellose Steigerungen bei der Informationsverarbeitung und der Effizienz ermöglicht, zahllose neue Tools geschaffen und Branchen,

geschäftliche Prozesse und Lebensgewohnheiten tiefgreifend verändert.

Die zunehmende Digitalisierung der öffentlichen, privaten und persönlichen Sphären produziert immer grössere Datenmengen. Big Data bezieht sich im Allgemeinen auf Datensätze, deren Grössen und andere Eigenschaften Herausforderungen für die aktuellen Informations- und Kommunikationstechnologien darstellen und neue Lösungen erfordern. Neben der Grösse (oder dem Datenvolumen) beschreibt «Big Data» auch die Geschwindigkeit der Datenerstellung und -verarbeitung, ihre Vielfalt und ihre Wahrhaftigkeit (siehe «Wann sind Daten «gross»?» S. 11). Der Umfang von Big Data entwickelt sich daher mit dem technologischen Fortschritt ständig weiter.

Big Data erfassen einen immer grösseren Teil unseres gesellschaftlichen, beruflichen und individuellen Lebens, wie auch von Geschäftstransaktionen, Industrieprodukten und wissenschaftlicher Forschung. Einige grosse Datensätze sind persönlicher Natur, andere nicht. Die Daten werden von Websites, Apps, Kameras sowie von Sensoren in Smartphones, Fahrzeugen, industriellen Produktionslinien oder Umweltüberwachungssystemen erfasst.

Wertschöpfung generieren

Daten gelten als eine immens wertvolle Ressource, die bislang noch nicht vollständig genutzt wird. Sie werden als «das neue Öl» (oder «das neue Gold») bezeichnet, um ihre Rolle als unverzichtbare Ressource zu unterstreichen, die viele Prozesse antreibt und eine zentrale Rolle in der Gesellschaft spielt. Daten haben für sich genommen wenig Wert. Ihr Wert wird in Form von umsetzbaren Analyseergebnissen gewonnen. Es ist Planung erforderlich, um die

zu stellenden Fragen, aber auch um die Daten zu ermitteln, die Antworten liefern können. Es muss festgelegt werden, wie die Daten gesammelt oder abgerufen werden können, um damit effiziente Analysetools zu entwickeln und die Ergebnisse dieser Analysen in wertschöpfende Massnahmen umzusetzen. Dabei müssen auch die möglichen unbeabsichtigten Nebenwirkungen berücksichtigt werden.

Dieses Paradigma ist nicht neu. Es wird schon seit Jahrzehnten zur Unterstützung von Marktforschung, Kundenbefragungen, Finanzplanung oder Epidemiologie eingesetzt. Big Data bringt jedoch durch seine schiere Grösse, die erforderliche Rechen- und Kommunikationsinfrastruktur, die Komplexität der zur Analyse erforderlichen Algorithmen sowie den Umfang und die Herausforderungen der Anwendungen eine neue Dimension mit sich. Immer vielfältigere Beispiele zeigen, wie Big Data feinere Modelle ermöglichen kann, die wiederum verbesserte Geschäfts-, Industrie- oder Regierungsprozesse ermöglichen.

Die Nutzung von Daten hat jedoch neue Risiken und Herausforderungen für die Gesellschaft mit sich gebracht, was Sicherheit, Fairness und sozialen Zusammenhalt angeht. Diese erfordern angemessene und verhältnismässige Lösungen.

Anwendungen in der digitalen und realen Welt

Bekanntere Beispiele für Big-Data-Anwendungen sind internetbasierte

Dienste wie soziale Netzwerke, Messaging- und E-Mail-Dienste, Such- und Werbe- sowie E-Commerce- oder Unterhaltungsdienstleistungen.

Auch in den Bereichen Energie und Transport wird Big Data eingesetzt. So können beispielsweise Windturbinen, die mit Dutzenden von Sensoren ausgestattet sind, 20-mal pro Sekunde Datenpunkte erzeugen. Diese Informationen können in Echtzeit analysiert werden, um die Neigung der Turbinenblätter fein abzustimmen und ihre Effizienz zu maximieren. Während des Fluges können die Sensoren eines Flugzeugs ein Terabyte an Daten erzeugen. Wenn diese Daten ordnungsgemäss übertragen, gespeichert und analysiert werden, können Komponenten überwacht und ihre Wartung geplant werden (bevor es zu Ausfällen kommt), wodurch das Flottenmanagement optimiert und Ausfallzeiten reduziert werden.¹

Sehr grosse Datensätze werden routinemässig für die Umweltüberwachung und das Umweltmanagement verwendet. Sie helfen beispielsweise den Vereinten Nationen, den Verlust der Biodiversität zu bewerten und die Auswirkungen des Klimawandels zu verfolgen² oder die Verteilung von Schadstoffen in der Luft oder im Wasser vorherzusagen³. Ein Konsortium von Journalist*innen hat durch die Analyse von Millionen von durchgesickerten Dokumenten systematische internationale Steuerhinterziehung aufgedeckt⁴.

In der Forschung wird Big Data schon seit Jahrzehnten in internationalen Physik-Kollaborationen wie am CERN

Wann sind Daten «gross»?

Big Data ist ein sich fortentwickelndes Konzept, da es Datensätze beschreibt, deren Eigenschaften die aktuellen Technologien herausfordern.

Die **Datenmenge** (volume) geht in der Regel über Gigabytes (GB) hinaus und erreicht Terabytes (1000 GB) oder sogar Petabytes (1000 TB), was eine sehr leistungsfähige Speicher- und Verarbeitungsinfrastruktur erfordert. Die **Geschwindigkeit** (velocity) der Daten, d. h. die Produktions- oder Übertragungsraten oder die Analysegeschwindigkeit, kann ein GB pro Sekunde überschreiten, was sehr schnelle Hardware und effiziente Software erfordert.

Anwendungen kombinieren oft heterogene Datentypen (Text, Zahlen, Koordinaten, Bilder, Ton, Video usw.) mit sehr unterschiedlichen Eigenschaften – eine GPS-Ortung ist sehr präzise, während die Semantik von Texten oft mehrdeutig ist. Diese **Vielfalt** (variety) erfordert Algorithmen, die mit unterschiedlichen Datenformaten und -typen umgehen können.

Daten sind selten fehlerfrei, wahrheitsgemäss, genau, repräsentativ oder vollständig – Eigenschaften, die mit dem Begriff **Wahrhaftigkeit** (veracity) erfasst werden. Viele Big-Data-Anwendungen basieren auf mehr oder weniger genauen Modellen oder auf Techniken des maschinellen Lernens, die zunächst aus Trainingsdatensätzen von unterschiedlicher Qualität lernen, was die **Gültigkeit** (validity) der Ergebnisse beeinflusst.

Manchmal werden weitere «V» verwendet, um eine Big-Data-Anwendung zu beschreiben, darunter Datenvariabilität (variability), Anfälligkeit (vulnerability), Visualisierung (visualisation) oder Wert (value).

¹ The case for an industrial big data platform, General Electrics (2017)

² World environment situation room, United Nations Environment Programme, <https://data.unep.org/>

³ A new area of utilizing industrial Internet of Things in environmental monitoring, HH Lou et al. (2022) Front. Chem. Eng. 4:842514.

⁴ The Panama papers: exposing the rogue offshore finance industry, International Consortium of Investigative Journalists., <https://www.icij.org/investigations/panama-papers>

Wie gross ist Big Data?

Gegenwärtige Anwendungen können mit Datensätzen im Bereich Petabyte (1 Million GB) operieren

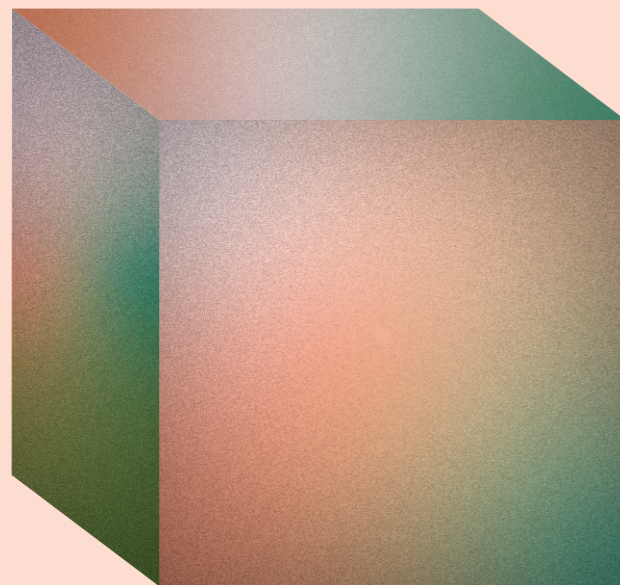
1 Gigabyte (GB)

- 1 Film
- 1000 Bücher



1 Terabyte (TB)

- 1 externe Festplatte
- 1000 Filme
- Trainingssatz künstlicher Sprachmodelle



1 Petabyte (PB)

- medizinische Bilder, die monatlich in einem Krankenhaus erstellt werden
- täglich auf YouTube hochgeladene Videos

1 Exabyte (EB)

- weltweiter täglicher mobiler Datenverkehr

oder durch leistungsstarke Teleskope erzeugt. Heute tragen sie zu einer wachsenden Zahl von Disziplinen bei, von der Biologie bis zu den Geisteswissenschaften. Ein grosser Durchbruch in den Biowissenschaften gelang beispielsweise im Juli 2022 mit der Veröffentlichung einer riesigen Datenbank möglicher 3D-Formen von fast allen der 200 Millionen bekannten Proteinen, ermittelt von Algorithmen maschinellen Lernens. Dies sind entscheidende Informationen, um die Rolle von Proteinen in physiologischen Prozessen zu verstehen, ihre Wirkung vorherzusagen und neue Behandlungsmethoden zu entwickeln⁵.

Verantwortungsvolle Nutzung sicherstellen

Die Leistungsfähigkeit von Big-Data-Anwendungen birgt auch zahlreiche Risiken. Eine verantwortungsvolle Wertschöpfung erfordert angemessene und verhältnismässige Lösungen in den Bereichen Datenschutz, Ethik, Wirtschaft, Recht und Unternehmensführung.

Hacker*innen geben regelmässig die privaten Daten von Millionen von Menschen preis. Internetunternehmen haben in einigen Fällen Kund*innendaten, inklusive Krankenakten, missbraucht, und die Erstellung von Kund*innenprofilen kann die Privatsphäre gefährden. Algorithmen maschinellen Lernens können die Verzerrungen in den Daten, die für das Training verwendet werden, verstärken und diskriminierende Ergebnisse erzeugen⁶. Unternehmen der Plattformökonomie haben den Arbeitsmarkt bereits auf den Kopf gestellt und stellen die Arbeitsvorschriften vor

völlig neue Aufgaben. Die Bewältigung dieser Probleme ist ein globales Unterfangen, an dem sowohl die privaten und öffentlichen Institutionen, die Big Data produzieren, speichern, übertragen, analysieren und nutzen, als auch regionale, nationale und internationale öffentliche Verwaltungen, Nichtregierungsorganisationen und schliesslich auch Bürger*innen beteiligt sind. Die rasche Anpassung des geltenden Rechts und die Verabschiedung neuer Gesetze sind notwendig, um sicherzustellen, dass bei der Nutzung von Big Data die Grundprinzipien des Datenschutzes, der Fairness, der Transparenz, der Rechenschaftspflicht und der Nichtdiskriminierung eingehalten werden.

1.2 Warum es eine starke Forschung braucht

Wissenschaftliche Forschung spielt eine zentrale Rolle bei der Entwicklung von Grundlagen für Big-Data-Anwendungen. Erstens trägt sie zur notwendigen Infrastruktur bei, z. B. mit stromsparenden Sensoren, die Daten für das Internet der Dinge produzieren, oder mit Technologien, die grosse Datenmengen speichern, übertragen und verarbeiten können. Zweitens stellt sie neue Werkzeuge zur Verfügung, um die von den Anwendungen benötigten Analysen und Vorhersagen zu erstellen, z. B. fortschrittliche statistische Methoden, Data-Mining-Algorithmen und maschinelles Lernen.

⁵ The entire protein universe, Ewen Callaway (2022), Nature 608, 15-16. Etwa 200 Millionen Proteinstrukturen werden vom internationalen Konsortium Uniprot gespeichert, das 2003 aus der Schweizer Initiative Swissprot hervorging.

⁶ Why algorithms can be racist and sexist, Rebecca Heilweil (2020), Vox.

Darüber hinaus ist Forschung dringend nötig, um die gesellschaftlichen Auswirkungen von Big Data besser zu verstehen. Es werden evidenzbasierte Empfehlungen benötigt, um bestehende und neue Herausforderungen anzugehen, einschliesslich Aufsicht und Regulierung sowie neuer Praktiken für Unternehmen und Behörden. Eine starke akademische Forschung bildet ausserdem die Grundlage für eine hochwertige Bildung; ein zentrales Element der «Strategie Digitale Schweiz»⁷.

Die Schweiz muss Zugang zu Spitzenkompetenzen im Bereich Big Data haben

Die Förderung einer äusserst wettbewerbsfähigen Big-Data-Forschung in der Schweiz ist aus mehreren Gründen von strategischer Bedeutung.

- Es ist entscheidend, am Puls der internationalen Entwicklungen zu bleiben, die Big Data prägen. Schweizer Forschungs- und Innovationsprojekte erhalten nur dann Zugang zu weltweiten Expert*innen und den neuesten Erkenntnissen, wenn sie selbst etwas zu bieten haben, denn die Besten arbeiten nur mit den Besten zusammen.
- Das derzeitige hohe Forschungs- und Ausbildungsniveau der Schweiz hat wichtige Big-Data-Unternehmen und einige grosse privatwirtschaftliche Forschungszentren ins Land gelockt.
- Die Schweiz muss sicherstellen, dass sie in der Lage ist, Spezialist*innen für Big Data, die international immer gefragter sein werden, auszubilden, anzuziehen und zu halten.
- Zahlreiche Aspekte von Big Data, insbesondere ethische, rechtliche und gesellschaftliche, sind für

die Schweiz spezifisch und erfordern daher Erkenntnisse von in der Schweiz tätigen Wissenschaftler*innen.

- Multinationale Unternehmen reizen oft die Grenzen von Big-Data-Technologien aus und bestimmen in diesem Bereich zunehmend den Kurs. Um eine gewisse Kontrolle über Technologien und die Richtung der Entwicklungen zu behalten, ist eine starke öffentliche Forschungsgemeinschaft notwendig.
- Die Forschung trägt dazu bei, dass die Öffentlichkeit informiert wird. Forschende tragen zur Verbreitung neuer Forschungsergebnisse bei und unterstützen die Öffentlichkeit, die Regierung und private Unternehmen dabei, fundierte Beiträge zu demokratischen Prozessen und Entscheidungsfindungen zu leisten.
- Qualitativ hochwertige Forschung trägt zu einer qualitativ hochwertigen Bildung bei, die für qualifizierte Arbeitskräfte im privaten und im öffentlichen Sektor entscheidend ist.

Dies sind die Hauptgründe, warum eine Stärkung der Big-Data-Forschung für die Schweiz von strategischer Bedeutung ist. Dazu hat das Nationale Forschungsprogramm «Big Data» einen wichtigen Beitrag geleistet.

⁷ <https://digital.swiss>

Die Big-Data-Pipeline

Daten werden über eine komplexe Wertschöpfungskette in Anwendungen verwandelt. Die **Daten** werden durch Online-Aktivitäten oder durch Sensoren erzeugt, bevor sie erfasst, geprüft, bereinigt und anonymisiert werden. Für die notwendige Big-Data-Infrastruktur werden Hardware- und Software**technologien** benötigt, die Daten speichern, den Zugriff verwalten und sichern, Vorverarbeitung und Analytik durchführen. **Anwendungen** arbeiten mit Datenmodellierungen, um Analysen, Vorhersagen oder

Empfehlungen zu erstellen. Diese wiederum müssen zur **konkreten Nutzung** validiert und in bestehende Prozesse integriert werden. Entlang der gesamten Wertschöpfungskette stellen sich **rechtliche, ethische und gesellschaftliche Fragen**, von Datenschutz bis hin zu potenzieller algorithmischer Verzerrung und Regulierung. Das Résumé des NFP 75 deckt den grössten Teil dieser Big-Data-Pipeline ab.



1.3

Das Nationale Forschungsprogramm «Big Data»

Neue Erkenntnisse zu Infrastrukturen, Anwendungen und gesellschaftlichen Aspekten

Das Nationale Forschungsprogramm «Big Data» (NFP 75)⁸ wurde 2014/2015 konzipiert. Es ergänzt die nationalen strategischen Programme zur Förderung der Digitalisierung, wie die «Strategie Digitale Schweiz» des Bundesrates, die branchenübergreifende Initiative DigitalSwitzerland, die Swiss Digital Initiative sowie spezifische Forschungsinitiativen wie Digital Lives.

Das NFP 75 wurde mit CHF 25 Millionen ausgestattet, womit ein Portfolio von Forschungsprojekten finanziert wurde, die strenge Kriterien der wissenschaftlichen Exzellenz erfüllten.⁹ Sie liefen zwischen 2017 und 2021 und liessen sich drei Kategorien zuordnen:

- Grundlegende Innovationen bei den für Big-Data-Anwendungen erforderlichen IT-Infrastrukturen
- Projekte anwendungsorientierter Forschung zur Entwicklung konkreter Anwendungen
- Forschung zum Zusammenspiel von Big Data und Gesellschaft, einschliesslich rechtlicher, ethischer und gesellschaftlicher Aspekte

Die Ausschreibung zur Einreichung von Vorschlägen des NFP 75 hielt vier Ziele fest:

- Das Erzielen von Fortschritten in den Bereichen Computing und Informationstechnologie
- Das Adressieren gesellschaftlicher, wirtschaftlicher, regulatorischer (inkl. regionaler und globaler) und bildungspolitischer Herausforderungen
- Die Erarbeitung von Anwendungen
- Die Stärkung der Forschungsressourcen

Das Programm erbrachte wissenschaftliche Fortschritte, die zu effizienteren Big-Data-Infrastrukturen beitragen, entwickelte konkrete Anwendungen in verschiedenen Bereichen, lieferte Erkenntnisse zu Möglichkeiten der Bewältigung gesellschaftlicher Aspekte und stärkte die Forschungs- und Innovationskapazitäten im Bereich Big Data in der Schweiz.

Rückblick: wichtige Beiträge des NFP 75

Das NFP 75 wurde 2014 konzipiert, zu einer Zeit, als viele Big-Data-Technologien und -Themen, die heute bekannt sind, gerade erst aufkamen. Die sehr schnellen Fortschritte in den Big-Data-Technologien und deren Einsatz in der Gesellschaft waren eine Herausforderung für die Forschungsprojekte, die flexibel sein und ihre Ziele anpassen mussten. Rückblickend lässt sich sagen, dass sowohl die Ausrichtung des Programms als auch die geförderten Projekte wesentliche Fragen entlang der gesamten Pipeline von

⁸ Nationale Forschungsprogramme (NFP) ermöglichen thematischen Forschungskonsortien, Themen zu behandeln, die für die Schweiz von Bedeutung sind. Sie werden dem Staatssekretariat für Bildung, Forschung und Innovation von Verwaltungseinheiten, Forschungsinstituten oder Einzelpersonen vorgeschlagen. Sie werden vom Bundesrat genehmigt und vom Schweizerischen Nationalfonds (SNF) umgesetzt.

⁹ Siehe auch Anhang: Das Nationale Forschungsprogramm «Big Data» (NFP 75), S. 96.

Big-Data-Anwendungen abgedeckt haben (siehe «Die Big-Data-Pipeline», S. 16).

Die vor acht Jahren getroffene Entscheidung, auch gesellschaftliche Herausforderungen in das Forschungsprogramm aufzunehmen, hat sich als richtig erwiesen. Das zeigen die mittlerweile zahlreichen Diskussionen zu Fairness und Bias bei künstlicher Intelligenz, Datensouveränität oder der Bedeutung neuer Anwendungen für Bürger*innen und Arbeitnehmer*innen.

Die Finanzierung von Projekten zur Entwicklung konkreter Anwendungen ermöglichte interdisziplinäre Zusammenarbeiten. Diese Kooperationen brachten Fachleute aus verschiedenen Disziplinen mit Computerwissenschaftler*innen und Partner*innen aus dem öffentlichen und dem privaten Sektor zusammen. Sie trugen zur Stärkung des interdisziplinären Know-hows bei, das für die Entwicklung von Big-Data-Anwendungen erforderlich ist, und lieferten wertvolle Erfahrungen für das kommende Jahrzehnt, nicht nur im Bereich Big Data, sondern auch in den Datenwissenschaften im Allgemeinen. Es wird erwartet, dass solche Kooperationen zunehmend wichtigere Rollen bei der Bewältigung globaler gesellschaftlicher Herausforderungen spielen werden, die in den Zielen für nachhaltige Entwicklung der UNO formuliert sind; darunter der Klimawandel, die Umweltkrise und die zunehmend älter werdende Bevölkerung.

Schliesslich haben Forschungsprojekte, die sich mit den technologischen Problemstellungen der Weiterentwicklung von Big-Data-Infrastrukturen befassten, das in der Schweiz vorhandene Fachwissen zur Entwicklung und Gestaltung von Big-Data-Technologien gestärkt.

Gesellschaftliche Auswirkungen des NFP 75

Die Outreach-Aktivitäten des NFP 75 brachten das Thema der Bedeutung von Big Data für die Gesellschaft einem breiteren Publikum nahe. Diese Aktivitäten umfassten Diskussionen über rechtliche und ethische Aspekte von Big Data, sie befassten sich mit Big Data und der Gleichstellung der Geschlechter, und Schulen wurde Unterrichtsmaterial zu Big Data zur Verfügung gestellt. Insgesamt hat das NFP 75 dazu beigetragen, dass die Schweiz auf verantwortungsvolle Weise von Big Data profitieren kann.

Einige Projekte brachten wertvolle Erkenntnisse für die Politikgestaltung über das Thema Big Data hervor, indem sie konkrete Daten analysierten, insbesondere in den Bereichen erneuerbare Energien, Umweltmanagement und Sozioökonomie (siehe «Erkenntnisse über Big Data hinaus», S. 35).

Die im Zusammenhang mit Big Data entstehenden Probleme können nicht ein für allemal gelöst werden, aber um sie zu bewältigen, muss die Schweiz in den Bereichen Forschung, Bildung und Innovation auf dem neuesten Stand der Technik bleiben. Das NFP 75 lenkte die Aufmerksamkeit Schweizer Wissenschaftler*innen sowie privater und öffentlicher Akteur*innen auf die Herausforderungen von Big Data, indem es über eine offene und wettbewerbsorientierte Ausschreibung Forschung mit hohen Qualitätsstandards finanzierte. Das Programm hat dazu beigetragen, der Schweizer Big-Data-Forschung ein solides Fundament zu geben und ihre Relevanz und ihren Einfluss zu stärken. Letztere werden auch stark vom laufenden Nationalen Forschungsprogramm «Digitale Transformation» (NFP 77) profitieren, das drei Jahre nach dem NFP 75 lanciert wurde

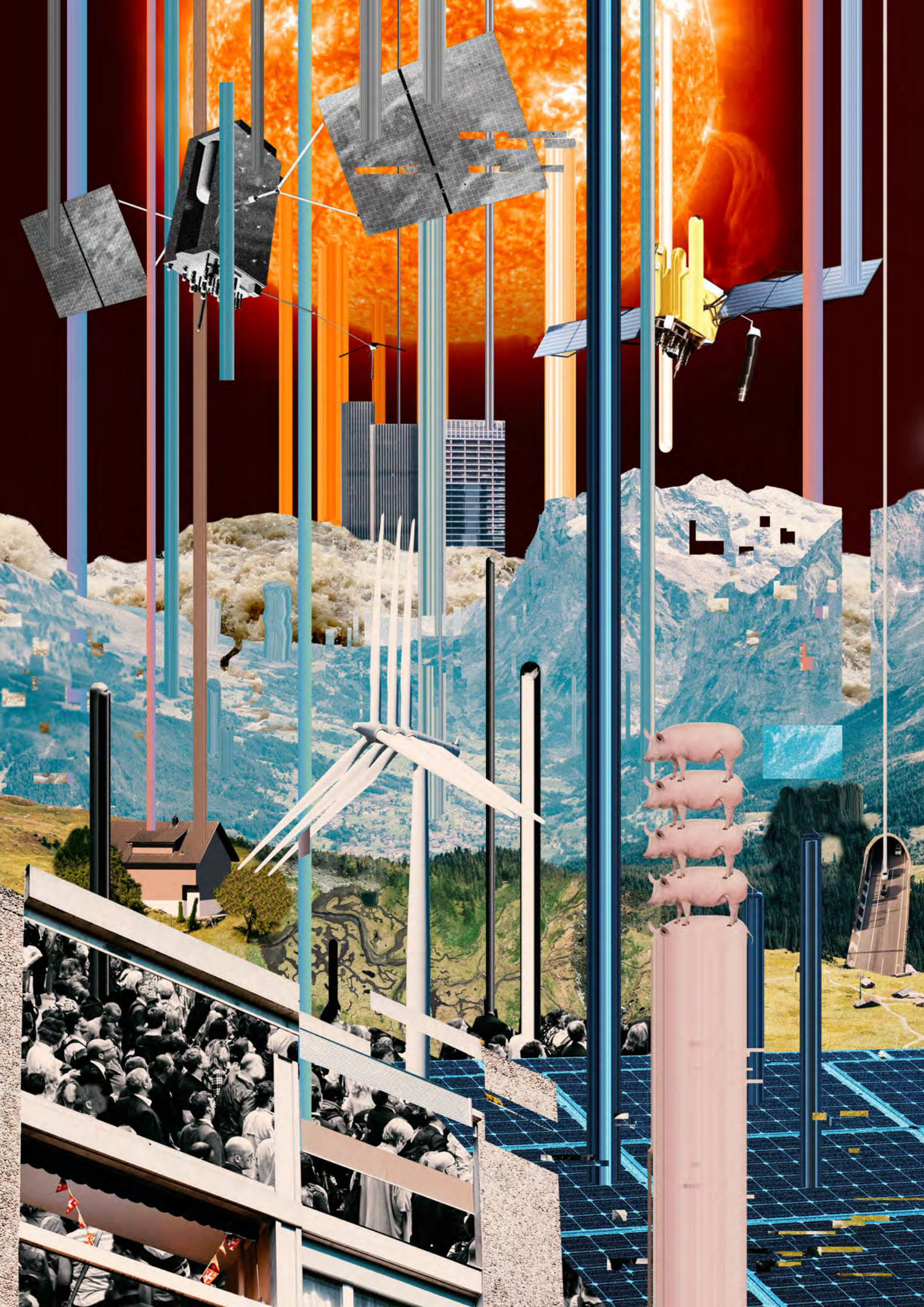
und sich auf die Auswirkungen der Digitalisierung in den Bereichen Bildung, Arbeitsmarkt, Vertrauenswürdigkeit und Governance konzentriert.

1.4 Die Struktur des NFP-75- Résumés

Das Résumé des NFP 75 «Big Data» fasst die wichtigsten Erkenntnisse aus den geförderten Forschungsprojekten zusammen und gibt einen Ausblick auf die Herausforderungen von Big Data.

Kapitel 2 «Big-Data-Anwendungen» beschreibt Möglichkeiten, die sich durch die Analyse grosser Datenmengen in konkreten Anwendungen ergeben, wie sie von NFP-75-Projekten zu Bereichen wie Gesundheit, Nachhaltigkeit und Sozioökonomie erforscht wurden. Kapitel 3 «Big-Data-Technologien» fasst die Fortschritte zusammen, die im Rahmen des NFP 75 erzielt wurden, um die technischen Herausforderungen von Big Data zu bewältigen, insbesondere effizientere Computerinfrastrukturen sowie neue Ansätze zur Datenanalyse. Kapitel 4 «Gesellschaftliche, rechtliche und ethische Aspekte von Big Data» bietet Erkenntnisse zu gesellschaftlichen Herausforderungen von Big Data, insbesondere Dateneigentum und Datenschutz sowie Fairness. Es enthält konkrete Leitlinien, die im Rahmen von NFP-75-Projekten entwickelt wurden. Kapitel 5 «Ausblicke» gibt einen allgemeineren Ausblick auf die Chancen und Risiken im Zusammenhang mit Big Data, die in den kommenden Jahren an Bedeutung gewinnen könnten. Die Schlussfolgerungen

der Leitungsgruppe des NFP 75 werden in Kapitel 6 vorgestellt. Der Anhang schliesslich liefert die wichtigsten Fakten zum Nationalen Forschungsprogramm «Big Data» (NFP 75). Das Résumé berichtet über die wichtigsten Aspekte von Big Data entlang der Pipeline, die von den Rohdaten zu den realen Anwendungen führt, wie dargestellt in «Die Big-Data-Pipeline», S. 16.



2.

Big-Data- Anwendungen

Big-Data-Anwendungen eröffnen Chancen in zahlreichen Bereichen. Ihre Entwicklung ist jedoch mit erheblichem Aufwand verbunden: dem Aufbau von Partnerschaften mit Interessengruppen, der technischen und rechtlichen Sicherung des Zugangs zu Daten, der Entwicklung nützlicher Analysemodelle und der Benutzendvalidierung. Dieses Kapitel stellt über ein Dutzend neuer Anwendungen vor, die im Rahmen von NFP-75-Projekten in den Bereichen Gesundheit, Nachhaltigkeit, Politik und wissenschaftliche Forschung entwickelt wurden, und zeigt die daraus entstandenen Lerneffekte auf.

Die immer stärkere Nutzung digitaler Technologien durch die Gesellschaft erzeugt immer grössere Datenmengen. Viele Akteur*innen möchten diese Ressource nutzen, um neue Instrumente zu schaffen oder bestehende Instrumente genauer und effizienter zu gestalten. Dies führt zu einer Flut von Big-Data-Anwendungen in zahlreichen Bereichen.

Trotz des derzeitigen Hypes bleibt die Entwicklung von Anwendungen auf der Grundlage von Big Data eine komplexe und langwierige Aufgabe, die mit vielen Herausforderungen verbunden ist. Der Entwicklungsprozess erfordert das Zusammenbringen von Partner*innen, die Adressierung rechtlicher und ethischer Herausforderungen im Zusammenhang mit Datenschutz und Fairness, den Zugang zu Daten und deren Analysen, die Entwicklung von Modellen, das Programmieren der Anwendungen und die Bewertung ihrer Genauigkeit und Nützlichkeit. Erst dann kommen die letzten Schritte: die Validierung der Anwendungen im Austausch mit den Nutzer*innen, ihre Einbindung in bestehende Abläufe und die Sicherstellung einer angemessenen Wartung.

Das Nationale Forschungsprogramm «Big Data» (NFP 75) hat Anwendungen auf der Grundlage von Big Data in Bereichen wie Gesundheit, Nachhaltigkeit, Sozioökonomie und wissenschaftliche Forschung hervorgebracht. Dies stellt nur einen kleinen Teil der Bereiche dar, die Big-Data-Anwendungen erforschen oder mit ihnen arbeiten. Sie reichen von traditionellen datenintensiven Bereichen wie Bankwesen, Marketing und Gesundheit bis hin zu neuen Gebieten wie Landwirtschaft, Journalismus oder politischer Entscheidungsfindung.

Die vom Programm geförderten Projekte haben bestehende Methodologien

verbessert, neue, auf bereichsspezifische Big-Data-Anwendungen ausgerichtete Methoden, entwickelt und den potenziellen Mehrwert für Gesellschaft und Wirtschaft hervorgehoben. Beispiele sind die Entwicklung von Strategien für die personalisierte Medizin, eine intelligentere Verkehrsplanung, der integrierte Einsatz erneuerbarer Energien und eine klarere Einschätzung der Auswirkungen sozioökonomischer Massnahmen.

Die verschiedenen Anwendungen des NFP 75 haben unterschiedliche Entwicklungsstadien erreicht, von ersten Modellen über Prototypen bis hin zu vollwertigen Systemen. Diese Vielfalt spiegelt die Möglichkeiten und Herausforderungen bei der Entwicklung praktischer Big-Data-Anwendungen im Allgemeinen wider. Die Anwendungen wie die Kernaussagen dieser Projekte sind in vier Bereiche unterteilt:

- Gesundheitsbereich in Abschnitt 2.1
- Nachhaltigkeit, einschliesslich Transport, Energie, Lebensmittelversorgung und Umweltmanagement, in Abschnitt 2.2
- Sozioökonomische Fragestellungen in Abschnitt 2.3
- Forschung in Abschnitt 2.4.

Der Abschnitt 2.5 enthält eine Zusammenfassung der Erkenntnisse aus dem NFP 75 über die Entwicklung von Big-Data-Anwendungen. Die Forschungsprojekte werden in Abschnitt 2.6 beschrieben.

2.1

Verbesserung und Personalisierung der Gesundheitsversorgung

Zahlreiche neue Ansätze sollen die Gesundheitsversorgung auf die spezifischen Merkmale und Bedürfnisse von Einzelpersonen und Bevölkerungsgruppen zuschneiden, und zwar nach dem sogenannten «P4»-Paradigma der prädiktiven, präventiven, personalisierten und partizipativen Medizin. Dieser Ansatz beruht auf einem besseren Zugang zu Daten, der Entwicklung robuster Analyseinstrumente und der Förderung einer engen Zusammenarbeit mit Ärzt*innen und Patient*innen, um die neuen digitalen Instrumente an deren Bedürfnisse anzupassen. Gesundheitsdaten stammen aus traditionellen Datenbanken wie elektronischen Krankenakten und aus neuen Quellen wie Smartphones und tragbaren Sensoren – Technologien, die immer alltäglicher werden. Da medizinische Aufzeichnungen als sensible Informationen gelten, müssen Big-Data-Anwendungen strenge Datenschutzvorgaben einhalten. Diese Gesetze variieren von Land zu Land und stellen eine Herausforderung für die Zusammenarbeit zwischen verschiedenen Institutionen dar, wie in Kapitel 4 beschrieben. Big-Data-Anwendungen können einen erheblichen Einfluss auf Gesundheitsforschung, -bildung und -versorgung haben und über medizinische Einrichtungen hinaus bis in die Wohnungen der Menschen reichen.

Die NFP-75-Forschungsprojekte zum Gesundheitsbereich (siehe ihre Beschreibung im Kapitel 2.6) stellen nur

eine sehr kleine Teilmenge möglicher Anwendungen dar. Sie liefern jedoch konkrete Beispiele dafür, wie der Gesundheitsbereich von Big Data profitieren kann:

- Durch die Ermöglichung eines effizienteren Managements kritischer Situationen auf Intensivstationen mittels Überwachung von Patient*innen und der Erstellung genauer Vorhersagen über ihren Zustand (siehe NFP-75-Forschungsprojekt *Intensivstationen*);
- Durch die Entwicklung eines Prototyps für ein Smartphone-basiertes System zur personalisierten Behandlung von Schmerzen im unteren Rückenbereich (siehe Projekt *Rückenschmerzen*);
- Durch die Entwicklung neuer Methoden biomedizinischer Forschung (siehe die in Abschnitt 2.4 diskutierten Projekte *Big Data Genetik*, *Genomvergleich* und *Bioinformatik-Datenbanken*).

Kernaussagen zu Anwendungen im Gesundheitsbereich

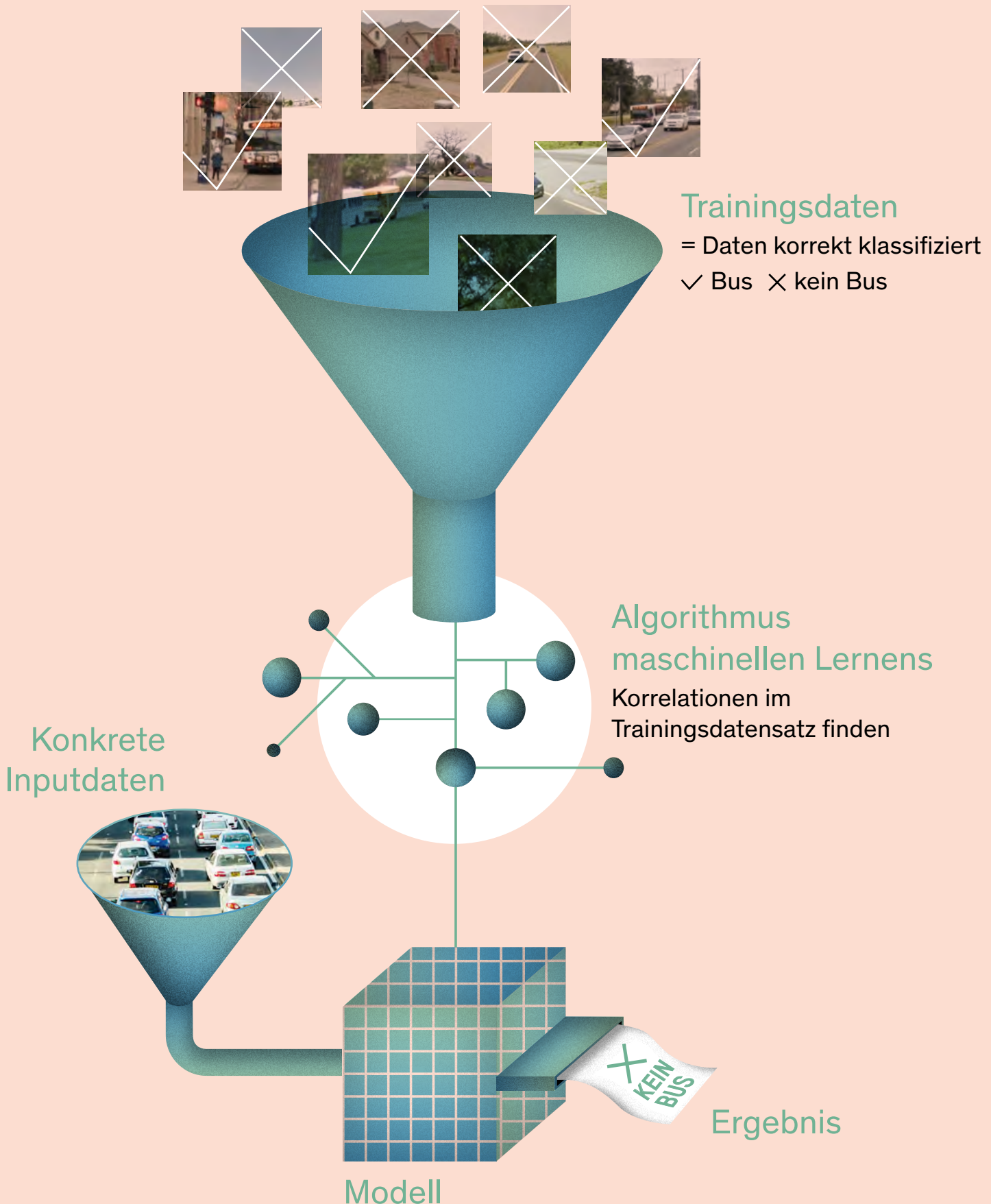
Das NFP 75 zeigt das Potenzial von Big-Data-Anwendungen zur Unterstützung einer modernen und effizienten Gesundheitsversorgung, aber auch viele Herausforderungen. Insbesondere die Entwicklung nützlicher und praktischer Anwendungen erfordert eine robuste Infrastruktur für die gemeinsame Nutzung von Daten, einschliesslich geeigneter Lösungen für technische, organisatorische und rechtliche Fragen.

Gesellschaftliche Auswirkungen

Die Automatisierung der Analyse der Patient*innenüberwachung kann die Qualität der Pflege verbessern, indem die Rate an Fehlalarmen, die das Personal warnen, verringert wird (*Intensivstationen*). Mobile Apps können

So funktioniert überwachtes maschinelles Lernen

Ein Algorithmus lernt, anhand klassifizierter Daten eine Aufgabe zu lösen (hier: einen Bus zu erkennen). Je grösser der Trainingsdatensatz ist, desto genauer sind die Ergebnisse in der konkreten Anwendung.



Patient*innen dabei helfen, physiotherapeutische Übungen zu Hause durchzuführen und Daten zu sammeln, mit deren Hilfe wiederum die Wirkung der Therapie bewertet werden kann (*Rückenschmerzen*). Auch die biomedizinische Forschung kann von Big Data profitieren (siehe Abschnitt 2.4). Mehrere methodische und technische Fortschritte verbessern die Wissenschaft der Genomik im Hinblick auf die klinische Forschung, die Epidemiologie und die Umweltbiologie (*Big Data Genetik, Genomvergleich*) sowie die Nutzung biomedizinischer Datenbanken durch ein breiteres Spektrum von Spezialist*innen (*Bioinformatikdatenbanken*).

Wissensgewinn zu Big Data im Gesundheitswesen

Einer Anwendung gelang es, multimodale Daten (Video, Sensordaten, medizinische Aufzeichnungen) in einen Algorithmus für maschinelles Lernen zu integrieren (*Intensivstationen*). Eine andere Anwendung kombinierte implizite und explizite Informationen und zeigte, dass von Sensoren erfasste Daten verwendet werden können, um den Wahrheitsgehalt von Beobachtungen, wie z. B. Tagebucheinträgen von Patient*innen über ihre Bewegungsgewohnheiten, zu bewerten (*Rückenschmerzen*).

Herausforderungen

Es ist notwendig, neue Wege zur Erfassung und Verwaltung von Gesundheitsdaten zu entwickeln, die im Allgemeinen in getrennten Datensilos gespeichert werden. Insbesondere die Umsetzung eines nationalen Gesundheitssystems, das auf der Grundlage transparenterer und gemeinsam genutzter Daten aufbaut, würde neue digitale Lösungen für die Gesundheitsversorgung erleichtern (*Rückenschmerzen*). Die Zusammenführung verschiedener Datenquellen wird durch die grosse Heterogenität der

Schweizer Gesundheitssysteme und -vorschriften sowie durch die Tatsache erschwert, dass die von individuellen medizinischen Geräten gesammelten Daten in hohem Masse proprietär sind (aufgrund fehlender kommerzieller und regulatorischer Anreize zur Integration der Daten) (*Intensivstationen*). Die prospektive Erfassung von klassifizierten Gesundheitsdaten und die Verwaltung der regulatorischen Prozesse für den Umgang mit den Daten nehmen viel Zeit und Mühe in Anspruch.

Mobile Gesundheits-Apps, die personalisierte körperliche Übungen zur medizinischen Rehabilitation oder Schmerzbehandlung anbieten, können Patient*innen helfen, sich an Behandlungsprotokolle zu halten, sind aber letztlich immer noch von deren Motivation abhängig (*Rückenschmerzen*). Die Digitalisierung von Gesundheitsdaten kann auch zu einem Misstrauen der Öffentlichkeit gegenüber staatlichen und privaten Gesundheitsunternehmen führen, wie die Kritik an der Anwendung SwissCovid App zur Ermittlung von Kontaktpersonen gezeigt hat.

Lösungsansätze

Eine Reihe von Massnahmen kann ergriffen werden, um im Gesundheitsbereich die Entwicklung von Anwendungen zu verbessern und die Nutzung von Big Data zu fördern.

— Das Verallgemeinern der Datenerfassung für bestimmte Krankheiten auf alle medizinischen Zentren in der Schweiz. Die traditionelle Struktur mit einzelnen Studienzentren, die an bestimmte Spitäler oder Kantone gebunden sind, ist für Big-Data-Anwendungen im Gesundheitswesen nicht geeignet. Dies ist notwendig, um genügend Daten zu generieren, die Erkenntnisse zu komplexen Krankheiten ermöglichen. Dabei sollten die Fortschritte beachtet werden, die multinationale

Technologieunternehmen bei der Erfassung, Speicherung und Analyse privater und öffentlicher Gesundheitsdaten machen.

- Das Rationalisieren der Prozesse, die für die Einhaltung von Gesetzen notwendig sind, z. B. die Verwendung digitaler Signaturen oder die Einholung der Zustimmung von Patient*innen zur Verwendung ihrer Daten in der medizinischen Forschung. Dabei muss sichergestellt werden, dass Regulierungsbehörden wie die Ethikkommissionen über die neuesten Entwicklungen bei Big-Data-Anwendungen informiert sind (*Rückenschmerzen*), siehe auch Kapitel 4.
- Die Notwendigkeit zur Vereinheitlichung der von medizinischen Geräten verwendeten Datenformate muss betont und eine stärkere Verwendung von Standardformaten in elektronischen Krankenakten gefördert werden.
- Alle Interessensgruppen müssen angehalten werden, qualitativ hochwertige Metadaten zu pflegen, die für eine effiziente Indexierung und Suche in grossen Datenbanken unerlässlich sind (*Intensivstationen*).

2.2 Förderung der Nachhaltigkeit

Der Umbau zu nachhaltigen Gesellschaften erfordert die Optimierung der Wechselwirkungen zwischen den zahlreichen Komponenten von Energie-, Transport-, Versorgungs- oder Lebensmittelsystemen. So hängen beispielsweise Mobilitätsmuster sowohl vom Verkehrsangebot als auch von der Nachfrage der Nutzer*innen (die unterschiedliche Bedürfnisse und Vorlieben

haben) ab, aber auch von komplexen Wechselwirkungen zwischen Faktoren wie dem Wetter, der Akzeptanz von Homeoffice und finanziellen Anreizen.

Big-Data-Anwendungen können Effizienz und Nachhaltigkeit von Systemen erheblich verbessern. Dazu ist es oft erforderlich, feinkörnige Daten in Echtzeit zu analysieren, um einzelne Aktivitäten wie die Mobilität von Menschen in öffentlichen Verkehrsmitteln oder die Leistung von Solarzellen auf bestimmten Dächern zu überwachen. Solche Erkenntnisse helfen dabei, das Angebot kontinuierlich zu optimieren, um die erwartete Nachfrage zu befriedigen, indem beispielsweise die Fahrpläne der Verkehrsmittel angepasst oder die Stromproduktion variiert wird.

Das NFP 75 hat ein halbes Dutzend innovativer Anwendungen vorgestellt, die eindeutig das Potenzial haben, Nachhaltigkeit zu fördern:

- Hochauflösende Karten auf regionaler und nationaler Ebene zeigen, wie Solar-, Wind- und geothermische Energie für einzelne Gebäude genutzt werden können (*Potenzial erneuerbarer Energien*).
- Hochauflösende digitale 3D-Modelle von Städten (*Digitale Zwillingstädte*) und die Verwendung von GPS-Daten ermöglichen Untersuchungen zur Optimierung des öffentlichen und des privaten Verkehrs (*Optimierung der Verkehrssysteme*).
- Neu entwickelte Tools quantifizieren anhand von Luftbildern automatisch das Ausmass erodierten Geländes (*Bodenerosion*) und können mittels der Verwendung von Smartphone- oder Überwachungsvideos zur automatischen Erkennung von Überschwemmungen eingesetzt werden (*Hochwassererkennung*).
- Eine Pilotplattform sammelt und analysiert für die Schweizer Schweineindustrie relevante Daten (*Pig Data*).

Kernaussagen zu Nachhaltigkeitsanwendungen

Bemühungen zur Nachhaltigkeit können stark von Big-Data-Anwendungen profitieren, da Daten im Maschinenbau und in der Optimierung von Effizienz und Sicherheit schon seit Langem eingesetzt werden. Die fortschreitende Digitalisierung, einschliesslich des Einsatzes von Geräten für das Internet der Dinge, erzeugt immer grössere Datensätze und verbessert damit die Spielräume für Big-Data-Anwendungen. Diese Anwendungen erfordern die Integration von Daten aus verschiedenen Quellen und die Entwicklung leistungsstarker neuer Analysetools. Das dringende Bedürfnis der Gesellschaft nach Nachhaltigkeit erfordert eine stärkere Nutzung von Big Data.

Gesellschaftliche Auswirkungen

Die im Rahmen des NFP 75 entwickelten Anwendungen zeigen, dass der Einsatz von Big Data zur Förderung von Nachhaltigkeit beitragen kann. Insbesondere kann Big Data auch bei der Evaluierung und Modifizierung nationaler und lokaler Strategien für erneuerbare Energien helfen, realistische Szenarien für das Verkehrsmanagement entwickeln, 3D-Karten für die Stadtplanung erstellen, die Umwelterstörung überwachen und im Umgang mit Naturgefahren dienen.

Wissensgewinn zu Big Data im Bereich Nachhaltigkeit

Das Programm NFP 75 hat gezeigt, dass neue Techniken dünne Datensätze ergänzen können, um Modelle der Energieproduktion in hoher Auflösung und mit quantifizierbaren Unsicherheiten zu erstellen (*Potenzial erneuerbarer Energien*). Neue Algorithmen können hochauflösende numerische 3D-Modelle von städtischen Gebieten (digitale Zwillinge) erstellen, wobei auf den Fahrzeugen montierte Kameras

kontinuierlich in alle Richtungen aufzeichnen (*Digitale Zwillingstädte*). Es ist auch möglich, Erkenntnisse zu spezifischen Mobilitätsdaten aus globalen, aggregierten und anonymen Datenbanken zu gewinnen (*Optimierung der Verkehrssysteme*). Überwachtes maschinelles Lernen kann die wichtigsten Faktoren identifizieren, die Bodenerosion und Erdbeben beeinflussen (*Bodenerosion*). Effiziente Schnittstellen für die Visualisierung von und Interaktion mit Informationen auf verschiedenen Ebenen können entwickelt werden, indem Endnutzer*innen frühzeitig in die Entwicklung von Anwendungen einbezogen werden (*Hochwassererkennung*).

Herausforderungen

Wie viele andere sind auch die für Nachhaltigkeit konzipierten Anwendungen von der Verfügbarkeit umfassender und hochwertiger Daten abhängig. Die Umweltdaten, die für den Einsatz von Geräten zur Gewinnung erneuerbarer Energien benötigt werden, sind oft nicht vorhanden oder unvollständig (*Potenzial erneuerbarer Energien*). Der Zugang zu Daten über die Bewegungen der Menschen erfordert die Unterstützung der Wirtschaft, insbesondere der Kommunikationsanbieter. Die aktuellen Datenschutzbestimmungen können einen solchen Zugriff behindern, obgleich es Möglichkeiten gibt, das Risiko von Datenlecks zu begrenzen, indem beispielsweise ausschliesslich auf ein sicheres Speicher-Backend Zugriff gewährt wird (*Optimierung der Verkehrssysteme*).

Algorithmen des überwachten maschinellen Lernens, die mit klassifizierten Daten trainiert werden, funktionieren in realen Anwendungen nicht immer zuverlässig. Manchmal können diese von fortschrittlicheren Ansätzen profitieren, z. B. dem unüberwachten Lernen, bei dem keine Trainingsdaten verwendet

Förderung der Big-Data-Kompetenz in Schulen

werden. Wichtige Metadaten, wie der GPS-Standort von Videos, fehlen oft bei Crowdsourced-Daten, was die Analyse sehr viel schwieriger und zeitaufwändiger macht (*Hochwassererkennung*). Natürliche Variationen auf Luftbildern – Schatten, gemähte Felder etc. – machen es schwierig zu analysieren, wie sich das Terrain im Laufe der Zeit verändert (*Bodenerosion*).

Es ist schwierig, Big-Data-Anwendungen in Branchen zu entwickeln, in denen die Datennutzung fragmentiert bleibt und in denen Analysen durch die geringe Qualität der Daten, den begrenzten Umfang oder das Fehlen konsistenter Ontologien (die Art und Weise, wie Daten konzeptualisiert, kategorisiert und dargestellt werden) behindert werden. Dies ist auch in Branchen der Fall, in denen Vertraulichkeit wichtig ist und Daten nur widerwillig weitergegeben werden (*Pig Data*). Schliesslich können personelle Veränderungen in den Partnerorganisationen eine geplante Zusammenarbeit plötzlich zum Erliegen bringen.

Lösungsansätze

- Das Entwickeln von Privacy-by-Design-Ansätzen, die Lösungen für Datenschutzfragen von Beginn des Projekts an integrieren und Expert*innen für rechtliche und gesellschaftliche Aspekte frühzeitig einbeziehen (*Optimierung der Verkehrssysteme, Digitale Zwillingstädte*), siehe auch Kapitel 4.
- Strategien entwickeln, die den Austausch sensibler Daten zwischen Industrie und Wissenschaft zu Forschungszwecken unter Wahrung der Vertraulichkeit und des Datenschutzes (*Pig Data*) ermöglichen und erleichtern.
- Eine Kultur der Digitalisierung fördern. Das Bewusstsein für die Probleme und Herausforderungen der Datennutzung schärfen und die

Das Erlernen der Nutzung von Informations- und Computertechnologien ist Teil des Deutschschweizer Lehrplans 21 («Medien und Informatik»). Der Erwerb von Datenkompetenzen erfordert allerdings neue und optimierte Unterrichtsmedien. In Zusammenarbeit mit dem Museum für Kommunikation in Bern hat das NFP 75 Unterrichtsmaterial zum Thema Big Data für die Sekundarstufe I und II erstellt, das acht Module mit jeweils zwei Unterrichtsstunden umfasst. Das Anfang 2020 lancierte Material wurde bereits mehr als 12 000-mal heruntergeladen. Die Initiative wurde vom Museum für Kommunikation und von der Schweizerischen Akademie der Technischen Wissenschaften (SATW) gefördert.

Big Data: Lehrmittel für die Sekundarstufen / Big Data: outil pédagogique pour les cycles secondaires, NFP 75 und Museum für Kommunikation, Bern (2020)

kontinuierliche Weiterbildung in diesem Bereich unterstützen. Erfahrene Mitarbeitende in technischen und angewandten Wissenschaften einbinden, um die Kluft zwischen Akademiker*innen und Industrie zu überbrücken, insbesondere da, wo die Datenkultur noch im Entstehen begriffen ist (*Pig Data*). Vertrauenswürdige staatliche Plattformen für den Austausch von Crowdsourced-Daten die für die Öffentlichkeit von Interesse sind, einrichten und sicherstellen, dass diese die erforderlichen Metadaten enthalten (*Hochwassererkennung*).

- Endnutzer*innen frühzeitig einbeziehen, um ihre Bedürfnisse zu verstehen und praktische Tools zur Interpretation und Visualisierung von Daten zu entwickeln (*Hochwassererkennung*).
- In unüberwachtes Lernen und andere Ansätze investieren, die Abhängigkeiten von potenziell verzerrten Trainingsdaten verringern (*Hochwassererkennung*). Überwachtes maschinelles Lernen entwickeln, wenn angemessene Trainingsdaten verfügbar sind (*Bodenerosion*).

2.3

Besseres Verständnis sozioöko- nomischer Wechsel- wirkungen

Der Einsatz von ausgefeilten Datenanalyseansätzen wird zunehmend für soziale und politische Anwendungen erforscht. Die erforderlichen Daten sind jedoch nicht immer ohne Weiteres verfügbar. Und wenn doch, kann die Analyse komplizierter sein als in den Natur- oder technischen Wissenschaften: Ursachen und Wirkungen sind in der Regel schwer aus den Daten zu entschlüsseln, weil sie vielschichtig und in unterschiedliche sozio-politische Rahmenbedingungen eingebettet sind. Die Tatsache, dass immer mehr Daten gesammelt werden, bietet jedoch viele Möglichkeiten für neue Anwendungen, die dabei helfen, politische Massnahmen zu evaluieren, zu orientieren und möglicherweise zu verbessern.

Konkrete Anwendungen in der Sozioökonomie wurden im Rahmen des NFP 75 zu folgenden Zwecken entwickelt:

- Zur Verbesserung der Methoden zur Ermittlung kausaler Zusammenhänge in sozioökonomischen Datenansätzen, um die Politikgestaltung zu bereichern und zu verbessern (*Evidenzbasierte Politik*).
- Zur Erweiterung der Analysemethoden zur Untersuchung der Patentproduktion und von Patenzitaten, um Innovation und Wissensgenerierung weltweit abzubilden (*Mapping globaler Innovation*).

Kernaussagen zu sozioökonomischen Anwendungen

Gesellschaftliche Auswirkungen

Die zunehmende Verfügbarkeit sozioökonomischer Daten bietet grosse Chancen für eine bessere, evidenzbasierte Entscheidungs- und Politikgestaltung. Analysen sind jedoch oft schwieriger zu erstellen, als man denkt. Das Projekt *Evidenzbasierte Politik* hat Techniken entwickelt, die auf maschinellem Lernen basieren, um kausale Effekte von blossen Korrelationen zu trennen. Big-Data-Techniken können auch die weltweite Verbreitung von Ideen und Einflüssen abbilden, was zu einem besseren Verständnis dessen beitragen sollte, was zu erfolgreicher Innovation führt (*Mapping globaler Innovation*).

Auswirkungen auf Big-Data- Kapazitäten in der Sozioökonomie

Der Einsatz von Algorithmen zur Analyse politisch relevanter Daten kann dazu beitragen, Verzerrungen zu verringern, zum Beispiel wenn Forschende unbewusst zu Erkenntnissen tendieren, die ihren eigenen Ansichten entsprechen. Mit diesem Ansatz lässt sich auch feststellen, ob bestimmte Untergruppen von Menschen anders auf politische Massnahmen reagieren als die Durchschnittsbevölkerung, was detailliertere Erkenntnisse und politische Massnahmen ermöglicht. Solche Techniken stehen jetzt bereit, um in der Praxis angewendet zu werden (*Evidenzbasierte Politik*). Neue Methoden können zur Analyse grosser Netzwerke eingesetzt werden. Sie sind in der Lage, zeitliche Effekte zu berücksichtigen, und helfen dabei, Knotenpunkte zu identifizieren, die schnelle oder dauerhafte Auswirkungen haben (*Mapping globaler Innovation*).

Herausforderungen

Selbst sorgfältige Datenanalysen können durch die eingesetzten Methoden verzerrt werden. Korrelationen können verborgen bleiben, weil die Daten zu granulär sind (*Evidenzbasierte Politik*). Ausserdem haben die Datensätze inhärente Beschränkungen – so zum Beispiel durch die Tatsache, dass Patentzitate nicht über verschiedene nationale Register hinweg verfolgt werden können (*Mapping globaler Innovation*).

Lösungsansätze

- Die Förderung des Einsatzes hochentwickelter Analysen auf der Grundlage sozioökonomischer Daten bei der Bewertung politischer Massnahmen (*Evidenzbasierte Politik*)
- Die Förderung der Datenaggregation auf verschiedenen Granularitätsebenen, z. B. auf der Ebene der Bevölkerung und der von Teilpopulationen, und die Förderung der Kausalanalyse, um aus den Daten fundierte Beziehungen abzuleiten (*Evidenzbasierte Politik*)
- Die Förderung der internationalen Kompatibilität nationaler Datensätze (*Mapping globaler Innovation*)

-weitergabe festgelegt. Der zunehmende Umfang der Datensätze – die bis zu mehreren Petabytes gross sein können – erschwert den Zugriff, die Speicherung, die Manipulation und die Analyse der betreffenden Daten. Dies führt zu einem zunehmenden Bedarf an effizienten Analysemethoden und Algorithmen maschinellen Lernens.

Das NFP 75 hat neue Methoden und Werkzeuge entwickelt, um den Umgang mit und die Analyse von umfangreichen Forschungsdatenbanken in Chemie, Sonnenphysik und Genetik zu verbessern:

- Die Entdeckung neuer Moleküle wird durch eine effiziente Simulation ihrer Eigenschaften beschleunigt (*Computergestützte Chemie*).
- Das Verständnis und die Vorhersage von Sonneneruptionen, die geomagnetische Stürme auf der Erde auslösen können, wird verbessert (*Sonneneruptionen*).
- Die Analyse von genomischen und biologischen Daten in der biomedizinischen Forschung wird leichter und unterstützt so die Entwicklung neuer therapeutischer und diagnostischer Ansätze (*Big Data Genetik, Genomvergleich, Bioinformatikdatenbanken*).

2.4 Beschleunigung der Forschung

Big Data fördert auch Innovation, indem es in der Grundlagenforschung eine immer wichtigere Rolle spielt. Dies zeigt sich vor allem in sehr grossen internationalen Kooperationen wie dem CERN in Genf, in Astronomieprojekten, der Erdbeobachtung mittels Satelliten oder der Genomik. Immer mehr wissenschaftliche Disziplinen haben Standards zur Datenerhebung und

Kernaussagen zu Forschungsanwendungen

Gesellschaftliche Auswirkungen

Die Entwicklung der richtigen Analysewerkzeuge, insbesondere im Bereich des maschinellen Lernens, erlaubt es, immer grössere Forschungsdatensätze zu nutzen. Dies ermöglicht schnellere und umfassendere Entdeckungen sowie genauere Vorhersagen in den Biowissenschaften, der Chemie und der Weltraumphysik.

Auswirkungen auf Big-Data-Kapazitäten in der Forschung

Selbst mit sehr grossen Datensätzen, die komprimiert wurden, um den Zeitaufwand und die Kosten für die Datenübertragung, -speicherung und -verarbeitung zu reduzieren, ist es möglich, hohe Vorhersagegenauigkeiten zu erreichen (*Sonneneruptionen, Big Data Genetik*). Techniken des maschinellen Lernens können die Zeit, die für die genaue Lösung sehr komplexer mathematischer Probleme benötigt wird, wie bei der Identifizierung der Eigenschaften von Molekülen, drastisch reduzieren (*Computergestützte Chemie*). Mehrere NFP-75-Projekte haben Tools zur Verwaltung und Durchsuchung riesiger biologischer Datensätze im Petabyte-Bereich (Millionen Gigabyte) entwickelt (*Big Data Genetik, Bioinformatikdatenbanken*).

Herausforderungen

In Forschungsprojekten stehen oft keine klassifizierte Daten zur Verfügung, was die Verwendung von Standardtools für überwachtes maschinelles Lernen einschränkt (*Sonneneruptionen*). Gleichzeitig sind die Metadaten – Informationen über die Daten selbst – oft von geringer Qualität, was eine effiziente Indexierung und Analyse des Datenzugriffs verhindert (*Big Data Genetik*).

Lösungsansätze

- Die Wissenschaft und die Industrie dazu ermutigen, grosse Forschungsdatensätze im Rahmen des Paradigmas der «Open Research Data» gemeinsam zu nutzen und zu pflegen (siehe «Die Hürden zu offenen Forschungsdaten», S. 51).
- Reale Datensätze nutzen, um entwickelte Methoden frühzeitig zu evaluieren, so dass der Forschungsumfang bei Bedarf angepasst werden kann.
- In erstklassigen Projekten der Grundlagenforschung und der angewandten

Forschung, die grosse Datensätze nutzen, investieren, um Big-Data-Fähigkeiten im akademischen und im privaten Umfeld zu fördern.

2.5 Kernaussagen zu Big-Data- Anwendungen

Vorsicht vor dem Hype

Die Berichterstattung in den Medien, durch Industrie und Think-Tanks kann den Eindruck erwecken, dass Big Data ein magisches Werkzeug sei: Finden Sie einfach die Daten, fügen Sie ein wenig maschinelles Lernen hinzu, trainieren Sie die Algorithmen, erstellen Sie eine Anwendung und schon sind Sie bereit, Geschäftspraktiken und ganze Branchen völlig zu verändern. Diese allzu simple Sichtweise übersieht zahlreiche Hürden konzeptioneller, technischer, rechtlicher, organisatorischer und kollaborativer Art.

Die Entwicklung einer Big-Data-Anwendung erfordert einen immensen Aufwand in vielen Schritten: den Aufbau von Partnerschaften mit Interessengruppen; das Recherchieren und Identifizieren geeigneter Datenquellen; das Bewerten der Datenqualität, ihre sichere Speicherung; die Vorbereitung der Daten für die Analyse; geeignete Algorithmen finden und anpassen oder neue entwickeln; ihre Performanz und Zuverlässigkeit prüfen, Schnittstellen zur Verwendung der Ergebnisse schaffen; und schliesslich die neue Anwendung in bestehende Arbeitsabläufe integrieren.

Immer wenn Big Data übermässig gefeiert wird, werden sehr tiefgehende, wenn auch scheinbar offensichtliche Fragen verschleiert. Existieren die Daten überhaupt? Sind sie zugänglich? Sind sie mit Metadaten richtig beschrieben? Können die Privatsphäre gewahrt und die Vorschriften eingehalten werden? Brauchen die Endnutzer*innen die beabsichtigte Anwendung tatsächlich? Solche Fragen müssen von Anfang an berücksichtigt werden, um den Umfang der anstehenden Arbeit realistisch einschätzen zu können.

Ist Ihr Bereich reif für Big Data?

Big-Data-Anwendungen können relativ effizient entwickelt werden, wenn sie auf bestehenden Prototypen aufbauen (*Potenzial erneuerbarer Energien, Optimierung der Verkehrssysteme*) oder wenn qualitativ hochwertige und standardisierte Daten vorliegen, wie in den Bereichen Wetter, Kartographie und Biologie (*Sonneneruptionen, Big Data Genetik*). In den Fällen können sich Datenwissenschaftler*innen hauptsächlich auf technische Fragen konzentrieren, z. B. auf den Aufbau einer Pipeline für den Datenzugriff in Echtzeit, die Entwicklung von Algorithmen oder auf die Schaffung benutzerfreundlicher interaktiver Schnittstellen. Die Projekte des NFP 75 erbrachten eine Reihe wichtiger Fortschritte bei der Konzeption, Umsetzung und Bewertung praktischer Ansätze für das Data Engineering, einschliesslich Datenmanagement, Analytik, Visualisierung, Bewertung, Prüfung, Integration und Mining.

Umgekehrt ist es viel schwieriger, Big-Data-Anwendungen in Bereichen zu entwickeln, die weniger digitalisiert sind, denen eine Datenkultur fehlt und die dem Austausch von Daten

abgeneigt oder darin ungeübt sind (*Pig Data, Mapping globaler Innovation*). In solchen Fällen ist ein erheblicher Zeit- und Arbeitsaufwand erforderlich, um nicht-technische Hürden zu überwinden, z. B. den Aufbau von Partnerschaften zwischen Akteur*innen, die nicht an eine gemeinsame Nutzung von Daten gewöhnt sind. Die Verfügbarkeit und die Qualität der Daten müssen frühzeitig bewertet werden, was möglicherweise zu einem Überdenken des Anwendungsbereichs führt.

Big Data verlangt nach Interdisziplinarität

Die Verwaltung von Datensätzen im Petabyte-Massstab erfordert viel Zeit, Arbeitskraft und gemeinsame Anstrengungen, um zahlreiche technische und sozialrechtliche Herausforderungen zu lösen. Der Aufbau einer Anwendung, die etwas bewirkt, erfordert in der Regel einen interdisziplinären Ansatz, bei dem sich die verschiedenen Interessengruppen frühzeitig mit allen potenziellen Problemen und Fragen befassen, einschliesslich der Frage, wie die geplante Lösung von den Endnutzer*innen verwendet werden soll (*Pig Data, Hochwassererkennung*). Wissenschaftler*innen, Expert*innen aus den entsprechenden Fachgebieten und Partner*innen aus Industrie oder Verwaltung müssen zusammenarbeiten, um sicherzustellen, dass qualitativ hochwertige Daten erfasst und gemeinsam genutzt werden und dass die Anwendungen den Bedürfnissen der vorgesehenen Nutzer*innen entsprechen. Die Forschenden des NFP 75 haben neue Wege einer effektiven Interaktion mit verschiedenen Interessengruppen erforscht. Die frühzeitige Einbeziehung von Nutzer*innen verbessert das Design von Anwendungen (*Hochwassererkennung*). Diese Art von Erfahrung stellt sicher, dass die

Erkenntnisse über Big Data hinaus

Einige NFP-75-Forschungsprojekte brachten Ergebnisse hervor, die von direkter gesellschaftlicher Relevanz sind, insbesondere in den Bereichen Nachhaltigkeit und Sozioökonomie. Sie veranschaulichen das Potenzial von Big Data zur Unterstützung der Politik

Das Projekt *Potenzial erneuerbarer Energien* lieferte konkrete Ergebnisse, welche die **Energiestrategie 2050 des Bundesrates** unterstützen können.

→ Es ist möglich, 50 Prozent des landesweiten Potenzials an Solarstrom mit nur 10 Prozent der Fläche bestehender Dächer zu erreichen, wenn man sich auf die Dächer mit dem höchsten Potenzial konzentriert. Dies entspricht etwa 12 TWh Strom pro Jahr oder etwa 20 Prozent des nationalen Verbrauchs.

→ Etwa 1000 Windturbinen könnten 4 TWh erzeugen: das vom Bundesamt für Energie gesetzte Ziel für die Windkraft im Jahr 2050.

→ Hochauflösende Karten des oberflächennahen geothermischen Potenzials in den Kantonen Waadt und Genf deuten auf eine potenzielle Wärmeproduktion von 4 TWh hin, was 40 Prozent des derzeitigen Bedarfs in den beiden Kantonen entspricht. Der Einsatz von Fernwärmenetzen zur Verteilung der Wärme auf die Siedlungen würde dieses Wärmepotenzial verdoppeln.

→ Die Wiedereinleitung von Wärme in den Boden während des Sommers ist wichtig für eine nachhaltige geothermische Energienutzung. In Kombination mit Fernwärmenetzen kann die oberflächennahe Geothermie bis 2050 mehr als 70 Prozent des Heiz- und Kühlbedarfs von Schweizer Gebäuden decken.

Das Projekt *Bodenerosion* brachte neue Erkenntnisse über **Erosionen in den Alpenregionen**.

→ Von 2007 bis 2016 wurde in einem 2000 km² grossen Gebiet um Martigny (VS), was 11 Prozent der Schweizer Alpen entspricht, eine Zunahme der Bodenerosion ausgesetzter Flächen um 80 Prozent festgestellt.

Das Projekt *Evidenzbasierte Politik* ermittelte kausale Effekte in **sozioökonomischen Zusammenhängen**.

→ Schulungen Arbeitssuchender führen im Durchschnitt nicht dazu, dass sie schneller eine Anstellung finden. Sie können jedoch die Chancen, einen Arbeitsplatz zu finden, von Untergruppen wie Migrant*innen um bis zu 60 Prozent erhöhen.

→ Musizieren wirkt sich positiv auf die kognitive und die nicht kognitive Entwicklung von Kindern aus.

→ Fussballschiedsrichter*innen bestrafen eher Teams aus bestimmten Sprachregionen in der Schweiz.

Das Projekt *Mapping globaler Innovation* hat durch die Analyse von Millionen von Patenten **Erkenntnisse über Innovation** gewonnen.

→ Patent-Ökosysteme sind weniger international als erwartet. Patentzitate sind oft geografisch und disziplinär konzentriert. Interdisziplinäre Patente scheinen indes nicht erfolgreicher zu sein. Patente aus einigen Ländern werden besonders häufig zitiert – die Schweiz zum Beispiel ist bekannt als ein fruchtbares Zentrum der Innovation.

→ Unternehmen und Organisationen spielen bei der Art und Weise, wie Patente untereinander zitiert werden, unterschiedliche Rollen. Einige von ihnen fungieren als zentrale Drehscheiben und Wissensvermittler.

Siehe Abschnitt 2.6

akademische Forschung bei Bedarf in der Lage ist, schnell praktische Anwendungen zu entwickeln.

Die Welt ist unordentlicher als Trainingsdaten

Algorithmen maschinellen Lernens können mit potenziell schwerwiegenden Folgen versagen, wenn es beispielsweise um Gesundheit oder autonomes Fahren geht. Dies ist ein Problem des überwachten Lernens, wenn die Algorithmen mit

unvollständigen, ungenauen oder zu homogenen Trainingsdaten gefüttert werden. Eine Hinwendung zu sparsamerem unüberwachtem Lernen könnte zu robusteren Systemen führen (*Hochwassererkennung*).

Von Anfang an auf den Datenschutz achten

Fragen im Zusammenhang mit Datenschutz und Regulierung müssen sorgfältig behandelt werden, insbesondere bedürfen sie der Hilfe von Rechtsexpert*innen (siehe auch

Kapitel 4). Privacy-by-Design-Ansätze sollten so früh wie möglich in Betracht gezogen und umgesetzt werden. Dies erfordert eine sorgfältige Berücksichtigung übergreifender Prinzipien wie Zweckbindung, Transparenz und Verhältnismässigkeit, sowie Datenminimierung, Datengenauigkeit und Datenschutz (*Intensivstationen, Digitale Zwillingstädte*). Der Austausch von Erfahrungen innerhalb und zwischen verschiedenen Bereichen hilft bei der Entwicklung von Best Practices.

2.6 Die Forschungsprojekte zu Big-Data-Anwendungen

Das NFP 75 brachte Anwendungen auf der Grundlage von Big Data in verschiedenen Bereichen hervor: zwei im Gesundheitsbereich, sechs im Bereich Nachhaltigkeit, zwei in der Sozioökonomie und fünf im Bereich Forschungsressourcen.

Forschungsprojekte zu Anwendungen im Gesundheitsbereich

Intensivstationen: ein automatisiertes Alarmsystem

In diesem Projekt wurde ein «ICU Cockpit» entwickelt, das dem Personal auf neurochirurgischen Intensivstationen hilft, in kritischen Situationen schnell zu reagieren. Dabei geht es insbesondere um ischämische (Hirn-) Schlaganfälle und epileptische Anfälle. Die Plattform hilft dem medizinischen

Personal, bei den Interventionen Prioritäten zu setzen und damit die Sicherheit der Patient*innen zu erhöhen. Es handelt sich um eine Zusammenarbeit zwischen der Neurochirurgischen Intensivstation des Universitätsspitals Zürich, der ETH Zürich und IBM Research Zurich.

Das System antizipiert kritische Situationen, indem es zahlreiche Arten von Daten integriert und analysiert, u.a. aus der Elektroenzephalographie (EEG), aus Videostreams und der Krankengeschichte der Patient*innen, einschliesslich bildgebender Untersuchungen des Gehirns und Laboranalysen. Das Team entwickelte Technologien zur Erfassung biomedizinischer Daten in Echtzeit mit einer hohen Auflösung von bis zu 200 Hertz. Es entwarf und implementierte Algorithmen, die epileptische Anfälle auf der Grundlage von Video- und Elektroenzephalographie automatisch erkennen und drohende sekundäre Hirnverletzungen vorhersagen können. Die Anwendung basierte auf den Daten von mehr als 100 Patient*innen mit einer Art von Schlaganfall, der sogenannten Subarachnoidalblutung. Es wurden zwei Algorithmen entwickelt, um die Rate der Fehlalarme zu reduzieren. Der eine basiert auf maschinellem Lernen, der andere auf der Videoüberwachung der Bewegungen der Patient*innen. Das System wurde integriert und in einer klinischen Umgebung getestet.

—
ICU-Cockpit: IT platform for multimodal patient monitoring and therapy support in intensive care and emergency medicine
Emanuela Keller (Universitätsspital Zürich)

Rückenschmerzen: eine personalisierte Smartphone-basierte Lösung

In diesem Projekt wurde ein Smartphone-basierter Ansatz für die Behandlung von Rückenschmerzen entwickelt. Das Team entwickelte die App «Swiss Health Challenge», um anonymisiert

Sensordaten zu sammeln, zu übertragen und zu speichern. Darüber hinaus bewertete es drei verschiedene Präventionsstrategien, um die hohen Behandlungskosten (die in der Regel aus Schmerzmitteln, Physiotherapie und Operationen bestehen) zu senken. Forschende der ETH Zürich entwickelten im Rahmen einer Zusammenarbeit mit der Universitätsklinik Balgrist und dem Schweizer Medizintechnikunternehmen Hocoma Methoden maschinellen Lernens, um die Daten zu analysieren. Im Rahmen eines präventiven Ansatzes wurden Patient*innen, die unter Schmerzen im unteren Rückenbereich litten, regelmässig aufgefordert, zu Hause interaktive Übungen durchzuführen. Während die regelmässige Beurteilung durch Physiotherapeut*innen keine Verbesserungen bei leicht erkrankten Patient*innen ergab, bot die Studie neue Erkenntnisse zur Einhaltung von Bewegungsprogrammen durch Patient*innen und zur Frage, wie sich die Angst vor Bewegung auf das posturale Schwanken auswirken kann (d.h. die kleinen unbewussten Bewegungen, die das Gleichgewicht erhalten).

Die App wurde durch Bewegungssensoren ergänzt, um körperliche Belastungen beim Skittraining zu erkennen, die zu Rückenverletzungen führen können. Dieser Ansatz liefert weitaus detailliertere Informationen über das Bewegungsverhalten als Selbstbeobachtungen. Das Projekt hat auch gezeigt, wie wichtig die Motivation der Patient*innen ist, wenn sie sich an ein Trainingsprogramm halten.

—
Personalized management of low back pain with mHealth: big data opportunities, challenges and solutions

Robert Riener (ETH Zürich) und Walter Karlen (Universität Ulm, davor ETH Zürich)

Forschungsprojekte im Bereich Nachhaltigkeit

Potenzial erneuerbarer Energien: Abschätzung für die Schweiz

Dieses Projekt schuf eine digitale Plattform zur Einschätzung des Potenzials für Erdwärme, Wind und Photovoltaik zur Heizung und Kühlung von Gebäuden. Die daraus resultierenden landesweiten Schätzungen haben eine hohe räumliche und zeitliche Auflösung und können dazu beitragen, Energiesysteme auf lokaler und regionaler Ebene zu planen, Anreize zu optimieren und die nationale Energiestrategie anzupassen. Das System integriert Daten über das Wetter (Wind und Sonneneinstrahlung), die Umwelt (Topographie, Geologie und Bodentemperaturen) und die bebaute Umwelt (Dachausrichtung und der für Bohrungen verfügbare Platz). Es nutzt diese Daten, um regionale und nationale Karten des Potenzials für erneuerbare Energien zu erstellen, mit einer räumlichen Auflösung auf der Ebene einzelner Gebäude und einer zeitlichen Auflösung von etwa einer Stunde. Mittels neuer Mining-Techniken für Big Data konnten die verfügbaren Messpunkte interpoliert werden, um Lücken in den Karten zu füllen. So wurden zum Beispiel Windkarten mit einer Auflösung von 250 Metern für das ganze Land erstellt, die nur auf den Daten von 208 MeteoSchweiz-Messstationen basierten. Das Team quantifizierte auch die Unsicherheit der erstellten Karten. Das Projekt lieferte zahlreiche Ergebnisse, die für die Energiepolitik relevant sind (siehe «Erkenntnisse über Big Data hinaus», S. 35).

—
Hybrid renewable energy potential for the built environment using big data: forecasting and uncertainty estimation
Jean-Louis Scartezini (EPFL)

Optimierung der Verkehrssysteme: anonymisierte individuelle Mobilitätsdaten

Dieses Projekt untersuchte die Möglichkeiten, GPS-Daten von Smartphones zu sammeln und zu analysieren, um Erkenntnisse über Mobilitätsmuster der Menschen zu gewinnen. Mehr als 4000 Teilnehmende installierten eine spezielle App, die das Aufzeichnen von mehr als einer Million Fahrten ermöglichte. Die beteiligten Wissenschaftler*innen entwickelten neue Verfahren, um die Daten zu anonymisieren und die Art der Fortbewegung (zu Fuss, mit dem Velo, dem Bus, dem Auto usw.) sowie die Aktivität (Sport, Arbeit, Ausbildung usw.) zu erkennen. Die Ergebnisse des Projekts wurden in anderen grossen Mobilitätsstudien in der Schweiz verwendet, z. B. zur Untersuchung der Genauigkeit von Selbsterfassungen (Mikrozensus Mobilität und Verkehr in der Schweiz) oder der Reaktion auf die Mobilitätspreisgestaltung (Mobilitätsverhalten in der Schweiz). Die Ergebnisse wurden auch verwendet, um verschiedene Szenarien, wie die Änderung von Fahrplänen des öffentlichen Verkehrs oder die Einführung einer neuen Verkehrsplanung, in einer Mobilitätssimulationsplattform namens Matsim zu erstellen und zu bewerten. Sie zeigten zum Beispiel, dass während der Covid-19-Pandemie die Nutzung der öffentlichen Verkehrsmittel in Zürich stark zurückging und die Nutzung von Velos deutlich zunahm. Das Projekt wurde in Zusammenarbeit mit dem Telekommunikationsanbieter Swisscom durchgeführt. Es erschloss das Potenzial von GSM-Spuren für die Verkehrsmodellierung und die Optimierung der damit verbundenen Massnahmen. Es zeigte, dass individuelle – aber anonyme – Mobilitätsmuster aus aggregierten GSM-Traces extrahiert werden können, sodass die Notwendigkeit entfällt, aktiv Daten von Einzelpersonen zu sammeln und die

Privatsphäre zu gefährden.

–

Big data transport models: the example of road pricing

Kay W. Axhausen (ETH Zürich)

Digitale Zwillingstädte: 3D-Modelle aus einem scannenden Auto

Im Rahmen dieses Projekts wurde ein auf einem Auto montiertes Kamerasystem entwickelt, das städtische Gebiete scannt. Es entstanden eine Kamera, die in alle Richtungen aufnimmt, und ein Algorithmus, der aus diesen kontinuierlichen Scans ein vollständiges numerisches 3D-Modell einer Stadt erstellt. Solche virtuellen «digitalen Zwillinge» von Stadtgebieten können die Stadtplanung und Verkehrsstrategien unterstützen.

Das Projekt untersuchte auch rechtliche und ethische Fragen und entwickelte einen Ansatz zum Schutz der Privatsphäre, der auf den Grundsätzen von Zweckbindung, Transparenz und Verhältnismässigkeit sowie von Datenminimierung, -genauigkeit und -sicherheit beruht. So wurden beispielsweise bei Tests in Sitten (VS) automatisch sensible Details wie Autokennzeichen und Gesichter entfernt. Das Team hat den Technologietransfer mit einem Schweizer Startup-Unternehmen besprochen, das Digitalisierungsdienste für Städte anbietet.

–

ScanVan – a distributed 3d digitalization platform for cities

Frédéric Kaplan (EPFL)

Bodenerosion: Quantifizierung durch Luftaufnahmen in der Schweiz

Im Rahmen dieses Projekts wurden Algorithmen maschinellen Lernens entwickelt, die es ermöglichen, erodierete Böden in offiziellen Luftaufnahmen automatisiert zu erkennen und zu kartieren. Es untersuchte zehn Standorte, hauptsächlich in Bergregionen, und zeigte zum Beispiel auf, dass sich die

erodierten Flächen im Urserental zwischen Realp und Hospental (UR) innerhalb von 16 Jahren auf 0,4 Quadratkilometer fast verdreifacht haben. Es wurden ausserdem die wichtigsten Faktoren ermittelt, die Erosion und Erdbeben beeinflussen, nämlich die Neigung, die Beschaffenheit und die Ausrichtung eines Geländes. Die automatisierte Analyse wurde verwendet, um eine detailliertere Studie über eine 2000 Quadratkilometer grosse Region um Martigny (VS) durchzuführen, die ein Zehntel der Schweizer Alpen umfasst. Es wurde geschätzt, dass die degradierte Fläche zwischen 2007 und 2016 um 80 Prozent zugenommen hat (siehe «Erkenntnisse über Big Data hinaus», S. 35).

Diese Ergebnisse machen die Erosion für politische Entscheidungsträger*innen sichtbar und können in Massnahmen zum Schutz des Bodens in der Landwirtschaft, im Tourismus und in der Flächennutzungsplanung einfließen. Dies ist von entscheidender Bedeutung, da der Boden eine nicht erneuerbare Ressource ist, die für die Nahrungsmittelproduktion, die Artenvielfalt und den Umgang mit Naturgefahren unerlässlich ist. Im Rahmen eines vom Bundesamt für Umwelt finanzierten Anschluss-Forschungsprojekts wird ein Instrument zur Kartierung der Erosion in grossem Massstab entwickelt.

—
WeObserve: integrating citizen observers and high throughput sensing devices for big data collection, integration, and analysis
Volker Roth (Universität Basel)

Hochwassererkennung: automatisches Geotagging von Crowdsourced-Videos

In diesem Projekt wurden erste Bausteine für eine Plattform zur Unterstützung des Managements von Hochwasserrisiken entwickelt. Es umfasst die automatische Videoanalyse zur Erkennung von Krisensituationen, die

Ermittlung des Standorts eines Videos und die Präsentation der Ergebnisse für Krisenmanager.

Das Projekt machte deutlich, dass neue unüberwachte Algorithmen für maschinelles Lernen erforderlich sind. Die existierende Bilderkennungstechnik, die auf überwachtem Lernen basiert und mit annotierten Datensätzen trainiert wird, kann bei der Analyse von Videos aus der alltäglichen Anwendung schnell versagen.

In einer Studie, die im Kanton Basel-Landschaft durchgeführt wurde, testete das Team, wie Videos automatisiert lokalisiert werden können, die z. B. mit einem Smartphone ohne GPS-Informationen aufgenommen wurden. Dies geschah durch den Vergleich der Videos mit vorhandenen Bildern auf Strassenebene. Doch diese Strategie hing von mehreren Faktoren ab: von ausreichend hochwertigen Quellen, vom richtigen Wetter und vom Vorhandensein erkennbarer Infrastrukturen oder Landmarken. In Zusammenarbeit mit Expert*innen, die sich mit Naturgefahren befassen, testeten die Forschenden Prototypen interaktiver Karten zur Visualisierung von Überschwemmungsszenen. Das Projekt hat mehrere Meilensteine bei der Nutzung von KI zur automatischen Erkennung von Krisensituationen erreicht.

—
EVAC – Employing video analytics for crisis management
Susanne Bleisch (FHNW)

Pig Data: Analytik für die Schweizer Schweinehaltung

Dieses Projekt brachte Akteur*innen der Schweizer Schweineindustrie zusammen, um Informationen zu einer Vielzahl von Themen wie Schweinetransport, Abrechnung, Gesundheit und Behandlung, Fleischqualität und Mastprozesse zu integrieren. Es entwickelte Methoden für Echtzeit-Vorhersagen. So wurde beispielsweise bestätigt, dass

die meisten Landwirt*innen Schlachtkörper liefern, die den vorgeschriebenen Qualitätsstandards entsprechen, während gleichzeitig Bereiche ermittelt wurden, in denen die Betriebe die Spezifikationen der Branche nicht erfüllen. Das Projekt lieferte Antworten auf 6 von 18 Fragen, die durch Interessenvertreter*innen gestellt wurden, z. B. was die Qualität der Schlachtkörper beeinflusst, wie das Netzwerk der Schweinezuchtindustrie strukturiert ist oder worin der Unterschied in Qualität und Ertrag zwischen gleichen Zuchtherden und gemischten Beständen besteht. Das Projekt hat wichtigen Interessenvertreter*innen, wie dem Schweizer Bundesamt für Lebensmittelsicherheit und Veterinärwesen, gezeigt, wie wichtig die Digitalisierung der Zuchtindustrie ist, und damit zur Schaffung eines Schweizer Kompetenz- und Informationszentrums für Schweinegesundheit beigetragen.

Das Projekt veranschaulicht, wie schwierig die Durchführung von Analysen in einer Branche mit vielen kleinen Akteur*innen ist, die es nicht gewohnt sind, mit Daten zu arbeiten – im Vergleich zu anderen Bereichen oder zur Schweinehaltung in anderen Ländern. Zu den besonderen Herausforderungen gehören Qualität und Kompatibilität der Daten sowie die mangelnde Bereitschaft, Informationen auszutauschen.

–
Pig data: health analytics for the Swiss swine industry
John Berezowski (Universität Bern)

Forschungsprojekte zur Sozioökonomie

Evidenzbasierte Politik: Kausalität aus Daten entdecken

Dieses Projekt hat dazu beigetragen, Ursache-Wirkungs-Beziehungen in grossen sozioökonomischen Datensätzen zu identifizieren, indem neue

Techniken maschinellen Lernens entwickelt wurden, um Kausalität im Gegensatz zu blossen Korrelationen nachzuweisen. Konkret wurden Ursache-Wirkungs-Beziehungen in realen Datensätzen zu Arbeitslosenunterstützung, Bildung und Sport aufgedeckt (siehe «Erkenntnisse über Big Data hinaus», S. 35).

Die neuen Methoden können Untergruppen aufdecken, die von Massnahmen profitieren, selbst wenn die Gesamtbevölkerung davon unberührt bleibt – Erkenntnisse, die in gemittelten Daten verborgen bleiben würden und die politische Entscheidungen möglicherweise stärker an personalisierten Massnahmen ausrichten könnten. Bei diesem Projekt wurden bedeutende methodologische Fortschritte erzielt, beispielsweise das Vergleichen und Evaluieren bestehender statistischer Ansätze.

Sozioökonomische Datensätze werden immer komplexer, was zu grösseren Erkenntnissen, aber auch zu schwierigeren Analysen führen kann. Im Rahmen des Projekts wurden Ansätze maschinellen Lernens entwickelt, um eine zentrale Frage der Modellierung damit beantworten zu können: Welche Faktoren sollten explizit analysiert werden und welche sollten stattdessen als Störfaktoren betrachtet werden? Solche Fragen sind von zentraler Bedeutung für die Entwicklung einer evidenzbasierten Politik und werden in einem Folgeprojekt untersucht, das vom Nationalen Forschungsprogramm «Digitale Transformation» (NFP 77) finanziert wird.

–
Causal analysis with big data
Michael Lechner (Universität St. Gallen)

Mapping globaler Innovation: Analyse von Patenten

In diesem Projekt wurden Millionen von Patenten analysiert, um herauszufinden, wie sich Innovationen weltweit

verbreiten. Es führte mehrere Datenbanken zusammen, die Informationen über Patente und Unternehmen weltweit enthalten, und erstellte ein Modell, um zeitliche Veränderungen und deren Auswirkungen in komplexen Netzwerken zu erfassen. So half es, folgenreiche Patente zu identifizieren und damit Einblicke zur globalen Innovation zu gewinnen (siehe «Erkenntnisse über Big Data hinaus», S. 35). Das Team passte ein statistisches Modell für Big-Data-Netzwerke mit über einer Million Knoten an und stellte es als Open-Source-Software zur Verfügung.

—

The global structure of knowledge networks: data, models and empirical results
Alessandro Lomi (Università della Svizzera italiana)

Forschungsprojekte zu Forschungsressourcen

Computergestützte Chemie: die Entdeckung neuer Moleküle

Dieses Projekt verbesserte Methoden künstlicher Intelligenz, um Moleküle zu simulieren: ein Sprungbrett, um die Entdeckung nützlicher Verbindungen für Gesundheit und Industrie zu beschleunigen. Es wurden neuartige Methoden maschinellen Lernens entwickelt, um die wichtigsten Eigenschaften von Molekülen zu berechnen, wobei bestehende Datenbanken mit Informationen über Hunderte von Milliarden von Molekülen verwendet wurden. Es ist sehr schwierig, solche riesigen Datenbanken effizient zu analysieren, aber das Team erstellte Trainings- und Testdatensätze und war in der Lage, Moleküle im Vergleich zu bestehenden Methoden viel schneller und genauer zu charakterisieren, selbst für grössere Verbindungen.

—

Big data for computational chemistry: unified machine learning and sparse grid

combination technique for quantum based molecular design

Helmut Harbrecht (Universität Basel)

Sonneneruptionen: geomagnetische Stürme vorhersagen

Dieses Projekt zeigte, wie Sonneneruptionen vorhergesagt werden können, um sich auf mögliche geomagnetische Stürme auf der Erde vorzubereiten. Diese Stürme können kritische Infrastrukturen wie Telekommunikation, Stromnetze, Satellitenbetrieb und Flugrouten stören. Das Team nutzte 30 Terabyte an Sonnenbeobachtungen, entwickelte Methoden zur Analyse der Daten in komprimierter Form und zeigte, dass Signale im ultravioletten Wellenlängenbereich, die derzeit kaum genutzt werden, helfen können, die Entstehung von Sonneneruptionen vorherzusagen. Der Algorithmus konnte eine Sonneneruption eine halbe Stunde vor ihrem Auftreten voraussagen. Das Team entwickelte sowohl überwachte Algorithmen (mit Trainingsdaten) als auch unüberwachte Methoden, um das Problem der begrenzten Anzahl von klassifizierten Daten zu überwinden.

—

Machine learning based analytics for big data in astronomy

Svyatoslav Voloshynovskiy (Universität Genf)

Big Data Genetik: leistungsstarke Indexierung

In diesem Projekt wurde eine neue Methode entwickelt, um sehr grosse Datenbanken mit genetischen Sequenzen zu indexieren und gleichzeitig tausendfach zu komprimieren. Das öffentliche Tool, Metagraph, kann Nutzenden auf der ganzen Welt dabei helfen, effizient Bestände von genetischen Sequenzen zu durchsuchen und umfassende Analysen durchzuführen. Der Ansatz findet zahlreiche Anwendungen in der personalisierten Medizin oder bei der Suche nach Varianten von

Krankheitserregern. Er hilft dabei, die immer grösser werdenden genetischen Datenbanken wie den Krebsgenom-Atlas zu bewältigen, die oft Petabytes (Millionen von Gigabytes) an Daten enthalten. Das Projekt indexierte über 1,4 Millionen ganze Genomsequenzen. Eine interaktive Plattform enthält die genetischen Informationen von 500 000 Pflanzen, 450 000 Bakterien und 120 000 Pilzen sowie von 240 000 menschlichen Darm-Metagenomen.

—

Scalable genome graph data structures for metagenomics and genome annotation
Gunnar Rätsch (ETH Zürich)

Genomvergleich: schnellere Analyse

Dieses Projekt entwickelte neue Methoden maschinellen Lernens, um Genome verschiedener Organismen trotz unterschiedlicher Datenqualität zu vergleichen. Solche Vergleiche verbessern das Verständnis der Evolution von Gen-Gruppen, die an bestimmten Stoffwechselprozessen beteiligt sind, und zeigen auf, welche statt mit Evolution mit essenziellen Versorgungsfunktionen (engl.: housekeeping gene) assoziiert sind. Im Rahmen des Projekts wurden insbesondere Methoden entwickelt, um Gene (oder Proteine) mit gleichen Funktionen bei Menschen und Modellorganismen, wie Fliegen oder Mäusen, zu finden. Diese Art der Forschung ist entscheidend für gross angelegte Projekte, wie das European Reference Genome Atlas Consortium oder das Earth BioGenome Project, die darauf abzielen, die Genome aller bekannten Tiere, Pflanzen und Pilze zu sequenzieren. Der Ansatz des Projekts könnte die Genomvergleiche möglicherweise erheblich beschleunigen.

—

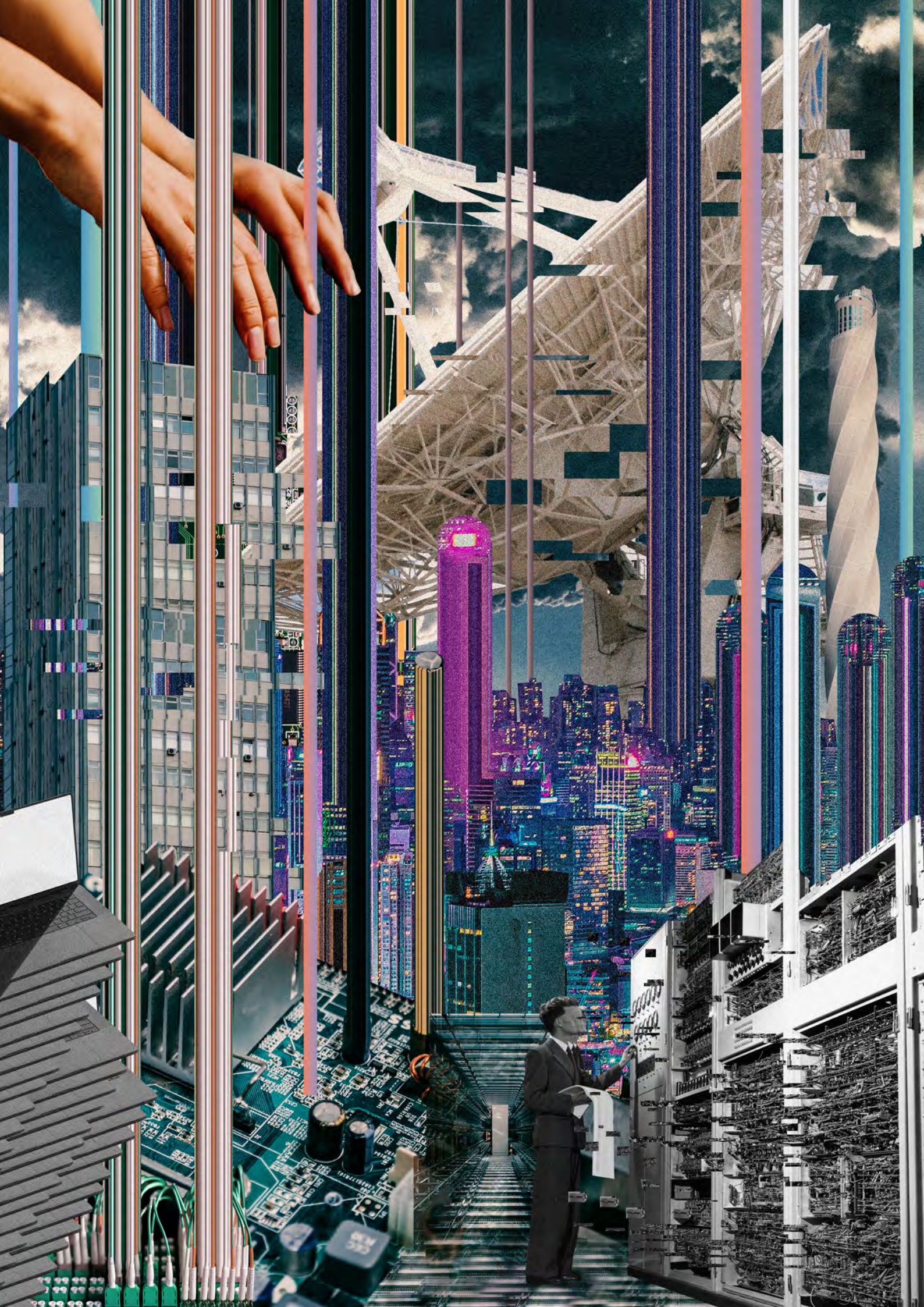
Efficient and accurate comparative genomics to make sense of high-volume low-quality data in biology
Nicolas Salamin (Universität Lausanne)

Bioinformatikdatenbanken: Abfragen in natürlicher Sprache

In diesem Projekt wurde eine benutzerfreundliche Schnittstelle zur Abfrage von Bioinformatikdatenbanken in natürlicher Sprache entwickelt. Die Plattform bietet visuelle Hilfestellung bei der Keyword-Suche während der ersten kursorischen Abfrage und beinhaltet Feedback, um Abfragen zu verfeinern und Ergebnisse zu verbessern. Das System integrierte Datenbanken für Proteine (UniProt), Genexpression (Bgee) und artenübergreifende Genexpression (OMA). Der Ansatz könnte auch über die Biowissenschaften hinaus verwendet werden und wird im europäischen Projekt INODE verfolgt, das Krebsforschung, Astrophysik und politische Entscheidungsfindung für Innovation umfasst.

—

BIO-SODA: enabling complex, semantic queries to bioinformatics databases through intuitive searching over data
Kurt Stockinger (ZHAW)



3.

Big-Data- Technologien

Big-Data-Anwendungen bedingen Big-Data-Technologien: Hardware- und Softwarelösungen, die in der Lage sind, riesige Datenmengen zu verarbeiten und sie zuverlässig und effizient zu analysieren. Die in diesem Kapitel vorgestellten Forschungsergebnisse des NFP 75 zeigen, wie Schweizer Forschende einen wichtigen Beitrag zur Entwicklung neuer Technologien für Big-Data-Anwendungen leisten und dabei helfen können, zukunftsweisende Lösungen für die notwendige Infrastruktur und Analytik erfolgreich einzusetzen.

Die reale Anwendung von Big Data steht vor tiefgreifenden technologischen Herausforderungen. Eine davon ist die schiere Menge der Daten: Nach gängiger Definition übersteigt Big Data die Möglichkeiten herkömmlicher Technologien zur Speicherung, Verwaltung und Analyse der Daten. Ein Grossteil der derzeitigen Recheninfrastrukturen werden bald veraltet sein und ersetzt werden müssen. Big Data erfordert neue Methoden der Datenverarbeitung und Datenanalyse.

Darum ist die Grundlagenforschung im Bereich der Big-Data-Infrastruktur und -Analysetechnologien so wichtig. Sie bereitet das Fundament vor, auf dem künftige Anwendungen aufbauen werden, und stellt ausserdem sicher, dass weiterhin hochqualifizierte Expert*innen zur Verfügung stehen. Die Schweiz betreibt Spitzenforschung in den Datenwissenschaften und muss diese auch weiterhin unterstützen. Sie sollte auch neue Infrastrukturen sowie die Erschliessung und die Betreuung zuverlässiger Datensätze unterstützen – und sich so auf die technologischen Herausforderungen von Big Data vorbereiten.

Das NFP 75 hat die Schweizer Grundlagenforschung in diesem Bereich gestärkt. Es hat ein Dutzend neuer Ansätze für die Entwicklung von Technologien hervorgebracht, die Big-Data-Anwendungen zugrunde liegen. Diese Ansätze lassen sich in zwei Bereichen zusammenfassen:

- IT-Infrastruktur, hauptsächlich Software, die für den Zugriff auf Daten und deren Bereinigung, Speicherung, Indexierung, Kontrolle, Vorverarbeitung und Überwachung benötigt wird (siehe Abschnitt 3.1)
- Datenanalytik, das meint vor allem Algorithmen zur Extraktion von Wissen aus Daten (siehe Abschnitt 3.2), zusätzlich zur Datenverwaltung.

Der Abschnitt 3.3 beschreibt die Herausforderungen bei denen die Forschung zu Big-Data-Technologien unterstützt werden kann. Der Abschnitt 3.4 fasst die Kernaussagen zusammen, während im Abschnitt 3.5 die einzelnen Projekte vorgestellt werden.

3.1 Effizientere Big-Data- Infrastrukturen

Big Data erfordert eine leistungsstarke Infrastruktur, insbesondere jene Low-Level-Prozesse, die als Rückgrat für die Datenanalyse auf höherer Ebene dienen. Diese Infrastruktur besteht aus Hardware und Software:

- Die Hardware umfasst Prozessoren wie CPUs, GPUs und TPUs (Central, Graphical sowie Tensor Processing Units), Transient und Persistent Memory, Kommunikationskomponenten, Stromversorgungs- und Kühl-elemente usw.
- Die Software-Infrastruktur ermöglicht den Datenzugriff und die Vorverarbeitung, die Programmierung, die Überwachung von Datenströmen oder Hardware usw.

Die Verbesserung der Infrastruktur für Big Data erfordert daher Fortschritte sowohl bei der Hardware als auch bei der Software. Während die Industrie den Fortschritt bei der Hardware vorantreibt, trägt die akademische Forschung massgeblich zu neuer Software bei, die eine schnellere und effizientere Verarbeitung grosser Datenmengen ermöglicht.

Wichtige technologische Konzepte für Big Data

Metadaten Informationen über einen Datenpunkt, z. B. wo und wann er erfasst wurde, seine Art oder seine Kategorisierung

Anonymisierung Entfernen aller Datenpunkte, welche die Identität einer Person preisgeben könnten, sodass die Daten anonym – oder besser «pseudonym» – werden

Re-Identifizierung Kombination mehrerer anonymisierter Datensätze, um Personen zu identifizieren

Künstliche Intelligenz Algorithmen und Maschinen, die «intelligentes» Verhalten zeigen, sowie die zugrunde liegenden Methoden und die realen Anwendungen

Maschinelles Lernen Computingmethoden, die es Algorithmen ermöglichen, selbstständig zu lernen, z. B. durch Trainingsdaten

Überwachtes Lernen Ein Ansatz des maschinellen Lernens, bei dem Algorithmen aus klassifizierten Trainingsdaten lernen

Unüberwachtes Lernen Ein Ansatz des maschinellen Lernens, bei dem Algorithmen Merkmale in Datensätzen entdecken, ohne klassifizierte Trainingsdaten zu verwenden

Das NFP 75 hat ein halbes Dutzend Ansätze zur Verbesserung der Infrastruktur, vor allem der Software, erforscht, vorgeschlagen und getestet:

- Ein Beitrag zum aufkommenden Bereich des In-Network-Computing, d. h. der Verarbeitung von Daten während ihrer Übertragung, durchgeführt an Kommunikationsknotenpunkten (*In-Network-Computing*)
- verbesserte Analyse dynamischer Netzwerke aus Streaming-Daten (*Graph Analytics und Mining*)
- Automatisierte Analysen der Beschreibungen heterogener Daten, um sie zu kombinieren und für die weitere Verarbeitung vorzubereiten (*Wenig strukturierte Daten*)
- Eine Echtzeit-Überwachung der Datenkonformität in Bezug auf vorgegebene Regeln (*Datenströme*)
- Die Schaffung verbesserter Schnittstellen für die Verarbeitung grosser Datenmengen in der Programmiersprache Scala (*Scala-Programmiersprache*)

Das Programm deckte nur einen kleinen Teil der vielen Fragen ab, die durch Big-Data-Technologien aufgeworfen werden, und der Ansätze, die derzeit weltweit erforscht werden.

Auswirkungen auf Big-Data-Infrastrukturen

Die Verarbeitung von Daten innerhalb von Kommunikationsknotenpunkten, während sie übertragen werden, kann die Verarbeitungsgeschwindigkeit deutlich erhöhen und die Latenzzeit für Operationen mit Big Data verringern (*In-Network-Computing*). Neue parallelisierte Algorithmen können in Echtzeit überwachen, ob ein schneller Strom von grossen Daten bestimmten Regeln entspricht. Methoden, die auf formaler Logik beruhen, gelten zwar oft als viel langsamer als Ansätze des

maschinellen Lernens, können jedoch hochskaliert werden (*Datenströme*).

Eine neuartige Computing-Architektur kann die Analyse grosser dynamischer Netzwerke verbessern (*Graph Analytics und Mining*). Ein Tool zur automatisierten Klassifizierung verschiedener Datentypen kann bei der Entwicklung von Anwendungen helfen, die Zahlen, Texte, Bilder oder Videos kombinieren (*Wenig strukturierte Daten*).

Eine neue Version der Programmiersprache Scala wurde im Jahr 2021 veröffentlicht. Sie wird den Umgang mit grossen Datensätzen erleichtern und neue Anwendungen von Big Data in vielen Bereichen weltweit ermöglichen. Scala wurde in der Schweiz erfunden und wird von grossen Technologie-, Bank- und Medienunternehmen eingesetzt (*Scala-Programmiersprache*).

3.2 **Neuartige Ansätze für Big-Data-Analytik**

Die Analytik ist die sichtbarste Komponente von Big-Data-Anwendungen. Sie schafft aus den Daten einen Wert, indem sie Wissen und Erkenntnisse extrahiert, die für Benutzer*innen oder Kund*innen wertvoll sind. Moderne Algorithmen für Datenanalyse, Data Mining und multidimensionale analytische Abfragen kombinieren fortschrittliche statistische Methoden und maschinelle Lernalgorithmen wie Deep Learning. Die Herausforderung für solche Algorithmen besteht darin, riesige

Datensätze zuverlässig, genau und in angemessener Zeit zu analysieren.

Auswirkungen auf die Big-Data-Analytik

Trotz der Verarbeitungsleistung, die durch verteilte Rechenleistungen und spezielle Prozessoren bereitgestellt wird, kann es immer noch Tage dauern, bis Algorithmen aus Trainingsdaten lernen. Um diesen Schritt zu beschleunigen, bedarf es entweder effizienterer Algorithmen oder cleverer Methoden, die Eingabedatensätze beschneiden oder komprimieren und so mit weniger Trainingsdaten gute Ergebnisse zu erzielen.

Das NFP 75 hat die Big-Data-Analytik auf sechs verschiedene Arten verbessert. Dazu gehört die Entwicklung effizienter Approximationen zur Beschleunigung der Datenverarbeitung und zur Reduktion der Menge an Trainingsdaten, ohne die Genauigkeit übermäßig zu beeinträchtigen. Die Forschung des NFP 75

- verbesserte die Methode der laufenden Analyse sehr grosser Datensätze unter Wahrung der Privatsphäre (*Datenstromanalytik*);
- reduzierte die Grösse des Datensatzes, der zum Trainieren von Modellen für maschinelles Lernen verwendet wird (*Coresets*);
- reduzierte den Zeitaufwand für das Training von Algorithmen, die zur Modellierung von Daten und zur Erstellung von Vorhersagen verwendet werden (*Schnelle Vorhersagealgorithmen*);
- ermöglichte effiziente Echtzeit-Überwachungen von Rechenfarmen (*Rechenzentren*);
- erweiterte unser Verständnis der Grenzen von tiefen neuronalen Netzen (*Modelle maschinellen Lernens*);
- verbesserte Sprachmodelle für dialogfähige Softwareagenten (*Sprachmodelle*).

Die Forschung im Rahmen des NFP 75 hat Fortschritte bei verschiedenen Arten von Big-Data-Analytik erzielt. Neue Methoden ermöglichen ein schnelleres Training von Algorithmen zur Analyse sogenannter Gauss'scher Prozesse, die zur Modellierung von Daten und zur Erstellung von Vorhersagen verwendet werden, wobei deren Unsicherheit präziser quantifiziert wird (*Schnelle Vorhersagealgorithmen*). Einige Datenverarbeitungen, wie z. B. die Bildanalyse, können jetzt in Echtzeit an einem eingehenden Datenstrom durchgeführt werden, ohne zu warten, bis alle Daten gespeichert sind. Diese Idee kann auch genutzt werden, um die Privatsphäre kontinuierlich zu schützen, während Daten erzeugt und übertragen werden: Durch Hinzufügen von Zufallsrauschen werden einzelne Datensätze unkenntlich gemacht (*Datenstromanalytik*).

Ein besseres Verständnis der Grenzen von tiefen neuronalen Netzwerken, z.B. inwiefern sie verallgemeinert und in verschiedenen Kontexten verwendet werden können, hilft, sie robust und zuverlässig zu halten (*Modelle maschinellen Lernens*). Theoretische Fortschritte auf dem Gebiet dialogfähiger Agenten erhöhen mittlerweile die Raffinesse und die Zuverlässigkeit der Softwareagenten (*Sprachmodelle*).

Daten können reduziert werden, ohne die Zuverlässigkeit zu beeinträchtigen. Die zum Trainieren von Modellen maschinellen Lernens verwendeten Daten können erheblich reduziert werden und damit können auch die erforderlichen Rechenressourcen verringert werden (*Coresets*). In Rechenzentren können selbst begrenzte Daten über Rechenressourcen zur Überwachung und Optimierung von Prozessen verwendet werden (*Rechenzentren*). Eine neue

Methode zur Synthese tabellarischer Daten minimiert das Risiko, dass urheberrechtlich geschützte oder vertrauliche Informationen nach aussen dringen (*Rechenzentren*).

3.3 Herausforderungen für die Forschung zu Big-Data-Technologien

Herausforderungen

Forschungsprojekte zur Verbesserung von Big-Data-Infrastrukturen und -Analytik stehen nicht selten vor ähnlichen Herausforderungen. Die Daten, die für die Entwicklung neuer Infrastruktursoftware benötigt werden, sind mitunter schwer zu beschaffen, da Unternehmen nicht immer bereit – oder in der Lage – sind, umfangreiche und realistische Fallstudien zur Verfügung zu stellen. Das kann an Vorschriften zur Wahrung der Vertraulichkeit und des Datenschutzes liegen, wie z.B. durch die DSGVO (*Rechenzentren*), aber auch an der Rotation von Mitarbeitenden, an Änderungen der internen Richtlinien oder an unrealistischen Erwartungen an Big-Data-Anwendungen (*Datenströme*). Daher mussten einige Projekte auf synthetische und Benchmarkdaten zurückgreifen, die für reale Anwendungen nicht immer geeignet sind. Insgesamt werden die Verfahren zur Regelung des Zugriffs, der Verarbeitung und der gemeinsamen Nutzung von Daten, insbesondere in Bezug auf den Datenschutz und die Einwilligung, als starr, schwerfällig und

zeitaufwändig empfunden (*Graph Analytics und Mining*).

Die Forschung im Bereich Big Data ist sehr kompetitiv, insbesondere aufgrund der Beteiligung multinationaler Technologieunternehmen. Diese Unternehmen investieren viel in Forschung und Entwicklung, verfügen über eine konkurrenzlose Computerinfrastruktur, Zugang zu den grössten Datensätzen der Welt und rekrutieren Expert*innen von Weltrang. Die akademische Welt kann auf hohem Niveau mit der Privatwirtschaft zusammenarbeiten, muss aber auch Nischen finden, in denen sie mit den wesentlich besser finanzierten und ausgestatteten Teams der Wirtschaft konkurrieren kann (Sprachmodelle). Langfristige Finanzierungen und flexible Forschungspläne sind notwendig, um die Chancen in diesem sich schnell entwickelnden Bereich zu nutzen.

Um auf dem neuesten Stand der Forschung zu bleiben, müssen die besten Forschenden eingestellt werden. Doch die Konkurrenz durch multinationale Technologieunternehmen ist gross. Ausserdem misst die akademische Welt Innovationen, wie der Entwicklung von Prototypen mit öffentlichen und privaten Partnern oder der Erstellung und Weitergabe von Open-Source-Software, wenig Bedeutung bei. Dadurch besteht wenig Anreiz, solche Projekte zu verfolgen (*Scala-Programmiersprache*).

Lösungsansätze

Die derzeitigen administrativen Prozesse in der Schweiz für den Zugriff, die gemeinsame Nutzung und die Verarbeitung von Daten können gestrafft werden, vor allem wenn sie sich auf die öffentliche Forschung beziehen. Die Datennutzung von Anwendungen kann sich negativ auf die Privatsphäre auswirken, aber eine pauschale

Einschränkung der Datennutzung hat auch Nachteile, z. B. macht sie Innovationen sehr zeitaufwändig. Datenschutz hat seinen Preis, der ebenfalls berücksichtigt werden muss.

Lösungsansätze bestehen darin, den Datenschutz zu einem inhärenten und möglicherweise obligatorischen Aspekt der Big-Data-Verarbeitung zu machen. Entwickler*innen und Nutzer*innen von Big-Data-Anwendungen müssen über die verschiedenen Techniken zur Wahrung der Privatsphäre und deren Vor- und Nachteile informiert werden. Idealerweise hätten sie Zugang zu Tools, die ihnen bei der Optimierung von Algorithmen helfen, wenn es darum geht, ein Gleichgewicht zwischen Datenschutz, Effizienz und Qualität der Dienste herzustellen. Die Digitalisierung von Informationen erfordert eine sorgfältige Analyse, um eine zielgerichtete Nutzung von Formaten und Metadaten zu gewährleisten.

Messwerte des wissenschaftlichen Erfolgs bei der Förderung und Mittelvergabe sollten über die übliche Anzahl von Veröffentlichungen und Anmeldungen hinausgehen und möglicherweise auch Auswirkungen der Forschung ausserhalb der akademischen Welt einbeziehen (vor allem, wenn die Arbeit mit Open-Source-Protokollen erfolgt). Wissenschaftler*innen müssen über Freiheit und Flexibilität in der Forschung verfügen, damit sie ihre Pläne so gestalten können, dass sie das Beste aus den sich schnell entwickelnden Bereichen wie Big Data heraussholen können.

Der Faktor Mensch kann ebenso wichtig sein wie der Zugang zu Technologie, zumal Letzterer oft «open source» und verfügbar ist. Die Unterstützung der akademischen Forschung ermöglicht nicht nur Fortschritte in der Big-Data-Technologie auf regionaler und

internationaler Ebene, sondern bildet auch Spezialist*innen aus, welche die Gesellschaft braucht. Diese Expert*innen werden nicht nur Technologien entwickeln, sondern auch ein Verständnis für die Probleme im Zusammenhang mit Big Data mitbringen, z. B. für die Verfügbarkeit von Technologien, den Datenschutz, die Cybersicherheit und die Beteiligung von Interessengruppen. So gesehen werden sie zu den strategischen Entscheidungen öffentlicher und privater Organisationen über die Digitalisierung beitragen.

3.4 Kernaussagen zu Big-Data- Technologien

Das NFP 75 hat zahlreiche Forschungsergebnisse von globaler Bedeutung hervorgebracht, indem neue Wege zur Verbesserung von Infrastruktur und Analytik verfolgt wurden, die schliesslich für die gewinnbringende Nutzung von Big Data erforderlich sind. Diese Grundlagenforschung ist zwar an sich sehr anspruchsvoll, allerdings ist der Weg zum Erfolg bekannt und somit letztlich direkter als die Entwicklung von Anwendungen. So kann beispielsweise der eingeschränkte Zugang zu Daten manchmal umgangen werden, indem künstlich erzeugte Datensätze verwendet werden, deren bekannte Eigenschaften das Testen und Abstimmen der neuen Systeme ermöglichen. Die Forschung beschäftigt sich nach wie vor mit der Frage, wie schnell die Systeme Daten verarbeiten und analysieren und das erwartete Ergebnis innerhalb einer bestimmten Fehlerspanne erreichen können. Mit anderen

Die Hürden zu offenen Forschungsdaten

Resultate und Daten zu teilen, kann die Forschung günstiger und effizienter machen. Daten können wiederverwendet und kombiniert werden, während neue Studien aus dem Bestand wissenschaftlicher Ergebnisse lernen können. Die geteilte Nutzung von Daten steigert die Produktivität, da sie den Zeitaufwand zur Datenerfassung reduziert und zugleich die Kreativität anregt, weil sie kostengünstige und innovative Experimente fördert. Ausserdem wird der Zugang zu Forschungsergebnissen geöffnet, auch für Personen ausserhalb der Wissenschaft. Tatsächlich sind Forschungsdaten jedoch oft unzugänglich und werden wenig genutzt. Wissenschaftler*innen, die gewillt sind, ihre Daten zu teilen, stehen immer noch vor vielen praktischen, institutionellen und finanziellen Hürden.

Diesen Problematiken widmete sich das NFP 75 mit der Querschnitt-Aktivität *Big Data: Offene Daten und rechtliche Herausforderungen*. Es befasste sich mit den konkreten Herausforderungen, denen sich Forschende bei der geteilten Nutzung, Veröffentlichung und Wiederverwendung von Daten gegenübersehen. Es führte Interviews durch, formulierte klare Empfehlungen für Forschungseinrichtungen und schlug konkrete Ratsschlüsse in Form eines Leitfadens vor.

Die wichtigsten Herausforderungen für Forschende

- Praktische und rechtliche Hürden erschweren den Zugang zu vorhandenen Forschungsdaten: mangelnde Datenqualität, veraltete Formate, uneindeutige Identifizierung der Quellen, unterschiedliche Speicherorte, unzugängliche Repositorien, veraltete Websites, unklarer rechtlicher Status der Daten und Einschränkungen bei der Weiterverwendung, Datenschutzgesetze, Verletzung der Rechte anderer usw.
- Forschende sehen sich bei der Veröffentlichung von Forschungsdaten auch mit organisatorischen, finanziellen und rechtlichen Hürden konfrontiert: fehlende Anreize, Kosten, technisches und juristisches Know-how, langfristiges Engagement, Risiko, dass andere Forschende ihnen zuvorkommen, oder das Risiko des Datenmissbrauchs durch externe Partner.
- Die fehlende Standardisierung von Forschungsdaten schränkt das Potenzial ausserdem für die Wiederverwendung ein.

Mögliche Lösungen

- Rechtsberatung zu Dateneigentum, geistigem Eigentum und Urheberrecht, vertraglichen Vereinbarungen mit Dritten und Datenschutzrecht
- Standardisierung von Datenformaten, Speicherung, Anonymisierungsprozessen oder regulatorischen Verfahren
- Identifizierung und Veröffentlichung von Best Practices und eines How-to-Leitfadens
- Finanzielle und reputationsbezogene Anreize für die gemeinsame Nutzung

Big Data: Offene Daten und rechtliche Herausforderungen
Sabine Gless (Universität Basel)

Worten, die Problemstellungen der Forschung sind klar definiert. Jedoch sind sie eingebettet in ein schnelllebiges Umfeld mit Akteurinnen und Akteuren in der Industrie und anderswo, die unterschiedliche Ziele verfolgen.

Wettbewerb und Zusammenarbeit zwischen privater und öffentlicher Forschung

Der intensive internationale Wettbewerb im Bereich der Big-Data-Technologien bedroht die digitale Autonomie

der Nationen, bedeutet aber auch eine Chance zur Zusammenarbeit. Die Wirtschaft, insbesondere in den USA und in China, macht viele der Fortschritte bei der Infrastruktur und der Analyse von Big Data und präsentiert ein gutes Drittel der Arbeit auf wissenschaftlichen Spitzenkonferenzen. Sie führten zur Entwicklung verbesserter Sprachmodelle und optimierter Hardware wie Tensor Processing Units, die auf die Ausführung neuronaler Netzwerke ausgerichtet sind. Die private Forschung und Entwicklung steht auch der weltweit besten akademischen Forschung in nichts nach.

Big-Data-Technologien der Unternehmen mögen universell erscheinen, in Anbetracht ihres Strebens nach einer breiten Akzeptanz. Eine CPU oder ein Vorverarbeitungsalgorithmus mag im Wesentlichen unabhängig von der Nutzung sein. Big-Data-Technologien werden jedoch zunehmend spezialisiert, um das jeweilige Problem und insbesondere die Art der Daten – ob dynamisch oder statisch, homogen oder heterogen usw. – bestmöglich zu bewältigen. Das bedeutet, dass auch die Industrie Einfluss auf das mögliche Spektrum der Big-Data-Anwendungen hat. Es ist daher entscheidend, dass die öffentlich finanzierte Forschung mit der Industrie Schritt hält, wenn die Gesellschaft ein Mitspracherecht bei der Zukunft der Digitalisierung haben soll.

Akademische Forschung bleibt für die Entwicklung von Big-Data-Technologien unverzichtbar, vor allem wenn es um Grundsätze von gesellschaftlicher Relevanz geht, die für grosse Technologieunternehmen jedoch von geringerer Bedeutung sind, wie die Senkung des Energieverbrauchs oder die Sicherstellung des Datenschutzes durch Technikgestaltung (Privacy-by-Design). Darüber hinaus kann die öffentliche Forschung mutiger sein, indem sie Wege mit hohem Risiko und hohem Gewinn beschreitet. Während die Wirtschaft oft Einheitslösungen verfolgt, hat die akademische Forschung erfolgreich eine breitere Palette von Hard- und Software für die Big-Data-Technologie entwickelt. Dazu gehören programmierbare Netzwerk-Switches, netzinterne Analysen, neue Programmiermodelle für domänenspezifische Geräte und Algorithmen, die auf formaler Logik anstelle von maschinellem Lernen basieren.

Akademische Forschung ist der Privatwirtschaft mitunter weit voraus, denn Letztere verlässt sich auf die

Innovationen von Spin-offs aus Universitäten. Bei einigen Forschungsthemen im Bereich Big Data ist dieser Abstand jedoch viel geringer, sodass die Zusammenarbeit zwischen Hochschulen und Wirtschaft gefördert wird – zumal Erstere die Rechenleistung, Speicherkapazitäten und den Datenzugang der Letzteren benötigen. Solche Kooperationen sind im Prinzip eine «Win-win-Situation», bei der Hochschulen von den Ressourcen der Wirtschaft, realen Problemen und grossen Herausforderung profitieren, während die Wirtschaft von modernster Forschung und innovativeren Ideen profitiert.

Ein schlummerndes Problem in der akademischen Welt ist allerdings die mangelnde Anerkennung von Forschenden, die Anwendungen entwickeln, Zusammenarbeit fördern und Open-Source-Software einsetzen. Sie kann Forschende von Weltrang dazu bringen, sich nicht mit konkreten Problemen zu befassen und nicht mit der Wirtschaft zusammenzuarbeiten. Daher sind vielfältigere Karrierewege in der öffentlichen Forschung gefordert sowie Bewertungen, die über die traditionellen wissenschaftlichen Veröffentlichungen und Förderungserfolge hinausgehen.

Die Personalproblematik

Eine grosse Herausforderung beim Einsatz von Big Data stellt der Mangel an qualifiziertem Personal entlang der gesamten Wertschöpfungskette dar, von den Infrastrukturtechnologien über die Anwendungen bis hin zu Unternehmensintegration und Regulierung. Es herrscht ein harter Wettbewerb um Talente. Viele der klügsten Köpfe werden von grossen multinationalen Unternehmen, aber auch von KMU und Start-ups abgeworben. Die akademische Forschung verliert infolgedessen, da sie

Mühe hat, die besten Wissenschaftler*innen anzuziehen – selbst auf der Ebene der Doktorierenden. Universitäten drohen daher, talentierte Forschende zu verlieren, wenn sie mit grossen Technologieunternehmen zusammenarbeiten. Schnelle und häufige Karrierewechsel bringen zwar neue Perspektiven und Verbindungen, sind aber ein Problem für Forschungsprojekte.

Auf der anderen Seite sorgt die weltweit führende Schweizer Forschung im Bereich Big Data dafür, dass viele Spezialist*innen, die von öffentlichen und privaten Organisationen benötigt werden, ausgebildet und geschult werden und gute Kontakte zu Hochschulen und Industrie pflegen. Das macht die Schweiz innovativ und attraktiv für multinationale Unternehmen und internationale Organisationen.

Die Beschaffung der Daten

Die zweite grosse Herausforderung ist die Verfügbarkeit grosser, hochwertiger Datensätze, die für eine realistische Bewertung von Big-Data-Analysen und -Anwendungen unerlässlich sind (siehe Kapitel 2).

Dieses Problem wird sich in dem Masse verringern, wie öffentliche und private Organisationen eine Datenkultur entwickeln, seine Lösung erfordert jedoch eine solide Strategie, die sicherstellt, dass Daten von hoher Qualität sind, ordnungsgemäss mit Metadaten versehen und durch Datenschutzpraktiken geschützt werden. Entwickler*innen und Nutzer*innen von Big-Data-Technologien müssen mit den verschiedenen Methoden zur Wahrung der Privatsphäre vertraut und in der Lage sein, das richtige Gleichgewicht zwischen Privatsphäre, Effizienz und Qualität der Dienste zu finden.

Privatsphäre und Datenschutz werfen viele Fragen auf, z. B. ob das schweizerische oder europäische Datenschutzgesetz die richtigen Grenzen im Datenmanagement setzt oder wie die Privatsphäre geschützt und gleichzeitig Innovation gefördert werden kann. Die derzeitigen ethischen, genehmigungsrechtlichen und administrativen Prozesse, die Verwendungen medizinischer und wissenschaftlicher Daten in der Schweiz regeln, werden als langsam und komplex empfunden und könnten gestrafft und vereinfacht werden. Dies ist jedoch eine vielschichtige Diskussion, die multidisziplinäre Ansätze erfordert, einschliesslich der Einbeziehung der Sozialwissenschaften. Sie wird in Kapitel 4 behandelt.

3.5 Die Forschungsprojekte zu Big-Data-Technologien

Das NFP 75 erschloss ein Dutzend verschiedener Ansätze zur Verbesserung der für Big Data benötigten Technologien, die Hälfte davon im Bereich der Infrastruktur, die andere Hälfte im Bereich der Analytik.

Big-Data-Infrastruktur

In-Network-Computing: Lösungen für Graph Analytics

Dieses Projekt hat mehrere Fortschritte bei der Analyse grosser Graphen (Netzwerke) mithilfe von In-Network-Computing gemacht, d. h. der Verarbeitung von Daten während der Übertragung

und vor der Speicherung. So bilden beispielsweise Telefongespräche und Beiträge in sozialen Medien hochkomplexe Graphen, deren Grösse und Verfügbarkeit rapide zunehmen.

Die neuen Methoden stützen sich insbesondere auf Netzwerkgeräte wie Field Programmable Gate Arrays (FPGAs) der neuen Generation oder auf programmierbare anwendungsspezifische integrierte Schaltungen (ASICs). Die Ergebnisse übertreffen die Leistung weit verbreiteter softwarebasierter Systeme um mehrere Grössenordnungen und erzielen netzwerkinterne Verarbeitungen von über vier Milliarden Ereignissen pro Sekunde.

–
Exploratory visual analytics for interaction graphs

Robert Soulé (Università della Svizzera italiana)

Graph Analytics und Mining

Dieses Projekt extrahierte Inferenzen aus einem Netzwerk und untersuchte die Graphen-Analytik auf verschiedenen Plattformen – einschliesslich Kombinationen von In-Core- und Out-of-Core-Verarbeitung. Einige Ergebnisse haben sich auch als wertvoll für Speichersysteme erwiesen, die bei der allgemeinen Big-Data-Verarbeitung eingesetzt werden.

–
Building flexible large-graph analytics and mining systems on commodity hardware

Willy Zwaenepoel (University of Sydney, davor EPFL)

Wenig strukturierte Daten: neue Techniken für die Datenintegration

In diesem Projekt wurden Instrumente entwickelt, um verschiedene Datentypen – wie Textdateien, PDFs oder Bilder – zu kombinieren und für die Verarbeitung und Analyse vorzubereiten. Die Datenvielfalt ist ein heikles Thema, vor allem angesichts der

Inkonsistenzen in den dazugehörigen Metadaten. Die neuen Tools beschleunigen den Prozess der Umwandlung von Rohdaten in Modelle und Visualisierungen, indem sie die Datensätze entweder automatisch oder halbautomatisch integrieren – auch wenn Menschen weiterhin Input liefern.

Die Instrumente wurden mit zahlreichen Datensätzen wie Twitter-Posts, Nachrichten, PDF-Dokumenten und medizinischen Bildern getestet. In Zusammenarbeit mit dem Schweizerischen Bundesarchiv wurde ein Tool zur Extraktion wenig strukturierter Daten in Archivdokumenten entwickelt, das zur Identifizierung von Dokumenten ohne offen zugängliche Daten verwendet wurde. Die Zusammenarbeit mit den Kantonsspitalern von Freiburg und Bern führte zu einem Prototyp einer neuartigen Architektur zur Integration von Informationen über Prostatakrebs, einschliesslich medizinischer Bilder.

–
Tighten-it-all: big data Integration for loosely-structured data

Philippe Cudré-Mauroux (Universität Freiburg)

Datenströme: Überwachung in Echtzeit

Im Rahmen dieses Projekts wurden effiziente parallelisierte Algorithmen entwickelt, um in Echtzeit zu überwachen, ob ein schneller Strom grosser Daten den vorgegebenen Regeln entspricht. Je komplexer die Regeln sind, desto schwieriger ist es, sie anhand riesiger Datenmengen effizient zu überprüfen. Ausserdem müssen die Überwachungsalgorithmen für die parallelisierte Ausführung in Computerclustern skalierbar sein.

Im Rahmen des Projekts wurden Überwachungsalgorithmen für sehr ausdrucksstarke und daher nützliche Eingabesprachen entwickelt und in zwei konkreten Anwendungen getestet. So entstanden Instrumente zur Prüfung

von Spesenabrechnungen und Zahlungen. Ausserdem wurde ein Proof-of-Concept-Tool entwickelt, um zu überprüfen, ob eine Verarbeitung datenneutral ist, d.h. nicht vom Anbieter kontrolliert werden kann. Es wurde gezeigt, dass die neuen Algorithmen mit industrietauglichen Überwachungssystemen konkurrieren können.

—

Big data monitoring

David Basin, Dmytro Traytel (ETH Zürich)

Scala-Programmiersprache: Big-Data-Analytik ermöglichen

Dieses Projekt führte mehrere neue Konzepte für Scala ein, eine an der EPFL entwickelte Programmiersprache, die führend wurde für Data-Science-Plattformen und -Tools. Das Team hat mehrere neue Technologien in eine kohärente Reihe von Abstraktionen für die Schnittstelle zu einem grossen Datensatz integriert und sie in Open-Source-Projekten validiert. Es ist zu erwarten, dass diese neue, 2021 veröffentlichte Version von Scala von Hunderttausenden von Softwareentwickler*innen auf der ganzen Welt übernommen wird.

—

Programming language abstractions for big data

Martin Odersky (EPFL)

Big-Data-Analytik

Datenstromanalytik: Tools zur schnellen und datenschutzgerechten Verarbeitung

In diesem Projekt wurden Tools entwickelt, die zur laufenden Echtzeit-Verarbeitung grosser Datenströme unter Einbezug des Schutzes der Privatsphäre mittels Hinzufügen von Rauschen zu Datensätzen beitragen. Das erste dieser Tools ist ein neuer Algorithmus zur Bildverarbeitung, der erfolgreich mit Daten des australischen Square

Kilometre Array Pathfinder getestet wurde – eines Radioteleskops, das bis zu 2 Gigabyte Rohdaten pro Sekunde erzeugt. Der Algorithmus konnte astronomische Bilder in Echtzeit rekonstruieren, ohne auf die gesamten Bilddaten warten zu müssen, und ebnete so den Weg für On-the-fly-Analysen im Petascale.

Das zweite Tool erleichtert die Implementierung und Anwendung der sogenannten differenziellen Privatsphäre, bei der randomisiert Datenpunkte verändert werden, um das Risiko einer erneuten Identifizierung zu vermeiden. Das Team entwickelte Methoden, um einem kontinuierlichen Datenstrom in Echtzeit differenzielle Privatsphäre hinzuzufügen, und schlug einen neuen Weg vor, um den Parameter darzustellen, der die Menge des hinzugefügten Zufallsrauschens steuert. Dadurch konnten die Benutzer*innen selbst über den Kompromiss zwischen Privatsphäre (starkes Rauschen) und Genauigkeit (geringeres Rauschen) entscheiden. Das Team veröffentlichte auch eine modifizierte Programmiersprache, um die Integration und die Steuerung von Techniken der differenziellen Privatsphäre für nicht spezialisierte Benutzer*innen zu vereinfachen. Das Team testete diese Konzepte anhand einer Analyse der Fernsehgewohnheiten von rund 3 Millionen Personen.

—

Privacy preserving, peta-scale stream analytics for domain-experts

Michael Böhlen (Universität Zürich)

Coresets: Big Data mit weniger Daten

Dieses Projekt hat gezeigt, dass es möglich ist, die zum Trainieren von maschinellen Lernmodellen benötigte Datenmenge zu reduzieren und dabei die Genauigkeit der wichtigsten statistischen Analysen und Lernprozesse nur geringfügig zu verringern – etwas, das auch bei dynamischen Datenströmen funktionieren kann. Das Team

verwendete sogenannte Coresets, die eine kleine Stichprobe verwenden, um grosse Datensätze zusammenzufassen, die jedoch robust und genau verarbeitet werden können. Dieser Ansatz kann auch die Privatsphäre besser schützen, da einzelne Datenteile in den Coreset-Stichproben erheblich verändert werden.

–

Scaling up by scaling down: big ML via small coresets

Andreas Krause (ETH Zürich)

Rechenzentren: effiziente Leistungsüberwachung

In diesem Projekt wurden neue Wege zur Auswertung der Leistung in Cloud-Rechenzentren entwickelt, eine wichtige Aufgabe bei der effizienten Verwaltung von Rechenressourcen bei gleichzeitiger Minimierung des Energieverbrauchs. Das Team verwendete approximative Analytik, die auf Teilmengen von Leistungsdaten – Protokolle virtueller und physischer Rechenressourcen – basiert, um komplexe Muster und Reihen der Ressourcennutzung vorherzusagen. Dieser Ansatz könnte die derzeitigen Überwachungssysteme verbessern, die in der Regel langsam und nicht sehr ausgereift sind.

Die Ergebnisse zeigen, dass es effektiver ist, aus einer kleinen Menge sauberer und informativer Daten zu lernen als aus einer grossen Menge von Daten geringer Qualität. Die Auswahl der richtigen Teilmenge ist von entscheidender Bedeutung, damit die Vorhersagen und Schlussfolgerungen nicht zu ungenau werden. Dies erfordert eine systematische Zusammenfassung des Datensatzes, anstatt zufällige oder einheitliche Unterproben zu nehmen. Das Team hat eine Methode entwickelt, um tabellarische Daten so zusammenzufassen, dass geschützte Daten, die von kommerziellen Unternehmen zur Verfügung gestellt werden, ohne Lecks weitergegeben werden können.

–

Dapprox: dependency-ware approximate analytics and processing platforms

Lydia Yiyu Chen (Technische Universität Delft, davor IBM Research Zurich)

Modelle maschinellen Lernens: Robustheit und Verallgemeinerung

In diesem Projekt wurden mehrere theoretische Fortschritte erzielt, mit deren Hilfe bewertet werden kann, ob Modelle maschinellen Lernens robust sind – d. h., ob sie auch dann noch verlässlich sind, wenn die Eingabedaten gestört sind – und ob sie verallgemeinerbar sind – d. h., ob sie reale Daten verarbeiten können, die sich von den Daten unterscheiden, die zum Trainieren der Modelle verwendet wurden. Das Team fand einen Kompromiss zwischen der Effizienz eines Modells und seiner Verallgemeinerbarkeit. Solche Fortschritte sind wichtig, um die Interpretierbarkeit und die Reproduzierbarkeit von Modellen zu verbessern, die entscheidend dafür sind, dass die Algorithmen unverzerrt (unbiased) und zuverlässig sind.

–

Theory and methods for accurate and scalable learning machines

Volkan Cevher (EPFL)

Schnelle Vorhersagealgorithmen

Im Rahmen dieses Projekts wurden Algorithmen entwickelt, die das Lernen aus grossen Datensätzen um ein Vielfaches schneller ermöglichen als die gegenwärtigen Methoden. Es handelt sich dabei um eine leistungsstarke statistische Methode namens Gauss'sche Prozesse, die zur Modellierung von Daten, zum Ziehen von Schlussfolgerungen und zur Erstellung von Vorhersagen verwendet wird. Vorhersagen werden in der Regel nicht als Einzelwerte, sondern als Wahrscheinlichkeitsverteilungen ausgedrückt, was besonders wichtig ist, wenn Unsicherheiten quantifiziert werden müssen – wie es

bei Wettervorhersagen der Fall ist. Das Team verwendete eine Kalman-Filter-Formulierung über einen reduzierten Satz von Trainingsdaten, sogenannte induzierende Punkte, und lokale Annäherungen unter Verwendung eines sogenannten korrelierten «Product of Experts» (PoE). Bei der neuen Methode skalieren Rechenzeit und Komplexität nur proportional zur Grösse der Datenproben, während frühere Methoden kubisch skalierten. Dies bietet gute Aussichten für die Ausweitung auf sehr grosse Datensätze. Es ermöglicht auch bessere Schätzungen der Unsicherheit von Vorhersagen und Schlussfolgerungen.

—
State space Gaussian processes for big data analytics

Marco Zaffalon (Istituto Dalle Molle di Studi sull'Intelligenza Artificiale USI-SUPSI)

Sprachmodelle: neue Methoden für dialogfähige Agenten

Im Rahmen dieses Projekts wurden mehrere theoretische Fortschritte im Bereich der Sprachmodelle – Systeme, die anhand bestimmter Eingaben Text generieren – erzielt, insbesondere für dialogfähige Agenten oder Dialogsysteme, die zur Beantwortung von Anfragen eingesetzt werden. Diese Aufgabe verlangt nach den Fragen angepassten Antworten, die auf ein bestimmtes Verständnis der in natürlichen Sprachen wie Englisch oder Mandarin ausgedrückten Eingaben hinweisen. Solche Systeme sind in der Kundenbetreuung, bei Suchmaschinen, in den sozialen Medien und im elektronischen Handel weit verbreitet und haben erhebliche wirtschaftliche und soziale Auswirkungen.

Das Team hat Technologien entwickelt, die direkt in Dialogsysteme integriert werden können. Die neuen Methoden können verwendet werden, um Textelemente zu verstehen, die sich auf eine Agentin oder einen Agenten beziehen

(Entity Detection und Linking), um Text durch tiefe neuronale Netze zu generieren, um die Leistung von Sprachalgorithmen zu bewerten und zu verbessern (Reinforcement Learning) wie auch von geometrischem maschinellem Lernen (das es ermöglicht, Komplexität und Struktur hinzuzufügen). Ein Teil der Arbeit wurde in Zusammenarbeit mit Google Zürich durchgeführt. Ein Spin-off-Unternehmen wurde gegründet, um Lösungen für die semantische Suche von Rechtsakten und die Schwärzung von Dokumenten für Rechtsteams zu kommerzialisieren.

—
Conversational agent for Interactive access to information
Thomas Hofmann (ETH Zürich)



4.

Gesellschaftliche, rechtliche und ethische Aspekte von Big Data

Sammlung, Analyse, Speicherung und geteilte Nutzung grosser Datenmengen werfen tiefgreifende gesellschaftliche, rechtliche und ethische Fragen auf. Die Suche nach Antworten erfordert interdisziplinäre Bemühungen, die mehrere Interessengruppen zusammenbringen. Dieses Kapitel gibt einen Überblick über die neuen Perspektiven, die das NFP 75 zu Fragen des Eigentums, der Kontrolle, des Zugangs und der Weitergabe von Daten, zu Datenschutz und digitaler Souveränität, Diskriminierung und Fairness sowie zu Wissensmanagement hervorgebracht hat.

Die immer umfangreichere Bereitstellung und Nutzung von Daten in der Gesellschaft hat die Aufmerksamkeit von Regierungen, Unternehmen, Organisationen und Einzelpersonen auf sich gezogen und wirft dadurch tiefgreifende ethische, rechtliche und soziale Fragen auf. Diese Fragen werden zwar breit diskutiert, sind aber noch nicht vollständig erschlossen. Angemessene Richtlinien für das Sammeln, Analysieren, Verwenden und Teilen von Daten müssen erst noch konzipiert, entwickelt, getestet, in staatlichen und institutionellen Regelungen berücksichtigt und in die Praxis umgesetzt werden.

Der Einfluss von Big Data ist in allen Bereichen der Gesellschaft zu beobachten, z. B. in den sozialen Netzwerken, deren Kombination zahlreicher Informationsströme ohne fortgeschrittene Analytik undenkbar wäre. Diese Technologien ermöglichen auch vorsätzliches Fehlverhalten und Manipulationen und schaffen damit neue Risiken, die bestehende in noch nie dagewesenem Umfang und Tempo verschärfen. Sie betreffen auch demokratische Prozesse, positiv wie negativ.

Die politische Sensibilität in Bezug auf mögliche negative Auswirkungen von Big Data hat in den letzten Jahren erheblich zugenommen. Dies wurde bei der Überarbeitung des neuen Datenschutzgesetzes deutlich, das am 1. September 2023 in Kraft treten wird.

Die vom NFP 75 gewonnenen Erkenntnisse

Durch das Nationale Forschungsprogramm «Big Data» wurden etliche ethische, regulatorische und rechtliche Fragen analysiert, die das rasante Wachstum von Big-Data-Anwendungen und -Praktiken hervorgebracht

hat, sowohl auf konzeptioneller Ebene als auch in spezifischen Kontexten.

Forschende des NFP 75 betrachteten spezifische Anwendungsbereiche wie das Gesundheitswesen, untersuchten, wie sich Vorschriften international ausbreiten, verfassten Richtlinien für die Versicherungsbranche, untersuchten die Wahrscheinlichkeit von Diskriminierung im Personalwesen, entwickelten Rahmenkonzepte für die ethische Nutzung von Daten im Gesundheitsbereich und untersuchten, wie der neue Beruf der Datenwissenschaftlerin und des Datenwissenschaftlers entstanden ist. Sie analysierten auch allgemeine Fragen im Zusammenhang mit der Souveränität und der Kontrolle von Daten, der Regulierung der Nutzung von Big Data in der Forschung und der Notwendigkeit, neue, durch Vorhersagemodelle entstandene Unsicherheiten anzugehen.

Die Ergebnisse der NFP-75-Forschung zu ethischen und rechtlichen Aspekten von Big Data sind nach vier wichtigen Themen geordnet:

- Dateneigentum, -kontrolle, -zugang und -weitergabe in Abschnitt 4.1
- Datenschutz und digitale Souveränität in Abschnitt 4.2
- Nicht-Diskriminierung, Fairness und Inklusion in Abschnitt 4.3
- Wissensproduktion und -management in Abschnitt 4.4

Ein Ausblick auf gesellschaftliche Fragen wird in Abschnitt 4.5 präsentiert. Der letzte Abschnitt 4.6 fasst die acht damit zusammenhängenden NFP-75-Projekte zusammen.

4.1

Dateneigentum, -kontrolle, -zugriff und -übertragung

Daten werden von Organisationen und Einzelpersonen mit einer Vielzahl von Aufgaben produziert, gesammelt, analysiert und weitergegeben wie nie zuvor. So veröffentlichen beispielsweise Regierungen einen Teil ihrer Daten – in der Regel in aggregierter Form und selten auf der Ebene von Individuen – im Einklang mit dem Paradigma der offenen Daten. Im Gegensatz dazu neigen kommerzielle Unternehmen dazu, Daten zu horten und sie nur widerwillig zu teilen.

Für Einzelpersonen haben persönliche Daten einen ganz besonderen Wert – und sie werden in grossem Umfang erzeugt. Die Menschen geben einen Grossteil ihrer Daten weiter und erhalten im Gegenzug Zugang zu kostenlosen Diensten, die bequem und effizient sind und ohne die man oft kaum leben kann, so z. B. E-Mail, Messaging, Bild- und Videofreigabe, Kartendienste, gezielte Empfehlungen und soziale Medien. Die Öffentlichkeit ist sich bis zu einem gewissen Grad der Risiken für die Privatsphäre sowie der möglichen Überwachung und des Missbrauchs bewusst. Dennoch geht die sehr liberale Weitergabe personenbezogener Daten unvermindert weiter. Nationale und internationale Vorschriften wie das Schweizer Bundesgesetz über den Datenschutz oder die DSGVO (Datenschutz-Grundverordnung) der EU werden zwar aktualisiert oder neu erlassen, jedoch verzögert. Ihre Wirksamkeit ist oft begrenzt, da sie erst Jahre oder

Jahrzehnte nach Beginn der Datenerfassung in Kraft treten.

Projekte des NFP 75 untersuchten Fragen des Eigentums und der Kontrolle über Daten sowie des grenzüberschreitenden Datenflusses (siehe Abschnitt 4.6 für weiteren Einzelheiten). Sie analysierten

- mehrere tiefgreifende rechtliche Herausforderungen, die durch Big Data aufgeworfen werden, insbesondere Eigentum, Schutz vor unzulässiger Überwachung und Selbstbelastung durch die eigenen Daten sowie die Durchsetzung von Rechtsansprüchen bei Verstössen im Zusammenhang mit der Datennutzung (*Rechtliche Herausforderungen von Big Data*);
- das Risiko, dass Menschen geschadet wird, deren Daten in der Forschung zu Big Data verwendet werden, einschliesslich Fragen wie Diskriminierung, Eingriff in die Privatsphäre und potenzieller Datenmissbrauch (*Regulierung der Big-Data-gestützten Forschung*);
- die Regeln der Welthandelsorganisation (WTO) und die seit 2000 abgeschlossenen zwischenstaatlichen Handelsabkommen sowie deren Auswirkungen auf nationale Big-Data-Vorschriften (*Handelsabkommen*).

Dateneigentum und -kontrolle

Bereits einige Konzepte bezüglich der Zuweisung von Rechten an die Inhaber*innen von Daten wurden in der Vergangenheit diskutiert. Was das Konzept des Dateneigentums angeht, stösst die normative Kategorie der Eigentumsrechte auf Probleme, da die gesetzlichen Bestimmungen in der Regel nur materielle Objekte, nicht aber Daten umfassen. Daten sind nicht materiell, sie können kopiert werden, sie

sind nicht rivalisierend (d.h. sie können von vielen Menschen gleichzeitig genutzt werden, ohne damit die Zugänglichkeit oder den Nutzen zu beeinträchtigen) und sie sind auch im Allgemeinen nicht exklusiv (d.h., sie stehen vielen zur Verfügung) – die beiden letztgenannten Eigenschaften charakterisieren gleichermaßen öffentliche Güter. Der Vorstellung von Dateneigentums fehlt daher die Relevanz (*Rechtliche Herausforderungen von Big Data, Regulierung der Big-Data-Forschung*), weshalb Regulierungsbehörden im Allgemeinen nicht mehr auf diesem Konzept aufbauen.

Die Zuweisung von Rechten geistigen Eigentums an Daten ist genauso schwierig. Solche Rechte müssen eine ausschliessliche Rechtsposition beinhalten, die gegenüber jeder anderen Person geltend gemacht werden kann. Ein immaterielles Gut muss aber die spezifischen Anforderungen der anwendbaren Gesetze erfüllen. Diese Bedingungen sind bei Daten nicht erfüllt, wie die Erfahrung in den meisten praktischen Fällen zeigt. Während Software dem Urheberrechtsschutz unterliegt, erreichen Daten nicht die Ebene einer spezifischen Schöpfung eines individuellen Geistes, die für die Zuweisung von geistigem Eigentum erforderlich ist.

Andere Rechtskategorien wie das Leistungsschutzrecht (eine besondere Form des Urheberrechts), das Deliktsrecht (Schadensersatz) oder das Persönlichkeitsrecht (Kontrolle der kommerziellen Nutzung der eigenen Identität) können in bestimmten Situationen Anwendung finden. Diese normativen Bestimmungen bieten jedoch keine geeignete Grundlage für einen allgemeinen Rahmen des Eigentums an Daten.

Ausschlaggebend dafür, wie Daten verwendet werden, ist der Zugang zu den

Daten und die Kontrolle darüber – nicht ihr potenzielles Eigentum. Personen oder Institutionen, die über Daten verfügen und sie verarbeiten, sind de facto in einer Eigentümerposition und haben die Macht, zu entscheiden, wie sie verwendet, gespeichert, gelöscht und übertragen werden. Regulierungsbehörden bauen daher in der Regel auf das Konzept des Datenzugriffs.

Zugriff auf eigene Daten

Die Regelung des Datenzugriffs ist entscheidend für diejenigen, die direkt oder indirekt an der Verarbeitung von Daten beteiligt sind. Einige datenverwaltende Stellen gewähren Einzelpersonen freiwillig Zugang zu deren persönlichen Daten. Darüber hinaus gibt es verschiedene Regelungen von Zugangsrechten.

- Einige allgemeine Rechtsinstrumente befassen sich speziell mit dem Zugang zu den eigenen Daten. So wird beispielsweise das Recht auf Datenportabilität (Teil des neuen Bundesgesetzes über den Datenschutz, ab September 2023) den Nutzenden ermöglichen, die gesammelten Daten von einem Dienstleister auf einen anderen zu übertragen.
- Derzeitige Regelungen zu Datenzugriffsrechten in der Schweiz finden sich vor allem im Kartellrecht, teilweise aber auch im Recht des unlauteren Wettbewerbs. Deren Anwendung steht vor vielen Herausforderungen, z. B. bei der Marktabgrenzung, bei der Frage, ob Daten korrekt und angemessen sind, und bei der Definition von Marktmacht. Zudem sind Kartellrechtsverfahren in der Regel teuer und langwierig, während die Wettbewerbsbehörde ihre Entscheidung oft erst fällt, wenn die konkrete Situation bereits entstanden ist.
- Der Datenzugang kann durch

Gesellschaftliche und ethische Aspekte im Zusammenhang mit Big Data

Datenschutz Einzelpersonen sollen vor einem unzulässigen Zugriff auf ihre privaten Daten und vor der Weitergabe und der Analyse dieser Daten durch andere geschützt werden.

Zugriff Menschen sollen in der Lage sein, auf ihre von Dienstleistern gespeicherten persönlichen Daten zuzugreifen und sie zu löschen.

Autonomie der Nutzer*innen Nutzer*innen sollen kontrollieren können, welche Daten auf welche Weise und zu welchem Zweck gesammelt werden – und zwar über die bloße Erlaubnis zur Verwendung bestimmter Cookies hinaus.

Gesellschaftliche Autonomie Die Entwicklung von Big Data wird hauptsächlich von Unternehmen vorangetrieben, ohne dass Bürger*innen oder Behörden eine Kontrolle ausüben.

Asymmetrie der Macht Bürger*innen, Unternehmen und Regierungen sind in der Praxis oft nicht in der Lage, Anbieter zu wechseln.

Regulierung Selbst Algorithmen, die mit grosser Verantwortung einhergehen, sind weitgehend unreguliert; im Gegensatz zu medizinischen Produkten oder Fahrzeugen. Unterschiedliche Vorschriften behindern internationale Forschungsprojekte.

Bias Daten sind nicht neutral: Sie spiegeln bestehende Verzerrungen in der Gesellschaft, z. B. eine eingeschränkte Repräsentation von Minderheiten oder diskriminierende Korrelationen.

Fairness Algorithmen, die mit verzerrten Daten trainiert wurden, liefern wahrscheinlich unfaire Ergebnisse.

Blackbox Das von einem Algorithmus erzeugte Ergebnis kann oft nicht erklärt werden. Dies beeinträchtigt die Zuverlässigkeit und das Vertrauen.

Vertrauen Die Gesellschaft muss Vertrauen in Big-Data-Anwendungen haben können. Dazu bedarf es Vertrauen in den gesamten Prozess der Datengenerierung und -nutzung und betrifft somit die Daten selbst, die Algorithmen sowie die Implementierung von Big-Data-Systemen.

Innovation Innovation braucht eine klare, stabile und ausgewogene Regulierung.

Geschäftspraktiken Big-Data-Anwendungen erfordern eine geteilte Nutzung von Daten, was Konflikte mit Geschäftsgeheimnissen hervorbringt.

Erkenntnisse zu ethischen, rechtlichen und sozialen Aspekten von Big Data

Das NFP 75 hat die *ELSI-Task-Force* eingesetzt, um spezifische ethische, rechtliche und soziale Aspekte im Zusammenhang mit Big Data zu untersuchen. Die aus dieser Studie resultierenden Kernaussagen werden im Folgenden zusammengefasst.

Big Data und digitale Souveränität

Das Konzept der digitalen Souveränität kann zwei Bedeutungen haben: die Autonomie eines Staates bei der Regulierung und dem Schutz der Daten seiner Bürger*innen oder die Selbstbestimmung der Nutzenden, wie ihre persönlichen Daten verwendet werden sollen. Diese beiden Bedeutungen stehen im Widerspruch zueinander, denn Big Data ist ein Gut, über das sowohl Individuen als auch Staaten Kontrolle ausüben wollen. Da Big Data immateriell ist, können Regierungen sie nicht auf der Grundlage der Souveränität über einen begrenzten physischen Raum regulieren. Stattdessen müssen sie miteinander kooperieren, um die Erhebung, Speicherung, Weitergabe und Übertragung von Big Data zu regeln. Der Schutz der digitalen Infrastruktur durch den Staat muss gegen die Autonomie der Bürger*innen abgewogen werden, um ungerechtfertigte Eingriffe in das Privatleben der Menschen zu vermeiden.

Die Herausforderungen informierter Einwilligung

Es könnte transparenter sein, zumindest teilweise auf die Notwendigkeit zu verzichten, die informierte Einwilligung der Bürger*innen zur Nutzung ihrer Daten einzuholen. Wenn dies der Fall ist, sollte eine solche Einwilligung durch Mechanismen ersetzt werden, die den Datenschutz durch Anonymisierung gewährleisten und eine angemessene Entschädigung garantieren, wenn diese Mechanismen versagen. Die Menschen würden klare Informationen darüber erhalten, wie ihre Daten verwendet werden, während Missbrauch entweder unwahrscheinlich wäre oder früh genug erkannt werden könnte, um grösseren Schaden zu verhindern – unabhängig davon, wie die Menschen ihre Daten verwenden möchten.

Die Bedeutung von Richtlinien

Datenschutzbestimmungen sind wichtig, um die Rechte des Einzelnen zu schützen, aber die Heterogenität der heutigen Rechtslandschaft macht solche Bestimmungen schwer verständlich. Es ist daher wichtig, eine Brücke zwischen dem Gesetz und praktischen Fragen zu schlagen. Angesichts der rasanten Veränderungen bei

Big Data ist es vielleicht besser, wenn die Regierungen Leitlinien vorgeben, anstatt Gesetze zu formulieren, insbesondere im Fall der informierten Einwilligung.

Ethik von Big Data im Gesundheitsbereich und in der biomedizinischen Forschung

Big Data könnte die Gesundheitsversorgung verbessern, wirft aber Fragen zu bestehenden Aufsichtsmechanismen wie dem Einsatz von Ethikkommissionen auf, insbesondere zur Frage, ob solche Gremien im Falle öffentlicher oder anonymisierter Daten Prüfungen durchführen sollten. Eine Klärung der Rolle dieser Kommissionen, die über Fachwissen in Bereichen der Datenwissenschaft verfügen sollten, könnte das Verantwortungsbewusstsein der Forschenden erhöhen. Ergänzende Gremien wie Datenethikausschüsse könnten ebenfalls in Betracht gezogen werden.

Die Asymmetrie zwischen Unternehmen und Kund*innen bei Big Data

Es besteht eine Kluft zwischen denjenigen, die grosse Datenmengen sammeln, speichern und analysieren, wie den Unternehmen, und denjenigen, deren Daten gesammelt werden, den Kund*innen. Statt zu versuchen, diese Kluft zu überwinden, wäre es vielleicht besser, diejenigen zu schützen, die von ihr bedroht sind – angesichts der Tatsache, dass diese Kluft in unserer heutigen Gesellschaft fast unvermeidlich ist. In diesem Sinne wären Werte wie Schadensvermeidung und Fairness wichtiger als Autonomie und Privatsphäre.

Risiken der Diskriminierung bestimmter Gruppen

Anwendungen maschinellen Lernens können zu Verzerrungen, Diskriminierung und Ungerechtigkeit führen. Zu den Erkenntnissen, die zur Bekämpfung dieser Probleme erforderlich sind, gehört die Einsicht, dass Verzerrungen nicht nur ein technisches, sondern auch ein normatives Element haben, und dass politische Entscheidungsträger*innen Fairness als Voraussetzung für die Entwicklung von Algorithmen mit hohen Anforderungen einbeziehen sollten. Diskriminierung kann jedoch nicht dadurch vermieden werden, dass die Verarbeitung von Gruppeninformationen verboten wird, da solche Informationen für die Bewertung der Fairness von Algorithmen benötigt werden. Es können verschiedene statistische Standards verwendet werden, um zu beurteilen, ob ein Algorithmus verzerrt ist, und Antidiskriminierungsmassnahmen sollten es ermöglichen, je nach konkretem Fall unterschiedliche Standards zu verwenden.

ELSI-Task-Force für das NFP 75, Markus Christen (Universität Zürich) Ethical, Legal and Social Issues of Big Data – a Comprehensive Overview, Eleonora Viganò (Ed.) (NFP 75, 2022)

sektorspezifische Vorschriften eingeschränkt werden, zum Beispiel im Gesundheitswesen.

Internationalität

Personenbezogene Daten sind als Wirtschaftsgut immer wertvoller geworden. Die Übertragung von Daten über Ländergrenzen hinweg erfordert auf juristischer Ebene grenzüberschreitende Vereinbarungen und harmonisierte Vorschriften. Deren Ausarbeitung wird jedoch durch die Vielzahl nationaler Gesetze zur Erhebung und Nutzung von Daten behindert, die wiederum durch internationale Handelsabkommen – bilateral oder global – beeinflusst werden (*Handelsabkommen*).

4.2 Datenschutz und digitale Souveränität

Datenschutz und Datensicherheit gehören zu den am meisten diskutierten politischen Fragen, die durch Big Data aufgeworfen werden. Datenschutzgesetze regeln zwar die Erhebung und der Verarbeitung personenbezogener Daten. Aber es ist oft schwierig, diese Art von Daten von der nicht personenbezogenen Variante zu unterscheiden. Insbesondere können verschiedene Daten, die an sich nicht personenbezogen sind, persönliche Informationen offenbaren, wenn sie durch den Querverweis mehrerer Datenbanken kombiniert werden – ein Prozess, der als Deanonymisierung bekannt ist. Darüber hinaus kann die Analytik neue Erkenntnisse liefern, indem sie die komplexen Korrelationen ausnutzt, die in

sehr grossen Datensätzen bestehen. Insgesamt gesehen lassen die technologischen Fortschritte im Bereich Big Data Zweifel an der Angemessenheit der traditionellen Grundsätze des Datenschutzes aufkommen.

In mehreren NFP-75-Projekten wurden Fragen des Datenschutzes und der digitalen Souveränität in einem konkreten Umfeld untersucht. Forschende

- untersuchten die ethischen, rechtlichen und gesellschaftlichen Fragen, die die Nutzung von Big Data im privaten Versicherungswesen aufwirft, und formulierten Empfehlungen dazu (*Big Data im Versicherungswesen*),
- erfassten die ethischen Fragen der Nutzung von Big Data im Gesundheitswesen, bewerteten die bestehenden Überwachungsmechanismen und entwickelten sowohl einen ethischen Rahmen als auch politische Empfehlungen (*Big Data im Gesundheitsbereich*),
- analysierten, wie Personalabteilungen die Nutzung von Big Data erschlossen haben, und bewerteten die Auswirkungen dieser Initiativen auf das Vertrauen der Mitarbeitenden (*Big Data im HR-Bereich*).

Grundsätze des Datenschutzes

Ein grundlegendes Problem ist die Frage, welche Personen oder Institutionen darüber entscheiden, wie Daten in verschiedenen Bereichen verwendet werden. In diesem Zusammenhang wird oft das Konzept der digitalen Souveränität verwendet, obwohl der Begriff mehrdeutig ist: Er kann informationelle Selbstbestimmung bedeuten, aber auch die persönliche Kontrolle über die eigenen Daten, sowie die Vorstellung, dass Daten die in einem Land gesammelt werden, nur den Gesetzen dieses Staates und nicht denen anderer

Staaten unterliegen sollten. Eine gründliche Analyse dieser Souveränität ist daher für künftige Regelungen wichtig (siehe das NFP 75 White Paper «Ethical, Legal and Social Issues of big data – a Comprehensive Overview»).

Die meisten Datenschutzgesetze beinhalten den Grundsatz der Datenminimierung, d. h. die Beschränkung der Erhebung, der Verarbeitung und der Nutzung von Daten auf das für einen bestimmten Zweck erforderliche Mass. Im Prinzip zielt dieser Gedanke darauf ab, die wahllose Anhäufung von Daten für zusätzliche Verwendungszwecke, ob vorhersehbar oder unvorhersehbar, zu minimieren. Allerdings sind grosse Datensätze in der Regel dazu gedacht, Korrelationen zu entdecken und neue Informationen zu generieren. Solche Ergebnisse sind unvorhersehbar. Dies kann wiederum als Rechtfertigung für die Erhebung und die Verarbeitung von Daten dienen, da die Einwilligung naturgemäss nur für ein Verfahren mit bekannten Ergebnissen gegeben werden kann. Eine parlamentarische Motion zur Ausarbeitung eines Gesetzes, das die Sekundärnutzung von Daten regelt, wurde am 22. August 2022 eingereicht und vom Ständerat am 14. Dezember 2022 angenommen¹⁰.

Der Grundsatz der informierten Einwilligung – dass Nutzende ihre ausdrückliche und informierte Zustimmung zur Erhebung, Verarbeitung, Nutzung und Weitergabe ihrer Daten geben müssen – ist in der Praxis nicht leicht anzuwenden. In der Tat sind die meisten Betroffenen nicht in der Lage, die vertragsrechtlichen Auswirkungen der Datenverarbeitung zu beurteilen oder die mit Big-Data-Analysen verbundenen Risiken zu bewerten.

Da die Grundsätze des Datenschutzes für den Umgang mit Big Data unangemessen sind, müssen diese Grundsätze und Gesetze dringend überdacht werden.

Neue Governance-Modelle und Technologien zur Verbesserung der Privatsphäre

Zukunftsweisende Datenschutzkonzepte sollten Governance-Elemente enthalten, z. B.:

- angemessene Analysen relevanter und potenzieller Risiken für den Datenschutz,
- die Festlegung einer geeigneten Strategie zur Einhaltung der Datenschutzgrundsätze,
- die Anwendung von Datenschutzrichtlinien und
- die Einsetzung geeigneter Verfahren zur Behebung von Datenschutzmängeln.

Der Schutz der Privatsphäre kann nicht nur durch Regulierung sichergestellt werden, sondern auch durch Technologien wie End-to-End-Verschlüsselung oder Vorgehen zur Gewährleistung der Anonymität. Geeignete Infrastrukturen, wie das Polybox-Ökosystem der ETH Zürich, können sowohl garantieren, dass Daten sicher geteilt werden, als auch sicherstellen, dass sie auffindbar, zugänglich, interoperabel und wiederverwendbar sind (*Big Data im Gesundheitsbereich*).

¹⁰ Rahmengesetz für die Sekundärnutzung von Daten, Motion 22.3890 (2022)

4.3

Fairness, Nicht-Diskriminierung und Inklusion

Die Themen Fairness, Nicht-Diskriminierung und Inklusion haben im Zusammenhang mit modernen Technologien viel Aufmerksamkeit erregt. Wie auch beim Datenschutz betreffen die Risiken besonders akut schwächere Teile der Gesellschaft. Ethische Implikationen von Big Data müssen sowohl allgemein als auch im Kontext spezifischer Anwendungen verstanden werden, z.B. bei Versicherungen, am Arbeitsplatz oder im Gesundheitswesen.

Datenfairness in konkreten Anwendungen

Versicherungswesen

Zunehmend detailliertere Daten über versicherte Personen ermöglichen eine stärker personalisierte Risikoprüfung. Dies untergräbt letztlich den Ansatz der Versicherer, Risiken zu bündeln, und bedroht damit das Prinzip der Solidarität, das dem Konzept der Versicherung zugrunde liegt (*Big Data im Versicherungswesen*). In der Tat wirft es die provokante Frage auf: Macht Big Data Menschen unversicherbar? Es ist nicht ausgeschlossen, dass manche Menschen tatsächlich vom Versicherungsmarkt ausgeschlossen werden, weil sie als zu riskant eingeschätzt werden. Hier spielen ethische Grundsätze eine wichtige Rolle, ebenso wie gesetzliche Regelungen. Für Versicherungsunternehmen konnten spezielle Verhaltenskodizes für den Einsatz von Big-Data-Analytik entwickelt werden (siehe «Vorschläge für die Versicherungsbranche», S. 67).

Beschäftigungsverhältnisse

Arbeitgebende müssen sich an die Grundsätze der Nicht-Diskriminierung und der Fairness halten und die Auswirkungen von diskriminierenden Algorithmen abmildern. Menschliche Integrität ist besonders wichtig in den Beziehungen zwischen Arbeitgebern und ihren Mitarbeitenden und für den Aufbau und die Aufrechterhaltung von Vertrauen am Arbeitsplatz (*Big Data im HR-Bereich*).

Das Monitoring der Tätigkeiten von Mitarbeitenden darf ihre Integrität und den moralischen Ruf nicht gefährden. Arbeitgebende müssen mit ihren Mitarbeitenden transparent kommunizieren und sie in die Gestaltung ihres Arbeitsumfelds einbeziehen. Das Personalwesen ist ein besonders sensibler Bereich, da seine Aktivitäten potenziell die Rechte der Mitarbeitenden verletzen oder deren Ruf schädigen können.

Gesundheitswesen

Die Konzepte der informierten Einwilligung, des minimierten Risikos und des Schutzes der Privatsphäre sind im Gesundheitswesen besonders wichtig (*Big Data im Gesundheitsbereich, Regulierung der Big-Data-gestützten Forschung*). Über die Vorschrift der Datenkonformität hinaus sollten Richtlinien und Toolkits eine ethisch vertretbare Nutzung von Daten gewährleisten.

An der Forschung zu Big Data für die Biomedizin und darüber hinaus sind eine Vielzahl von Interessengruppen beteiligt, darunter Forschende, Entwickler*innen, der private Sektor, Berufsverbände, Entscheidungsträger*innen und die Öffentlichkeit, die alle konkurrierende Interessen und unterschiedliche Erwartungen an eine solide Gesundheitsversorgung haben können. Daher sollten die bestehenden Aufsichtsmechanismen auf

verschiedenen Ebenen reformiert, gestärkt und an die neuen Technologien angepasst werden.

4.4 Wissens- produktion und -management

Entwickler*innen von Big-Data-Anwendungen spielen eine zentrale Rolle bei der Entwicklung und der Aufrechterhaltung allgemeiner und bereichsspezifischer ethischer Richtlinien und Praktiken. Neben dem Projekt (*Regulierung der Big-Data-gestützten Forschung*) untersuchten zwei weitere NFP-75-Projekte, wie sich Big Data auf die Forschungs- und Wissensberufe auswirkt.

- Ein Team wertete unter Rückbezug auf Erkenntnistheorie die Weise aus, wie Big Data in Computersimulationen für wissenschaftliche Forschung eingesetzt wird, insbesondere in der Klimawissenschaft (*Unsicherheit in Big-Data-Anwendungen*).
- Ein Projekt untersuchte mit soziologischen und ethnographischen Methoden die Rolle von Big Data in der soziologischen Forschung, der Datenwissenschaft und dem Datenjournalismus (*Big Data in der Praxis*).

Transdisziplinarität und neue Kompetenzen

Rechtliche, ethische und soziale Fragen rund um Big Data betreffen auch die Naturwissenschaften. Daher sind interdisziplinäre Ansätze erforderlich, um mit grossen und neuartigen Datensätzen umgehen zu können. Die Einbeziehung von Personen in

Angelegenheiten, die sie betreffen, muss verbessert werden, denn sie ist Teil eines grundlegenden Wandels in der Produktion von wissenschaftlichem Wissen sowie im Verständnis der Strukturen und Mechanismen, die bei der Erhaltung und Veränderung von Wissensdomänen eine Rolle spielen. Kompetenzen im Umgang mit Daten, Visualisierungen sowie computergestütztem

Vorschläge für die Versicherungsbranche

Das NFP-75-Projekt *Big Data im Versicherungswesen* analysierte die ethischen und rechtlichen Fragen, die durch Big-Data-Analysen in der Personenversicherung aufgeworfen werden. Es formulierte mehrere konkrete Vorschläge für die Branche und die Regulierungsbehörden:

- Ob und wie es Versicherungsunternehmen erlaubt sein sollte, ihre Versicherungsverträge auf der Grundlage von Big-Data-Analysen zu personalisieren, sind Fragen, die nicht indirekt durch die Anwendung allgemeiner Datenschutz- und Antidiskriminierungsgesetze gelöst werden sollten.
- Die Schweizer Aufsichtsbehörde sollte die Nutzung von Big Data zur Personalisierung von Versicherungsverträgen kontinuierlich überwachen und antizipieren. Das Versicherungsrecht sollte angepasst werden, um entweder unerwünschte Formen der Personalisierung zu verbieten oder die Art der zulässigen Personalisierung zu definieren.
- Versicherungsunternehmen sollten die Nutzung von Datenquellen vermeiden, die nichts mit dem versicherten Risiko zu tun haben, um das Vertrauen der Kund*innen in die Produkte und Dienstleistungen der Branche nicht zu erschüttern.
- Sie sollten ein Bewusstsein für potenziell diskriminierende Auswirkungen auf Vorhersagen, Preisgestaltung und Betrugserkennung durch den Einsatz von maschinellem Lernen etablieren.
- Sie sollten ihren Kund*innen zeigen, wie sie Privatsphäre, Fairness und Solidarität bei der Nutzung von Big-Data-Analytik schützen können.
- Sie sollten ihre ethischen Grundsätze so anpassen, dass sie verantwortlich sind bei der Behandlung von Fragen, die durch die Digitalisierung ihrer Branche aufgeworfen werden.

Big data ethics recommendations for the insurance industry, NRP 75 (2019)

Denken sind entscheidend für eine erweiterte Beteiligung an der Wissensproduktion (*Big Data in der Praxis*).

Kontext ist wichtig

Die Vielfalt der Big-Data-Forschung bedeutet, dass ethische Fragestellungen nicht durch alles übergreifende Einheitsregelungen angegangen werden können. Der Einbezug des Kontexts und seine Abwägung sollten im Vordergrund stehen, nicht eine starre Standardisierung (*Regulierung der Big-Data-gestützten Forschung*). Die Interdependenz rechtlicher, ethischer und sozialer Fragen und der Einbezug verschiedener Kulturen erfordern eine interdisziplinäre und internationale Forschung.

Big-Data-Wissen verstehen

Die Wissenschaft setzt Big-Data-Analytik auf vielfältige Weise ein, z. B. um Datenpunkte zu synthetisieren, Vorhersagen zu treffen und Beziehungen zu entdecken. Die Ergebnisse dieser Analysen sind mit zahlreichen Unsicherheiten behaftet, die bewertet, quantifiziert und angemessen kommuniziert werden müssen, wenn man den Ergebnissen vertrauen und sie nutzen will (*Rechtliche Herausforderungen von Big Data, Unsicherheit in Big-Data-Anwendungen*).

4.5

Herausforderungen und Kernaussagen

Herausforderungen bei der Erforschung der gesellschaftlichen, ethischen und rechtlichen Fragen von Big Data

Die Erforschung von Big Data in der Gesellschaft muss sich mit schnell ändernden Technologien und rechtlichen Fragen auseinandersetzen, so z. B. dem schwindenden Konzept des Dateneigentums (*Rechtliche Herausforderungen von Big Data*). Dies erfordert flexible Forschungsprogramme und Finanzierungen.

Wie auch bei Projekten zu Anwendungen und Infrastruktur kann es schwierig sein, Daten (insb. akademische) für sozialwissenschaftliche Forschung zu beschaffen (*Big Data in der Praxis, Big Data im Versicherungswesen, Unsicherheit in Big-Data-Anwendungen*) – wobei es in einem Bereich möglich war (*Handelsabkommen*).

Die ethische und rechtliche Betrachtung von Big-Data-Systemen ist im Wesentlichen interdisziplinär, was es schwierig macht, sich auf Terminologie, Methodik und Konzepte zu einigen, insbesondere zwischen Datenwissenschaftler*innen, Modellierer*innen sowie Rechtswissenschaftler*innen (*Rechtliche Herausforderungen von Big Data, Unsicherheit in Big-Data-Anwendungen*).

Kernaussagen

Obwohl das NFP 75 nur eine Teilmenge aller gesellschaftlichen, rechtlichen und ethischen Fragen im Zusammenhang mit Big Data behandelt hat, zeigt die Forschung dennoch einige allgemeine Probleme auf.

- Sowohl der öffentliche als auch der private Bereich muss zur Nutzung von Big Data transparenter auftreten.
- Weitere akademische Forschung sollte sich mit den möglichen Auswirkungen von Big Data auf die Demokratie befassen. Dazu gehört die immer wichtigere Rolle der Analytik in den sozialen Medien, die Verbreitung von Falschinformationen und Datenmanipulationen beschleunigt. Solche Entwicklungen sollten nicht dem Ermessen kommerzieller Unternehmen überlassen werden.
- Neben der Technologie sollten bei künftigen Überlegungen auch die gesellschaftlichen Aspekte der Big-Data-Analytik nicht vernachlässigt werden.

Gesellschaftliche Fragen angehen

Neue Datentechnologien verändern das Privat- und Geschäftsleben. Die Aggregation und Analyse von Daten, wie auch Analysetechniken haben grosse Auswirkungen auf Bereiche wie das Gesundheitswesen und den Arbeitsplatz. Dies erfordert die Durchführung kontextbezogener Analysen und das Vorhersehen gesellschaftlicher Folgen sowie die Ausarbeitung praktischer Leitlinien, die an unterschiedliche Umgebungen angepasst sind (*Big Data im Gesundheitsbereich, Big Data im HR-Bereich, Big Data im Versicherungswesen, Regulierung der Big-Data-gestützten Forschung*).

Angemessene Regulierungen entwerfen

Die Rechtsordnung muss einen verhältnismässigen normativen Rahmen

schaffen. Da das Konzept des Eigentums für (nicht physische) Daten schlecht geeignet ist, könnten die Gesetzgeberinnen und Gesetzgeber damit beginnen, das alternative Konzept der Datenrechtsinhaber*innen zu formulieren, das sich auf die Kontrolle von und den Zugang zu Daten konzentriert. In Bereichen wie der Blockchain werden neue normative Instrumente erforderlich sein, und es sollte weder zu viel noch zu wenig reguliert werden.

Ethische Richtlinien entwickeln

Ethische Richtlinien werden benötigt, da Inklusion und Fairness nicht ausreichend gesetzlich verankert sind und viele Grundsätze der Nicht-Diskriminierung nicht durch Verfassungen abgedeckt sind.

Solche Richtlinien sollten anhand konkreter Prozesse erstellt werden, die situationsbezogen sind und mehrere Interessengruppen einbeziehen – je mehr Interessengruppen es sind, desto wahrscheinlicher ist es, dass die Richtlinien befolgt werden (*Big Data im Gesundheitsbereich, Big Data im Versicherungswesen, Regulierung der Big-Data-gestützten Forschung*).

4.6

Die Forschungsprojekte zu gesellschaftlichen, rechtlichen und ethischen Aspekten von Big Data

Gesellschaftliche Aspekte

Big Data in der Praxis: Soziologie, Datenwissenschaften und Datenjournalismus

In diesem Projekt wurde mithilfe soziologischer und ethnographischer Methoden analysiert, wie Big Data in der Soziologie, den Datenwissenschaften und dem Datenjournalismus verstanden, gelehrt und genutzt wird. Es wurden mehr als 750 Lehrpläne von deutschen Universitäten untersucht und vier Kulturen in der Lehre der soziologischen Methodik identifiziert.

Um zu verstehen, wie die Datenwissenschaften in der Schweiz wahrgenommen und gelenkt werden, analysierte das Team etwa 4300 Online-Stellenanzeigen, 34 Politik- und Strategiepapiere sowie 40 neue Lehrpläne an Schweizer Hochschulen. Die Ergebnisse zeigen, dass die Datenwissenschaften eine Reihe von Methoden, Werkzeugen und Praktiken gemeinsam haben, dass sie die Grenzen zwischen den Disziplinen überschreiten und gleichzeitig an der vordersten Front der Konflikte zwischen Informatik und Statistik liegen. Ethnographische Studien zeigen, dass der Datenjournalismus spezifische epistemologische und professionelle

Kulturen erfordert, die mit den bestehenden journalistischen Praktiken koordiniert werden müssen. Insgesamt unterstreichen die Ergebnisse, dass Big Data Fähigkeiten erfordert, die disziplinäre Grenzen überschreiten.

—
Facing big data: methods and skills needed for a 21st century sociology
Sophie Mützel (Universität Luzern)

Unsicherheit bei Big-Data-Anwendungen: Lehren aus Klimasimulationen

In diesem Projekt wurden unter Rückbezug auf Erkenntnistheorie Computersimulationen untersucht, die für die Klimaforschung entwickelt wurden und auf Big Data basieren. Das Projekt zeigt, dass Klimaforschende immer vielfältigere Daten verwenden, so z. B. aus sozialen Medien oder Suchabfragen im Internet, was die Berechnungen effizienter macht und neue Beziehungen innerhalb der Modelle aufdeckt. Unsicherheiten ergeben sich jedoch aus Schwankungen in der Datenqualität und einem unvollständigen Verständnis der Rolle der Daten. Da die Forschungsergebnisse regelmässig mit neuen Beobachtungen verglichen werden, unterstreicht das Projekt, wie wichtig es ist, Big-Data-Methoden mit traditionellen wissenschaftlichen Ansätzen, wie dem Prozessverständnis, zu kombinieren. Es fordert eine transdisziplinäre Zusammenarbeit zwischen Fachexpert*innen sowie Datenwissenschaftler*innen, insbesondere zur Bewertung der Unsicherheit.

Das Team analysierte zwei Fallstudien zum Thema Klima: die hochauflösende Modellierung und Vorhersage von städtischen Wärmeinseln und die Abhängigkeit der städtischen Temperatur von der Vegetation und anderen Faktoren. Im ersten Fall unterscheidet die Forschung zwischen Unsicherheiten, die auf Beschränkungen des

Vorhersagealgorithmus selbst beruhen, und solchen, die sich aus endlichen Trainingsdaten ergeben. Letztere wirken sich eher auf spezifische Objekte als auf generische Objektkategorien aus.

—
*Combining theory with big data?
The case of uncertainty in prediction of
trends in extreme weather and impacts*
Reto Knutti (ETH Zürich)

Rechtliche Aspekte

Rechtliche Herausforderungen von Big Data

Das Projekt untersuchte mehrere rechtliche Fragen, die durch Big Data aufgeworfen werden. Es stellte fest, dass das Eigentum an Daten derzeit nicht definiert werden kann und dass Märkte für personenbezogene Daten, z.B. gezielte Werbung, im Wesentlichen ohne rechtliche Regulierung funktionieren. Das Forschungsteam evaluierte mögliche Alternativen zu bestehenden normativen Konzepten in Bezug auf Datenrechte.

Eine rechtliche Analyse zeigt, dass die Einwilligung zur Weitergabe von personenbezogenen Daten an eine Dienstleisterin oder einen Dienstleister nur teilweise auf die Rechte des Datenschutzes verzichtet. Eine andere Studie unterstreicht den Konflikt zwischen dem Auftrag der Behörden, die Gesellschaft zu schützen, und dem Grundrecht der Bürger*innen, nicht unangemessen überwacht zu werden. Fahrzeugdaten ermöglichen zum Beispiel die forensische Rekonstruktion von Unfällen, können aber gegen grundlegende Rechtsprinzipien wie das Recht gegen Selbstbelastung verstossen. In der Studie werden zahlreiche ungelöste Fragen im Zusammenhang mit Datenrechten erörtert, z.B. die Klärung der Frage, wer die Opfer von Datenkriminalität sind, und die

Tatsache, dass Bürger*innen in der Schweiz und in Deutschland derzeit Schwierigkeiten haben, in Strafverfahren berechnete Interessen geltend zu machen.

—
*Legal challenges in big data. Allocating
benefits. Averting risks*
Sabine Gless (Universität Basel)

Handelsabkommen: Auswirkungen auf nationales Recht

Das Projekt untersuchte Hunderte von Handelsabkommen der letzten zwei Jahrzehnte auf ihre Relevanz für die datengetriebene Wirtschaft. Es analysierte die bestehenden Normen und die immer zahlreicheren Bestimmungen, die direkt mit Big Data zu tun haben, z. B. die Bestimmungen zum elektronischen Handel, zum Datenschutz und zu offenen Regierungsdaten.

Das Team analysierte das Zusammenspiel zwischen internationalen Verpflichtungen und nationaler Politik und erstellte ausserdem eine offene zugängliche Datenbank, die nun von der OECD, dem britischen Handelsministerium, dem WEF und anderen genutzt wird. Die Untersuchung ergab, dass die Datenregulierung eine stärkere internationale Zusammenarbeit erfordert, wobei sich die Richtlinien in Bereichen wie Datenschutz und nationaler Sicherheit von Land zu Land erheblich unterscheiden. Die Ergebnisse unterstreichen die wachsende Bedeutung des Handelsrechts und zeigen Wege auf, wie dieses Recht in datengesteuerten Volkswirtschaften besser genutzt werden kann. Das Projekt argumentiert, dass die Schweiz als innovatives und global vernetztes Land eine wichtige Rolle spielen könnte.

—
*The governance of big data in trade
agreements: design, diffusion and
implications*
Mira Burri (Universität Luzern)

Regulierung der Big-Data-gestützten Forschung

Dieses Projekt analysierte die vielen ethischen Aspekte, die Forschung mit Big Data aufwirft, insbesondere Probleme der Diskriminierung, der Verletzung der Privatsphäre und des Datenmissbrauchs. Während ethische und regulatorische Prozesse in der Forschung am Menschen, etwa in der Medizin oder Psychologie, vertraut sind, wirft die Nutzung anonymer Daten in der Wissenschaft viele (wenig bekannte) Fragen auf.

Im Rahmen des Projekts wurde festgestellt, dass es aufgrund der Vielfalt der wissenschaftlichen Arbeit schwierig ist, einen umfassenden, harmonisierten und standardisierten Rahmen für Forschung mit Big Data zu schaffen. Es wird vorgeschlagen, dass die Regulierung stattdessen Entscheidungen auf der Grundlage des Kontextes, ethischer Überlegungen und der Analyse von Kompromissen umfassen sollte – ein potenziell fortlaufender Prozess. Prüfausschüsse sollten sich aus Fachleuten zusammensetzen, darunter auch Big-Data-Spezialist*innen, die ethische Aspekte von Forschungsprojekten während ihres gesamten Lebenszyklus bewerten. Die ethische Bewertung muss sich auch auf Forschung von Privatunternehmen erstrecken, die zunehmend mit der Wissenschaft zusammenarbeiten. Richtlinien, Verfahren und Verhaltenskodizes müssen regelmäßig überarbeitet werden, um mit dem technologischen und regulatorischen Wandel Schritt zu halten, z. B. mit der Entwicklung von Algorithmen, durch welche die Effizienz der Anonymisierung in Frage gestellt wird, oder mit der Umsetzung der europäischen DSGVO.

–
Ethical and legal regulation of big data research – Towards a sensible and efficient use of electronic health records and social media data

Bernice Simone Elger (Universität Basel)

Ethische Aspekte

Big Data im Gesundheitsbereich: ein ethisches Rahmenkonzept

Dieses Projekt befasste sich mit den ethischen Aspekten von Big Data im Gesundheitswesen und bewertete, ob solche Fragen von bestehenden Aufsichtsmechanismen wie Ethikkommissionen wirksam behandelt werden können. Ausserdem wurden ein ethisches Rahmenkonzept und politische Empfehlungen zur Unterstützung dieser Mechanismen entwickelt sowie ein praktisches Toolkit für Forschende und Ethikkommissionen vorgeschlagen.

Die Untersuchung zeigt, dass Wissenschaftler*innen sowie App-Entwickler*innen dazu neigen, Big-Data-Ethik als Einhaltung bestehender Datenschutzbestimmungen zu sehen, während sie Fragen wie die Verantwortlichkeit in der Forschung, Fairness, individuelle Autonomie und Schaden für Gruppen ignorieren. Ethikkommissionen in der Schweiz haben bestätigt, dass es ihnen an Fachwissen zu Big Data mangelt, und um Schulungen gebeten. Das Projekt kommt zum Schluss, dass die Ethikkommissionen ihre Vorschriften und Verfahren ändern müssen, damit sie die biomedizinische Forschung besser überwachen können. Das Team hat ein Toolkit entwickelt, um Ethikkommissionen zu unterstützen, ihre Bereitschaft für die Big-Data-Forschung zu beurteilen, sowie eine Checkliste, um die Überprüfung von Projekten zu erleichtern.

–
BEHALF – Big-data-ethics-health framework

Effy Vayena (ETH Zürich)

Big Data im Versicherungswesen

Ein interdisziplinäres Team untersuchte die ethischen, rechtlichen und gesellschaftlichen Aspekte von Big Data im privaten Versicherungsbereich und formulierte kurze Empfehlungen für

Versicherungsunternehmen mit grossen Kundendatensätzen (siehe «Vorschläge für die Versicherungsbranche», S. 67).

Im Rahmen des Projekts wurde festgestellt, dass der Datenschutz weniger ein Problem darstellt als prädiktive Analysen, die nicht nur das Risiko quantifizieren, sondern auch die Neigung der Kund*innen zur Zahlung der Prämien oder zum Betrug offenlegen. Die zunehmende Granularität der Risikobewertung macht eine Diskriminierung bestimmter Verhaltensweisen (z. B. keinen Sport auszuüben) wahrscheinlicher und diskriminiert indirekt die Bevölkerungsgruppen, die am meisten zu diesem Verhalten neigen. Eine Analyse der Gesetzgebung zeigt, dass die Schweiz eher liberal ist und den Grundsatz der Vertragsfreiheit viel stärker respektiert als beispielsweise Kalifornien.

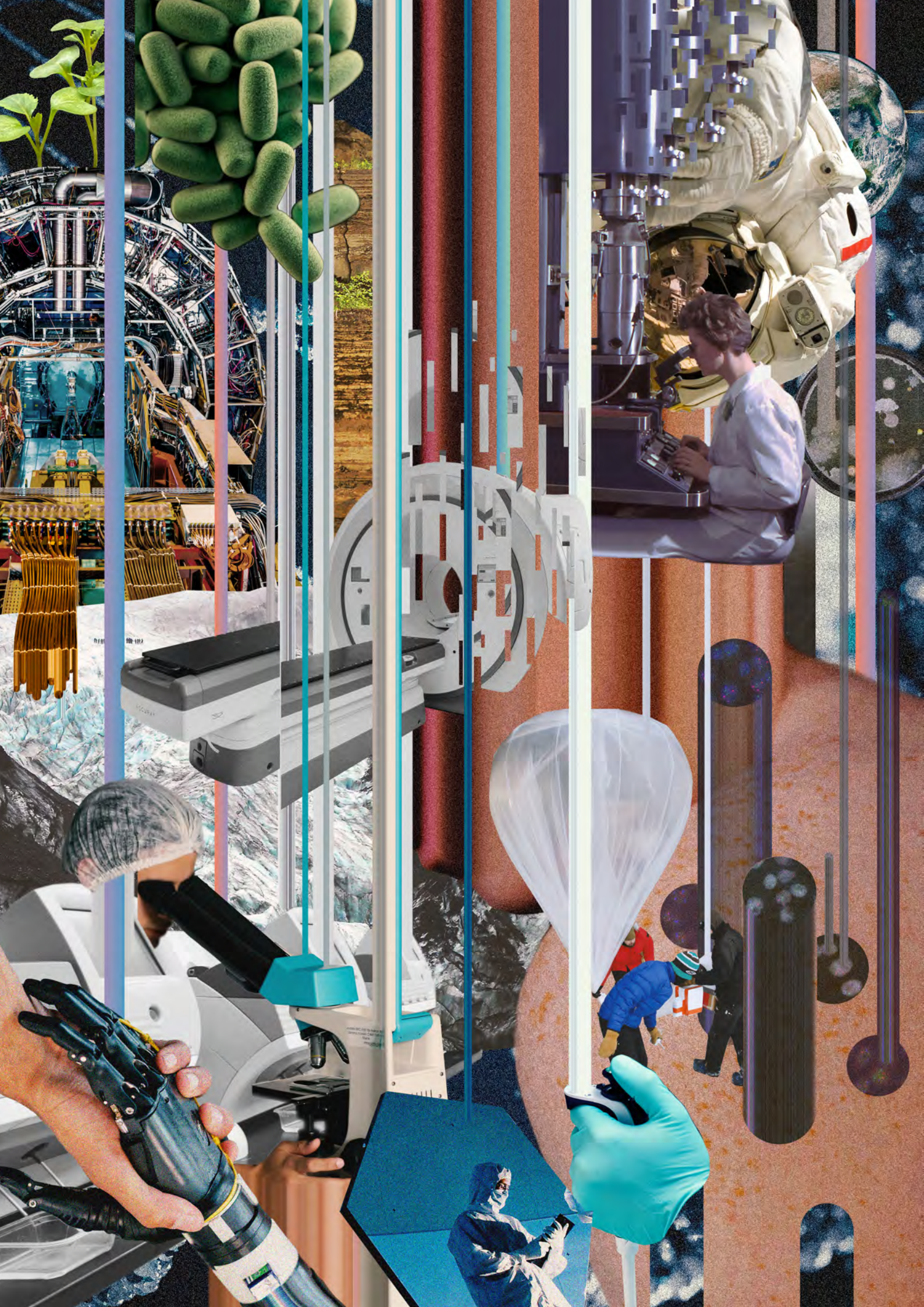
—
*Between solidarity and personalization –
Dealing with ethical and legal big data
challenges in the insurance industry*
Markus Christen (Universität Zürich)

Big Data im HR-Bereich

Dieses Projekt analysierte, wie Personalabteilungen, insbesondere in der Schweiz, Big Data implementiert haben und wie sich diese Integration auf das Vertrauen zwischen Arbeitnehmenden und Arbeitgebenden ausgewirkt hat. Dabei wurden grosse Unterschiede in der Transparenz der Unternehmen bei der Datenerfassung und z. B. der Befähigung der Mitarbeitenden aufgedeckt, was wiederum zu einem unterschiedlichen Mass an Vertrauen in die Arbeitgebenden führte. Der Aufbau von Vertrauen erfordert den Schutz der Autonomie und Handlungsfähigkeit der Mitarbeitenden sowie die Wahrung von Privatsphäre, Transparenz und Kontrolle. Die Mitarbeitenden müssen in strategische Entscheidungen einbezogen und ihre Beschwerden angehört werden.

Die Ergebnisse zeigen, dass insbesondere die schweizerischen Gesetze schlecht für den Umgang mit diskriminierenden algorithmischen Systemen am Arbeitsplatz gerüstet sind, was die Notwendigkeit angemessener rechtlicher Schutzmassnahmen und einer Modernisierung unterstreicht. Das Team hat ein Toolkit entwickelt, um Personalverantwortliche mit Big-Data-Techniken vertraut zu machen und die Bewertung von rechtlichen, ethischen und arbeitsplatzbezogenen Fragen zu ermöglichen.

—
*Big data or big brother? Big data
HR control practices and employee trust*
Antoinette Weibel (Universität St. Gallen)



5.

Ausblicke

Die Gesellschaft muss sich auf die potenziell disruptiven Veränderungen einstellen, die Big Data und Anwendungen maschinellen Lernens bewirken können. Im Folgenden wird ein Überblick über die wichtigsten Chancen und Herausforderungen gegeben, die vor uns liegen.

Die wichtigsten Errungenschaften des Nationalen Forschungsprogramms «Big Data» (NFP 75) haben die Fähigkeiten in der Schweiz in den Bereichen Technologien, Anwendungen und gesellschaftliche Aspekte von Big Data gestärkt. Das NFP 75 hat Technologien weiterentwickelt, die Big-Data-Infrastrukturen stützen und Big-Data-Forschende mit Expert*innen aus verschiedenen Bereichen zusammengebracht, um spezifische Anwendungen zu realisieren. Es hat auch das Bewusstsein für die gesellschaftlichen Herausforderungen geschärft, die mit der Produktion und Analyse grosser Datenmengen einhergehen, und schliesslich zur Entwicklung einer «Big-Data-Kultur» beigetragen, mit dem Ziel, die Vorteile von Big Data verantwortungsvoll zu nutzen.

Die 37 im Rahmen des Programms finanzierten Projekte deckten nur einen Teil dieses schnell wachsenden Bereichs ab. Dieses Kapitel geht darüber hinaus, um einen allgemeineren Überblick über Chancen und Risiken von Big Data zu geben, insbesondere über diejenigen, die in den kommenden Jahren an Bedeutung gewinnen könnten. Die folgende Einschätzung basiert auf den Erkenntnissen aus der Forschung des NFP 75 und den kollektiven Erkenntnissen der Mitglieder der Leitungsgruppe des Programms. Sie befasst sich sowohl mit der Aussicht auf eine stärkere Nutzung von Big Data in der Wirtschaft und im öffentlichen Sektor als auch mit den Herausforderungen in Bezug auf Nachhaltigkeit, Datenschutz und Rechenschaftspflicht.

5.1 Anwendungen wirken sich auf weitere Bereiche aus

Es ist zu erwarten, dass in den kommenden Jahren viele weitere Big-Data-Anwendungen entwickelt und eingesetzt werden. Neue private Sektoren, die über den elektronischen Handel hinausgehen, und öffentliche Verwaltungen werden sich auf Daten einstellen. Sie hoffen dadurch, wettbewerbsfähig zu bleiben, indem sie neue Fähigkeiten entwickeln, Kosten senken und gleichzeitig die Effizienz steigern. Wie einige Forschungsprojekte des NFP 75 gezeigt haben, erfordert die Entwicklung von realen Anwendungen die richtige Kombination von Fachwissen aus verschiedenen Bereichen. Es bedarf einer soliden Datenstrategie, die Privacy-by-design-Ansätze, analytisches Know-how bei Branchenexpert*innen und Umsetzungswissen in der Arbeitswelt umfasst. Ein entscheidender Faktor ist die Verfügbarkeit von Datenwissenschaftler*innen, die den jeweiligen Anwendungsbereich verstehen, wie von Fachleuten, die mit der Datenwissenschaft vertraut sind. Dies unterstreicht, wie wichtig es ist, die neuen – und älteren – Generationen mit dem Wissen und den Werkzeugen auszustatten, die zum Einsatz von Big-Data-Anwendungen erforderlich sind (siehe Schlussfolgerung 1 in Kapitel 6).

Es folgt eine Auswahl von Bereichen, die von Big-Data-Anwendungen stark betroffen sein könnten.

Produktion: Verbesserung des Outputs, Optimierung der Wartung

Viele Industrieprodukte enthalten integrierte Sensoren, die mit dem Internet der Dinge (IOT) verbunden werden können. Die Produkte können Leistungsinformationen in Echtzeit übermitteln, sodass Hersteller Komponenten identifizieren können, die ersetzt oder verbessert werden müssen, oder die Sicherheit und Kundenzufriedenheit erhöhen können.

In der Landwirtschaft nutzen autonome Robotersysteme Bilderkennung, um Unkraut zu entfernen, Krankheiten und Schädlinge zu erkennen, Früchte zu ernten, Dünger gezielt auszubringen und ganze Felder mit Drohnen zu überwachen. Solche Roboter könnten dazu beitragen, den Arbeitskräftemangel zu verringern, den Düngerverbrauch zu senken und Pestizide zu vermeiden¹¹.

Staat: Verbesserung der Infrastrukturen und Unterstützung der Energiewende

Regierungen können Big Data nutzen, um evidenzbasierte politische Entscheidungen zu treffen, z. B. bei der Zuweisung von Ressourcen, der Durchführung strategischer Planungen oder der Überwachung öffentlicher Infrastrukturen (Schlussfolgerung 5). Die Analytik von Big Data kann die Verkehrsplanung verbessern (*Optimierung der Verkehrssysteme*), Verkehrsstaus entschärfen, die Planung, den Bau und den Betrieb von Versorgungseinrichtungen verbessern, die z. B. Wasser, Strom und Beleuchtung bereitstellen, und Umweltüberwachungen durchführen (*Bodenerosion, Hochwassererkennung*). Hochentwickelte Analytik wird

dazu beitragen, unseren CO₂-Fussabdruck zu verringern, indem sie eine flexible Energieversorgung, -speicherung und -verteilung sicherstellt und insbesondere die Stromnetze in die Lage versetzt, mit dezentralen und intermittierenden erneuerbaren Energiequellen wie Sonnen- oder Windenergie umzugehen (*Potenzial erneuerbarer Energien*).

Dienstleistungen: Automatisierung in den Bereichen Finanzen und Cybersicherheit

Finanzinstitute können Echtzeit-Transaktionsanalysen und Marktprognosen für den schnellen automatisierten Handel nutzen, wofür effiziente Infrastrukturen erforderlich sind (*Schnelle Vorhersagealgorithmen, Graph Analytics und Mining*). Die Quantifizierung individueller Risiken erlaubt es Versicherungsunternehmen, ihre Policen besser anzupassen, bedroht aber möglicherweise das Solidaritätsprinzip, das der Versicherung zugrunde liegt (*Big Data im Versicherungswesen*). Tracking-Systeme in Fahrzeugen und anderswo könnten risikominderndes Verhalten belohnen und so den Schwerpunkt auf Risikoprävention statt auf Risikoschutz legen.

Analytik kann helfen, Cyberangriffe zu verhindern, indem sie nach Anomalien bei der Datenübertragung in Echtzeit sucht und dann automatisch Gefahrenquellen blockiert (*Datenströme*). Bilderkennung kann verwendet werden, um automatisch physische Sicherheitsverletzungen und andere Unregelmäßigkeiten zu erkennen.

¹¹ Siehe z. B. <https://www.agricultural-robotics.com> für einen Überblick.

Gesundheitsbereich: Unterstützung des medizinischen Personals und Personalisierung der Medizin

Es wird allgemein erwartet, dass maschinelles Lernen die Gesundheitsversorgung erheblich verbessern wird (Schlussfolgerung 4), wobei es bereits heute zur Erkennung von Anomalien in der klinischen Bildgebung eingesetzt wird¹². Neue Technologien könnten durch die Zusammenführung riesiger Datensätze aus Labortests, Krankenakten und Genetik grosse Fortschritte bei Prävention, Diagnose und gezielten Therapien ermöglichen. Insbesondere die fortschrittliche Verarbeitung der natürlichen Sprache (*Sprachmodelle*) ermöglicht automatische Extraktionen und Interpretationen von Daten aus unstrukturierten Texten in Krankenakten. Durch die Integration von Datenströmen aus verschiedenen klinischen Geräten in Echtzeit ist es auch möglich, den Gesundheitszustand von Patient*innen zu messen und Notfälle (*Intensivstationen*) zu erkennen.

Die Nutzung von Big Data für medizinische Anwendungen erfordert jedoch eine beachtliche Infrastruktur. Ausserdem werden innovative Methoden benötigt, um aus kleinen Teilmengen von Daten verlässliche Resultate zu generieren, angesichts der Tatsache, dass eine einzige Patientin oder ein einziger Patient Terabytes an Daten erzeugen kann. Bei genomischen Daten kann dies durch eine angemessene Vorverarbeitung geschehen (*Big Data Genetik*).

E-Commerce und Unterhaltung: Einbeziehung der Konsument:innen und synthetische Kunst

Kundendaten sammeln, analysieren und verwerten wird im E-Commerce wahrscheinlich eine immer grössere Rolle spielen. Online-Unternehmen nutzen bereits personalisierte Empfehlungen und Trendvorhersagen, aber neue datengesteuerte Anwendungen werden vermutlich Kundenerwartungen auch in den eigentlichen Produktgestaltungsprozess einbeziehen.

Sprachmodelle verbessern sich sehr schnell und werden immer besser darin, Bedeutung, Absicht und Kontext zu verstehen und wertvolle Informationen aus Texten zu extrahieren sowie synthetische Berichte oder Gespräche durch «Chatbots» zu erzeugen. Algorithmen können Musik auf der Grundlage von Stilen bestimmter Komponisten erzeugen. Computer erzeugen mithilfe von Textanweisungen überzeugend aussehende synthetische Bilder und Videos. Es ist zu erwarten, dass Software bald in der Lage sein wird, Filme zu erzeugen, die nicht als solche zu erkennen sind – komplett mit natürlich aussehenden Personen und Kulissen. Solche Systeme können die heutigen Medien- und Unterhaltungsprodukte ergänzen oder ersetzen, stellen aber auch grosse Herausforderungen für das geistige Eigentum sowie die Demokratie dar, da sie realistische computergenerierte Bild-, Audio- und Videofälschungen ermöglichen¹³.

¹² Siehe z. B. <https://grand-challenge.org/aiforradiology> für einen Überblick.

¹³ Der Rechtsstreit, der die Regeln des KI-Urheberrechts neu schreiben könnte, The Verge (2022)

Sicherstellung einer repräsentativen Big-Data-Forschung

Offene Forschung: Entdeckungen beschleunigen

Immer mehr Wissenschaftler*innen stellen ihre Datensätze anderen kostenlos zur Verfügung, um Entdeckungen zu beschleunigen und die Reproduzierbarkeit zu verbessern (Schlussfolgerung 6). Doch wie jedes andere Datenarchiv müssen auch sie die für eine offene Forschung bestimmten Standards einhalten – wie die «FAIR-Prinzipien» der Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit. Dies erfordert maschinenlesbare, standardisierte Metadaten, die notwendige Erklärungen und Beschreibungen enthalten – alles Teile eines neuen Paradigmas, an das sich die akademische Welt gewöhnen muss (*Big Data: Offene Daten und rechtliche Herausforderungen*).

Das NFP 75 förderte ein Programm zur Stärkung der Gemeinschaft von Wissenschaftlerinnen, die in der Big-Data-Forschung in der Schweiz tätig sind, da nur 22 Prozent der Absolvent*innen technischer Fächer weiblich sind – eine der niedrigsten Quoten in allen OECD-Ländern¹⁵. Diese Situation verstärkt den Fachkräftemangel, fördert Forschungsthemen, die nicht repräsentativ für die Gesellschaft sind, und formt Themen auf eine verzerrte Weise. Für Expertinnen bedeutet dies, dass es ihnen an Inspiration, Ermutigung und Unterstützung fehlt und dass Belästigung und Diskriminierung wahrscheinlicher werden.

Die NFP-75-Querschnitt-Aktivität (*Women in Big Data*) konzentrierte sich sowohl auf technische als auch auf soziale Wissenschaften. Sie lancierte verschiedene Aktionen zur Förderung von Karrieren, wie die Unterstützung beim Aufbau von Netzwerken, die Beseitigung von Hindernissen für technische Spitzenleistungen und die Förderung des interdisziplinären Austauschs über geschlechtsspezifische Probleme in Big Data.

Women in big data
Lydia Yiyu Chen (Technische Universität Delft)

5.2 Reduktion des Fussabdrucks der Dateninfrastrukturen

Big Data wird sicherlich eine wichtige Rolle bei der Bekämpfung des Klimawandels und der Verringerung unseres CO₂-Fussabdrucks spielen, trägt aber zugleich auch zu diesem Problem bei. Das Speichern und Verarbeiten grosser Datenmengen verbraucht erhebliche

Energiemengen: 3,6 Prozent des gesamten Stromverbrauchs der Schweiz im Jahr 2019 entfielen auf Rechenzentren, ein Anstieg um 30 Prozent in 6 Jahren¹⁴.

Die Verwaltung von Big Data ist mehr als nur das Sammeln und Speichern; die Daten müssen auch vor unbefugtem Zugriff, Korruption und Verlust geschützt werden. Dies erfordert eine Zugangskontrolle, Sicherungsprotokolle und Lösungen zur Korrektur beschädigter, unvollständiger oder ungenauer Daten. Datenbanken müssen bewahrt werden, indem sie kontinuierlich an neue Standards der Speicherung,

¹⁴ Das entspricht 2,1 TWh oder einem Viertel der Produktion des Kernkraftwerks Gösgen. Siehe Stromverbrauch der Rechenzentren in der Schweiz steigt weiter an, Schweizerisches Bundesamt für Energie (2021).

¹⁵ Schweiz: 22 Prozent der Absolvent*innen von MINT-Fächern (Mathematik, Informatik, Naturwissenschaften und Technik) sind weiblich, verglichen mit 26 Prozent in Deutschland, 32 Prozent in Frankreich und 40 Prozent in Italien. In der Informatik ist die Quote mit nur 16 Prozent sogar noch niedriger. Geschlechterunterschiede in MINT-Studiengängen: Eine deskriptive Analyse, KOF, ETH Zürich (2020)

Komprimierung und Analyse angepasst werden. Dies bedeutet Arbeit für Daten- und Fach-Expert*innen und erhöht die Kosten von Big-Data-Anwendungen. Sparsame oder leichtgewichtige künstliche Intelligenz zielt darauf ab, den Energieverbrauch zu senken, z. B. indem sie mit kleineren Datensätzen arbeiten kann und synthetische Trainingsdaten verwendet, die Ressourcen sparen. Dieses neue und wachsende Feld erfordert weitere Forschungsanstrengungen (*Coresets*).

5.3 Abwägung des Datenschutzes

Zahlreiche Big-Data-Anwendungen, z. B. im Finanzwesen, im Ingenieurwesen oder bei der Umweltüberwachung, werfen keine neuen Fragen zum Datenschutz auf, da sie keine personenbezogenen Daten verwenden. Doch viele andere Anwendungen tun dies, und die ständig wachsende Menge an Daten, die sie über Einzelpersonen sammeln, wirft ethische und rechtliche Bedenken auf. Anwender*innen haben in der Regel wenig Ahnung davon, welche Daten von ihnen Anbieter*innen von Online-Diensten kontrollieren. Dies hat zu den Begriffen «digitale Kluft» und «digitale Asymmetrie» geführt.

Obwohl Anbieter*innen derzeit verpflichtet sind, Nutzer*innen zu benachrichtigen und um Einwilligung zu bitten, wenn sie Daten sammeln, reichen diese Schritte nicht aus, um die Privatsphäre zu schützen, da die meisten Nutzer*innen ohne weiteres Nachdenken und mit wenig Wissen über die Konsequenzen zustimmen. Das Hauptproblem besteht darin, dass die

Nutzenden die Last tragen, die Auswirkungen ihrer Einwilligung zu verstehen, obwohl sie keinen unmittelbaren Nutzen aus der Datenerhebung ziehen. Die Behörden werden entscheiden müssen, inwieweit sie diese Praxis regulieren wollen (Schlussfolgerung 8).

Vollständige Anonymisierung ist oft unerreichbar

Bis vor Kurzem galt es als sicher, Daten weiterzugeben, die Informationen über Einzelpersonen enthielten, sobald sie anonymisiert wurden – durch Entfernen von Informationen, die eine Person direkt identifizieren könnten, wie Name, Geburtsdatum und Adresse. Es wird jedoch immer deutlicher, dass die Verknüpfung von Daten aus verschiedenen Quellen, auch wenn sie anonymisiert sind, die Wiedererkennung einzelner Personen ermöglichen. Bestimmte Arten von Daten, wie ganze Genome oder auch aufgezeichnete GPS-Daten eines Smartphones, enthalten ein so hohes Mass an sensiblen personenbezogenen Informationen, dass eine absolute Anonymisierung nicht realistisch ist. Die Freigabe von Daten, bei denen persönliche Informationen entfernt wurden, muss daher als Kontinuum betrachtet werden, bei dem der Verlust der Privatsphäre von Fall zu Fall gegen den geschaffenen Wert abgewogen werden muss.

Mehrere Ansätze können eine Re-Identifizierung verhindern. Der differenzielle Datenschutz beispielsweise verschleiert die Daten durch Hinzufügen von Zufallsrauschen, allerdings auf Kosten der Genauigkeit (*Datenstromanalytik*). Eine andere Möglichkeit besteht darin, bestimmte Datenpunkte zu unterdrücken oder zu umfassenderen Kategorien zusammenzufassen, wie dies bei der als k-Anonymität bezeichneten Methode geschieht.

Daten analysieren, ohne auf sie zuzugreifen

Sensible Daten können in Enklaven mit ausgeklügelter Zugriffskontrolle gespeichert werden. Dadurch wird sichergestellt, dass nur lokale Analysen durchgeführt werden können und dass nur aggregierte Ergebnisse, die Privatsphären schützen, nach ausserhalb der Enklaven gesendet werden. Eine weitere Option, die derzeit entwickelt wird, ist die föderierte Analyse, bei der Daten in mehreren lokalen Systemen ohne Austausch aufbewahrt werden. Die Berechnungen, einschliesslich des Trainings von Algorithmen maschinellen Lernens, werden lokal und gemeinsam durchgeführt. Auch hier werden nur partielle und aggregierte Ergebnisse oder Zwischenaktualisierungen des Modells geteilt, während die Originaldaten niemals übermittelt werden. Dies hilft, schwierige Fragen des grenzüberschreitenden Datentransfers zu lösen, der rechtliche Lösungen auf internationaler Ebene erfordert (Schlussfolgerung 9). Forschungsteams, die Anwendungen für Big Data entwickeln, sollten frühzeitig den ethischen und rechtlichen Rahmen der Datenverarbeitung in Betracht ziehen (Schlussfolgerung 2).

5.4 Algorithmen zur Rechenschaft ziehen

Big-Data-Anwendungen verwenden häufig Algorithmen maschinellen Lernens, die auf der Grundlage von Modellen, trainiert mit bestimmten Daten, Vorhersagen treffen können. Diese Algorithmen sind zwar oft sehr gut in der

Vorhersage, aber es ist oft unklar, wie sie genau zu den Vorhersagen gekommen sind. Dieses Problem kann ethische und rechtliche Fragen aufwerfen, wie in Kapitel 4 und im White Paper *Ethical, Legal and Social Issues of Big Data – A Comprehensive Overview* behandelt.

Das Risiko der Diskriminierung

Normale Software folgt einer strengen Reihe von Anweisungen, die (grösstenteils) von Menschen entworfen wurden. Programmierer*innen sowie Prüfer*innen können im Prinzip garantieren, dass sie wie erwartet funktioniert. Bei vielen Algorithmen maschinellen Lernens verhält es sich jedoch anders: Ihre Ergebnisse beruhen auf Modellen mit einer grossen Anzahl von Parametern, deren Werte automatisch aus Trainingsdaten generiert werden. Ihr Verhalten folgt nicht den von Menschen kodierten Regeln.

Das macht es schwierig herauszufinden, ob solche Ergebnisse mit etablierten ethischen Standards übereinstimmen oder ob sie z. B. bestimmte Bevölkerungsgruppen diskriminieren könnten. Dies kann geschehen, wenn die Trainingsdaten nicht repräsentativ, wenn sie verzerrt, veraltet oder fehlerhaft sind, was bei der Verwendung von Daten aus dem Internet der Fall sein kann. Modelle für maschinelles Lernen sind von Trainingsdaten abhängig, sodass Verzerrungen innerhalb dieser Daten in den Resultaten unter Umständen reproduziert werden. Wenn sie beispielsweise den Parameter «Geschlecht» aus den Trainingsdaten entfernen, kann das diskriminierende Ergebnisse nicht verhindern, da ein trainiertes Modell Ersatzwerte aus anderen Eingaben verwenden könnte, um die Kategorie Geschlecht wiederherzustellen. Ein solches Verhalten kann bei

frühen Tests unentdeckt bleiben, aber später auftauchen.

Maschinelles Lernen verstehen

Wie erwähnt können die Ergebnisse von tiefen neuronalen Netzen und anderen Techniken des maschinellen Lernens für Menschen sehr schwer zu verstehen sein, da die Milliarden von trainierbaren Parametern, aus denen die Modelle bestehen, die Mechanismen verschleiern, die zu bestimmten Ergebnissen führen. Derzeit gibt es keine akzeptierte Lösung, um dieses «Blackbox»-Problem der künstlichen Intelligenz vollständig zu überwinden.

Theoretiker*innen versuchen, diese automatisierten Systeme besser zu verstehen, um die Erklärbarkeit und die Nachvollziehbarkeit ihrer Entscheidungen zu verbessern. Diese Ziele sind entscheidend, um nachzuweisen, dass Algorithmen verantwortungsbewusst, vertrauenswürdig und nicht diskriminierend sind.

Normalerweise haben Personen oder Unternehmen, die von einem potenziell verzerrten Algorithmus betroffen sind, weder das Wissen noch die Fähigkeit, überzeugend zu argumentieren, dass das System einen Fehler gemacht oder sie diskriminiert hat. Eine Möglichkeit besteht darin, die Beweislast umzukehren, sodass Systemverantwortliche für die Algorithmen nachweisen müssen, dass sie sich korrekt verhalten. Dies könnte ein Zertifizierungsverfahren beinhalten, von einer öffentlichen oder privaten Organisation entwickelt (Schlussfolgerung 3), und ausserdem die absichtliche Veränderung von

Testdatensätzen, um zu prüfen, ob die Ausgabe mit ethischen Vorschriften übereinstimmt.

Wer ist für die Algorithmen verantwortlich?

Der rasante Fortschritt des maschinellen Lernens wirft die Frage der Haftung auf, wie sie bei selbstfahrenden Fahrzeugen ausführlich diskutiert wird. Wer sollte in diesem Fall für einen Unfall verantwortlich gemacht werden? Der Besitzer oder die Besitzerin des Fahrzeugs? Die Hersteller*innen? Niemand? Dies ist ein sich entwickelnder Bereich des Rechts und der Politik, und es besteht derzeit keine Einigkeit über die Antworten auf diese Art von Fragen. Zwar sind die Fahrzeuge so zu konstruieren, dass sie die Risiken in typischen Fahrsituationen minimieren, jedoch können nicht alle möglichen Umstände vorhergesehen werden. Es ist wichtig, die Verantwortlichkeiten genau zu definieren, damit Rechtsunsicherheit nicht die Innovation behindert.

Auf dem Weg zu einer neuen Regulierung

Die Gesetzgebung hinkt den rasanten Fortschritten im Bereich maschinellen Lernens und der immer umfangreicheren Datenerfassung oft hinterher. Sie hat sich bisher auf die Rechte Einzelner und die Vermeidung negativer Auswirkungen auf Einzelne konzentriert statt auf die Auswirkungen auf die Gesellschaft als Ganzes.

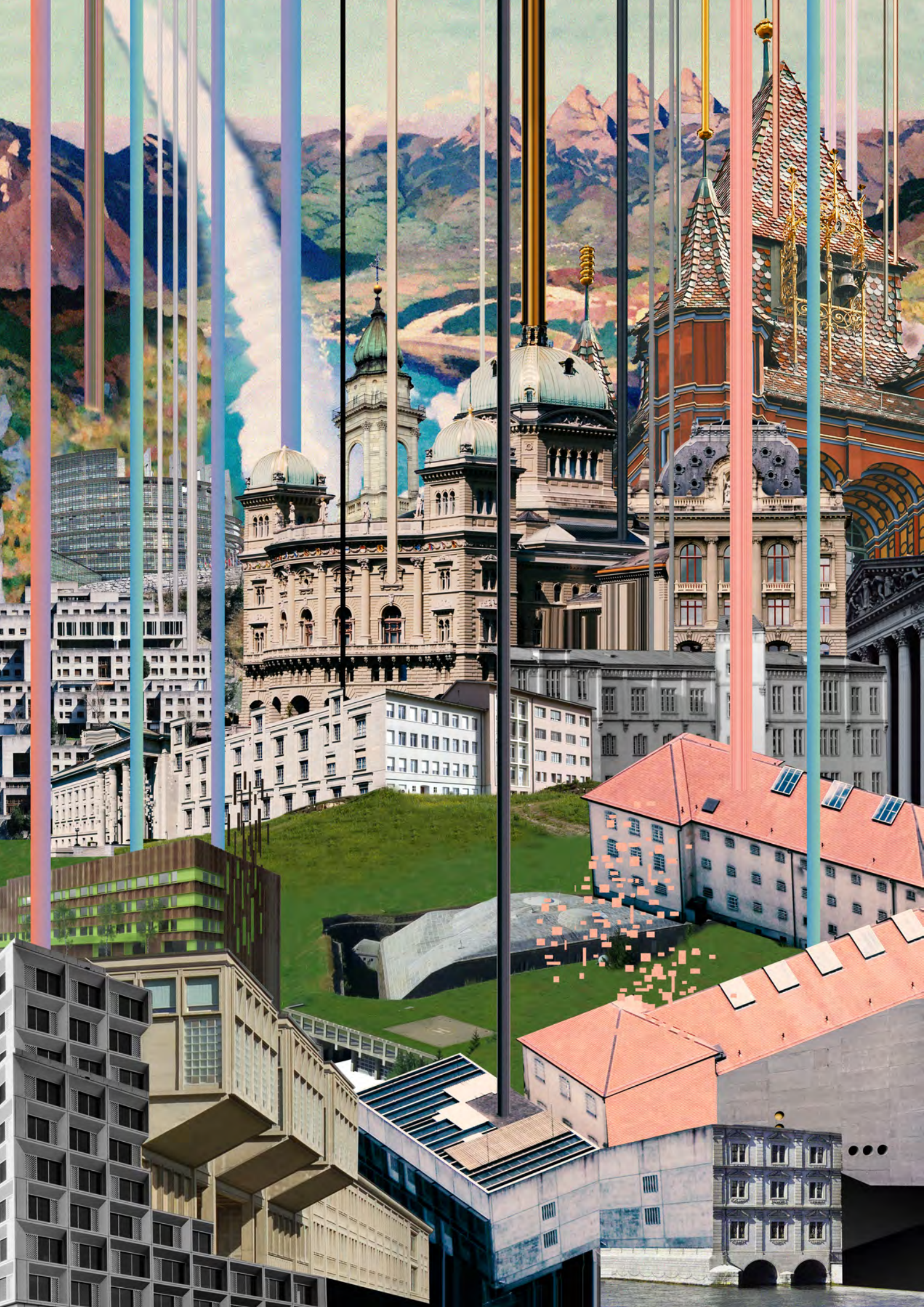
Die EU arbeitet derzeit an einem Gesetzesentwurf¹⁶ zur Regulierung von

¹⁶ Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften über künstliche Intelligenz (Rechtsakt über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union, Europäische Kommission (2021); siehe auch <https://artificialintelligenceact.eu/>

KI-Anwendungen. Es würde Anwendungen verbieten, die als inakzeptabel riskant gelten, wie manipulative Algorithmen oder Systeme zur sozialen Bewertung, und gleichzeitig solche einschränken, die als hoch riskant gelten, wie die Verwaltung kritischer Infrastrukturen oder die Strafverfolgung. China hat auch eine Ethikrichtlinie für KI formuliert, die soziale Sicherheit über individuelle Rechte stellt¹⁷. Diese Richtlinie schliesst den öffentlichen Sektor aus, dem es freisteht, Gesichtserkennung und soziales Profiling durchzuführen.

Die rasante Entwicklung der Technologie, die vor allem von internationalen Unternehmen vorangetrieben wird, stellt die Gesetzgebung vor eine grosse Herausforderung. Die Schweiz sollte proaktiv eine Gesetzgebung (Schlussfolgerung 7) entwerfen, die sicherstellt, dass die Regeln konkret angewandt werden können und dass ihre Einhaltung überwacht wird.

¹⁷ Ethikkodex für die neue Generation der künstlichen Intelligenz, Ministerium für Wissenschaft und Technologie (2021)



6.

Schlussfolgerungen der Leitungsgruppe

Die NFP-75-Leitungsgruppe hat zum Thema des Forschungsprogramms neun Schlussfolgerungen formuliert. Sie sollen für öffentliche Institutionen, private Organisationen und die Gesellschaft insgesamt einen nützlichen Beitrag bilden, zur Meinungsbildung über die notwendigen Entwicklungen im Zusammenhang mit der Nutzung von Big Data und ihrer verantwortungsvollen Anwendung.

Hand in Hand werden Big Data, künstliche Intelligenz und maschinelles Lernen die Gesellschaft tiefgreifend beeinflussen. Die potenziellen Vorteile dieser Technologien in allen Bereichen der Gesellschaft sind sehr vielfältig. Im Nationalen Forschungsprogramm «Big Data» (NFP 75) wurde eine Vielzahl von Möglichkeiten erforscht, um die Entwicklung neuer Technologien und Anwendungen zu beschleunigen, aber auch die damit verbundenen gesellschaftlichen Herausforderungen zu bewältigen. Die Kontrolle über diese Entwicklung zu behalten, stellt eine grosse Herausforderung für unsere öffentlichen und privaten Institutionen dar. Dies erfordert gezielte Anstrengungen in viele Richtungen, von der Bildung bis zur Regulierung, bereichsübergreifende Initiativen und öffentliche Diskussionen. Im Folgenden werden die Schlussfolgerungen der Leitungsgruppe des NFP 75 vorgestellt, die auf den Erkenntnissen aus fünf Jahren Forschung und den kollektiven Erkenntnissen der Leitungsgruppe beruhen.

Die gezogenen Schlussfolgerungen sollen geeignete Entwicklungen und Massnahmen anstossen und die bereits laufenden unterstützen. Konkrete Umsetzungsvorhaben sind aber Sache der entsprechenden Entscheidungsträger*innen. Die Schlussfolgerungen formulieren Erfordernisse und Aussichten, die sich aus der Perspektive der Forschung in den Bereichen der Datenwissenschaften und anderen Big-Data-bezogenen Forschungsfeldern ergeben. Sie sind ein Kondensat, das die Mitglieder der Leitungsgruppe auf der Grundlage der Ergebnisse der Forschungsprojekte und ihrer eigenen wissenschaftlichen Fachkenntnisse und Erfahrungen erarbeitet haben.

Forschung kann Antworten auf einzelne Fragen liefern und spezifische Lösungsansätze entwickeln. Es kann jedoch auch zu Überschneidungen und Interessenkonflikten zwischen den einzelnen Ansätzen führen. Es ist nicht Aufgabe der Forschenden, die gesellschaftlichen Prioritäten zu bewerten und eine Interessensabwägung vorzunehmen. Dies ist vielmehr eine Angelegenheit der Politik und der Stimmbevölkerung.

Dieses Programmrésumé des Nationalen Forschungsprogramms «Big Data» mit seinen Schlussfolgerungen ist daher ein Beitrag der Wissenschaft zur Meinungsbildung, zur politischen und fachlichen Debatte und zur Planung von Strategien und Massnahmen zur Umsetzung und Realisierung von datenanalytischen Anwendungen und Regelungen. Das Résumé wendet sich mit seinen Ansätzen und Schlussfolgerungen insbesondere an diejenigen Akteur*innen, die den Schweizer Datenraum massgeblich mitbestimmen und damit gestalten können.

Förderung eines geeigneten Umfelds für die Entwicklung von Big Data

1

Die Ausbildung von Big-Data-Fachleuten ausbauen

Die kompetente Nutzung von Big-Data-Technologien erfordert neue Kenntnisse und Fähigkeiten. So können heute IT-Fachleute, selbst solche, die innerhalb der letzten zehn Jahre ausgebildet wurden, mitunter Schwierigkeiten mit Aspekten von Big Data wie verantwortungsvollem Umgang mit Daten, Datenintegration und -technik, Analytik, maschinellem Lernen und Visualisierung haben. Entlang der gesamten Wertschöpfungskette und in den spezifischen Bereichen gibt es einen Mangel an qualifizierten Fachkräften. Es herrscht ein harter Wettbewerb um die besten Talente zwischen Hochschulen, grossen IT-Unternehmen und Start-ups. Um die Vorteile von Big Data in Unternehmen und der Industrie, in der Gesellschaft und in der Forschung zu nutzen, wird empfohlen, die schulische und universitäre Ausbildung im Bereich Big Data auszubauen, z. B. in Form von verbesserten Programmen und Weiterbildungsangeboten.

Die begrenzte Verfügbarkeit von qualifizierten IT-Fachkräften ist ein Engpass bei der Nutzung von Big-Data-Technologien und -Anwendungen. Dies verlangt nach einem Ausbau von Aus- und Weiterbildungsangeboten, sowohl in qualitativer als auch in quantitativer Hinsicht. Nur durch Bildung, Forschung und damit verbundene Öffentlichkeitsarbeit kann die Schweiz nationale und internationale Talente anziehen und damit ihre bereits starke Position im Bereich Big Data weiter stärken.

Auf universitärer Ebene, einschliesslich der Fachhochschulen, sollte die Zahl der IT-Studierenden erhöht werden, indem vermehrt spezielle Studiengänge für Big Data auf Bachelor-, Master- und Promotionsebene angeboten werden. In den Fachwissenschaften, die häufig Big Data nutzen, sollten Kurse zu einschlägigen Aspekten von Big Data angeboten werden. Dies könnte die Geschlechtervielfalt im IT-Bereich erhöhen, da Big Data über die Kerninformatik hinausgeht und interdisziplinäre Fähigkeiten erfordert, die auch gesellschaftliche, wirtschaftliche und rechtliche Aspekte umfassen.

Für IT-Fachleute und Auszubildende sollte das Weiterbildungsangebot zum Thema Big Data erweitert werden, wobei darauf zu achten ist, dass alle Aspekte der Big-Data-Pipeline abgedeckt werden. So werden Fachkenntnisse in der Datenerfassung und -aufbereitung benötigt, einschliesslich Fähigkeiten zur Integration, zur Bereinigung und zur Verknüpfung von Daten. Darüber hinaus ist ein geschulter Umgang mit Infrastrukturen für die Speicherung, die Verwaltung und die Wertschöpfung von Daten erforderlich, was eine Schulung in den neuen und sich schnell entwickelnden Open-Source-Bibliotheken erfordert, die für diese Zwecke zur Verfügung stehen. Darüber hinaus ist es oft von grosser Bedeutung, die Verwendung von Tools zur Visualisierung der Ergebnisse von Datenanalysen zu beherrschen. Schliesslich müssen die Fachleute die rechtlichen, regulatorischen und ethischen Fragen im Zusammenhang mit Big Data verstehen. Dies kann durch das Angebot interdisziplinärer Programme erreicht werden.

Da alle Ebenen der Gesellschaft von den rasanten Entwicklungen im Bereich Big Data betroffen sind, muss auch die Ausbildung im Bereich Big Data auf breiterer Ebene in Schulen und Berufsausbildungen intensiviert und kontinuierlich modernisiert werden, mit dem Ziel, die Datenwissenschaft mit ihren verschiedenen Facetten wie Recht, Ethik und gesellschaftlichen Fragen mit einzubeziehen und die Anforderungen zu berücksichtigen, die sich aus den technologischen und gesellschaftlichen Herausforderungen ergeben.

2

Rechtliche und ethische Beratung für Big-Data-Forschungs- und -Entwicklungsprojekte fördern

Viele Forschungs- und Entwicklungsprojekte benötigen rechtliche und ethische Beratung zu Aspekten wie der Frage, welche Daten sie nutzen oder weitergeben dürfen und wie solche Prozesse zu gestalten sind. Auch für den Fall, dass es zu einer öffentlichen Debatte über ein Projekt und seine Ergebnisse kommt, kann es für die Projektmanager*innen eine Herausforderung sein, erfolgreich zu kommunizieren und überzeugend darzulegen, dass ihre Bemühungen rechtmässig und ethisch einwandfrei sind. Dies kann und sollte dadurch ermöglicht werden, dass kompetente Beratung und vertrauenswürdige Audits zu erschwinglichen Kosten zur Verfügung gestellt werden.

Forschende in Forschungs- und Entwicklungsprojekten werden zunehmend rechtliche oder ethische Kenntnisse benötigen, um zu entscheiden, welche Daten genutzt oder weitergegeben werden können, und um bei der Gestaltung von Dienstleistungen zu helfen. Wenn weder die Projektteilnehmenden noch ihre unmittelbaren Kolleg*innen oder ihre Institutionen über das erforderliche rechtliche und ethische Fachwissen verfügen, besteht die vorsichtige Entscheidung darin, verfügbare Daten nicht zu nutzen oder weiterzugeben – selbst wenn die Nutzung oder die Weitergabe wertvoll und vollkommen rechtmässig und ethisch vertretbar wäre. Infolgedessen öffnet sich eine Lücke zwischen realen Problemen und Erkenntnissen sowie den untersuchten Versuchsumgebungen und Testeinrichtungen – mit negativen Folgen, wie sie zum Beispiel im Zusammenhang mit personalisierter medizinischer Behandlung, im öffentlichen Gesundheitswesen oder bei Nachhaltigkeitsmassnahmen auftreten können.

Diese Art von Problemen ist an Universitäten, öffentlichen Forschungseinrichtungen und Instituten für angewandte Wissenschaften besonders drängend. Infolgedessen werden Forschungschancen nicht wahrgenommen und Bildungsprogramme verlieren an Attraktivität. Die Situation wird durch das Risiko einer negativen Medienberichterstattung und den begrenzten Schutz von Universitätsmitarbeitenden, wenn dies geschieht, noch verschärft. Projektmanager*innen brauchen einen Dienst, entweder intern oder von der öffentlichen Verwaltung eingerichtet, der ihnen kompetente, vertrauenswürdige, verlässliche und zitierfähige Beratung zu rechtlichen und ethischen Fragen sowie zu kommunikativen Herausforderungen anbietet.

Wie sollten geplante Aktivitäten gestaltet werden, damit sie rechtskonform und ethisch einwandfrei sind? Was sind die Gründe für mögliche Einschränkungen? Und wie können die Aktivitäten transparent kommuniziert werden, insbesondere bei kontroversen öffentlichen Debatten? Beratung zu solchen Fragen, leicht nachvollziehbare Leitlinien für die praktische Umsetzung sowie leicht verständliche Erklärungen zu rechtlichen und ethischen Erwägungen werden das Ziel dieses Dienstes sein. Darüber hinaus sollte er Audits für Datennutzung und -weitergabe anbieten, die auf die spezifischen Herausforderungen eines Projekts eingehen und von qualifizierten, möglicherweise speziell zertifizierten Auditor*innen durchgeführt werden.

3

Zertifizierungen der Eigenschaften von Big-Data-Anwendungen ermöglichen

Big-Data-Anwendungen haben das Potenzial, eine Vielzahl unterschiedlicher Prozesse zu verbessern, auch in der öffentlichen Verwaltung und im privaten Sektor. In einigen Fällen können Anwendungen auf Bedenken stossen, die mit Konzepten von Fairness und Verzerrung, Diskriminierung, ethischen Standards, Datenschutz usw. zusammenhängen. Um die Anpassung solcher Anwendungen zu ermöglichen, wird empfohlen, Mittel zur Zertifizierung der relevanten Eigenschaften solcher Anwendungen bereitzustellen. Dazu gehört sowohl die Formulierung einschlägiger Merkmale als auch das Angebot von Verfahren für die Anbieter*innen von Anwendungen, mit denen sie die Einhaltung dieser Merkmale durch ihre Anwendungen zertifizieren können.

Die Funktionsweise und die Ergebnisse von Big-Data-Anwendungen sind in manchen Fällen nicht transparent, was zu Misstrauen führen kann. Dem kann einerseits durch die Definition wünschenswerter, sachdienlicher Merkmale von Big-Data-Anwendungen entgegengewirkt werden, z. B. von Konzepten wie Fairness, Verzerrung, Diskriminierung, Erklärbarkeit, Transparenz und Verantwortlichkeit. Andererseits können Wege geschaffen werden, die es Anbieter*innen von Anwendungen ermöglichen, zu dokumentieren, dass ihre Anwendungen diese Merkmale erfüllen. Um letztendlich ein angemessenes Vertrauen in Big-Data-Anwendungen aufzubauen und damit eine breitere Nutzung zu ermöglichen und ihren Nutzen zu erhöhen, wird empfohlen, Vorgehen zur Zertifizierung der relevanten Anwendungsmerkmale zu schaffen.

Big-Data-Anwendungen sind in nicht sensiblen Bereichen bereits weit verbreitet und führen oft zu einer verbesserten Leistung und Effizienz sowie zu Kosteneinsparungen. Sie haben jedoch auch das Potenzial, in sensiblen Bereichen beträchtliche Vorteile zu bewirken, z. B. bei der Strafverfolgung und im Sozial- und Gesundheitswesen. Prozesse für die Einhaltung von Vorschriften und die Zertifizierung der Einhaltung von Vorschriften gibt es bereits in wichtigen Bereichen wie Energie, Bauwesen oder Handel. Es wird empfohlen, solche Prozesse so zu erweitern, dass sie auch die spezifischen Eigenschaften von Big-Data-Anwendungen umfassen.

Integrieren von Big Data in öffentliche und private Organisationen

4

Verstärkte Nutzung von Big-Data-Technologien im Gesundheitssektor

Der Gesundheitssektor ist ein Paradebeispiel für einen Sektor, für den das Potenzial der Big-Data-Analytik von vielen Interessengruppen anerkannt wird, aber noch lange nicht ausgeschöpft ist. Eine stärkere Fokussierung auf datengestütztes Management und datengestützte Entscheidungsfindung könnte die gegenwärtigen Praktiken im Gesundheitswesen verändern und möglicherweise Transparenz, Qualität, Sicherheit, Effizienz und Koordination der Gesundheitsversorgung verbessern und die gesundheitsbezogenen Kompetenzen der Patienten steigern. Dieses Potenzial darf nicht ungenutzt bleiben. Die rechtlichen und ethischen Herausforderungen müssen daher angegangen werden, damit Big-Data-Analysen in grösserem Umfang im Gesundheitssektor eingesetzt werden können.

Die Erhebung und Nutzung qualitativ hochwertiger Gesundheitsdaten ist eine wichtige Voraussetzung für eine evidenzbasierte und personalisierte Medizin. Dies kann die Diagnosen verbessern, eine frühzeitige Erkennung von Risikopatient*innen ermöglichen, die Entdeckung neuartiger Wechselwirkungen zwischen Krankheiten oder Medikamenten und Risikofaktoren für das Fortschreiten der Krankheit ermöglichen und die Therapieverträglichkeit der Patient*innen verbessern. Um den Einsatz von Big-Data-Analysen im Gesundheitswesen zu ermöglichen, müssen jedoch bestehende Hindernisse (z. B. rechtlicher, administrativer, finanzieller oder ethischer Art oder IT-System- oder datenschutzbedingt) überwunden werden. Das Schweizer Gesundheitssystem ist relativ dezentralisiert und weitgehend auf lokaler Ebene organisiert, was die Gesetzgebung und die Verwaltung vor Herausforderungen stellt, um eine landesweite Datenerfassung und -nutzung zu ermöglichen. Aus datenwissenschaftlicher Sicht besteht die Gefahr, dass diese Organisation zu fragmentierten, lokalen IT-Lösungen mit verminderten Sicherheits- und Qualitätseigenschaften führt. Der Zugang zu Daten am richtigen Ort zur richtigen Zeit und die Schnittstellen zwischen den Akteur*innen müssen erheblich verbessert werden.

Die Schweiz hat das elektronische Patientendossier (EPD) und Qualitätsmassnahmen im Gesundheitswesen eingeführt, die Möglichkeiten für föderale Initiativen in diesem Bereich aufzeigen. In seiner derzeitigen Form stellt das EPD eine unstrukturierte Sammlung von (oft gescannten) Dokumenten dar, der es an strukturierten Zusammenfassungen und Indexierungen mangelt, dessen Datenformate nicht ausreichend harmonisiert sind und dem semantische Interoperabilität und standardisierte Terminologie fehlen. Darüber hinaus ist es nicht mit den Anforderungen der automatisierten Datenanalyse und den Standardisierungsrichtlinien kompatibel. Dennoch hat das EPD ein enormes Potenzial, wenn eine Anonymisierungsebene eingeführt wird, um kollektive Analysen von Patient*innendaten zu ermöglichen. Es gibt Technologien zum Schutz der Privatsphäre, die in dieser Hinsicht hilfreich sein können. So gibt es zum Beispiel Datenklaven, in denen die Daten nur verarbeitet werden, aber niemals direkt von aussen zugänglich sind. Ein weiterer Ansatz könnten föderierte Analysen sein, die eine Verarbeitung verteilter Daten ermöglichen, ohne deren Sicherheit zu beeinträchtigen. Solche Technologien müssen jedoch weiterentwickelt werden und erfordern ethisch, rechtlich, wissenschaftlich und statistisch versierte Gremien, welche die Verwendung der Daten überwachen.

5

Die politische Entscheidungsfindung und deren Evaluation mit Big Data stärken

Die Erhebung von Daten und die zunehmende Verfügbarkeit fortschrittlicher Datenanalysen bieten gemeinsam eine leistungsstarke Basis für die Stärkung der evidenzbasierten politischen Entscheidungsfindung. Dies ermöglicht es, soziale und wirtschaftliche Probleme zunehmend zu quantifizieren und die Wirksamkeit von politischen Massnahmen und Vorschriften immer besser auszuwerten. Dieses Potenzial sollte in einer Weise genutzt werden, die sowohl verantwortungsvoll als auch nutzbringend ist.

Es gibt ein enormes Potenzial für den Einsatz von Big-Data-Technologien in vielen Bereichen von Industrie und Wirtschaft sowie in der öffentlichen Verwaltung. Insbesondere kann der Einsatz von Big-Data-Technologien in Verbindung mit der Datenerfassung für die öffentliche Politikgestaltung und das Monitoring ihrer Wirksamkeit genutzt werden. Dies erfordert eine sorgfältige Berücksichtigung von Datenschutz- und Sicherheitsaspekten und deren angemessene Handhabung. Zu den Sektoren, auf die dies zutrifft, gehören zum Beispiel Gesundheit, Energie, Finanzen, Verkehr, Raumplanung und Sport. Neben anderen positiven Effekten ermöglichen die Erhebung und die Auswertung von Daten Wirksamkeitsvergleiche mit Massnahmen in anderen Ländern oder die Ausarbeitung verbesserter Massnahmen und bewährter Verfahren.

Im Gegenzug sollten die für die Gestaltung und die Evaluation der Politik verwendeten Daten öffentlich zugänglich sein, was wiederum Überlegungen zum Datenschutz und zur Anonymisierung erfordert. Insgesamt müssen die entsprechenden Ämter und Einrichtungen gestärkt und ausgebaut werden, um die zunehmende Arbeitsbelastung und die notwendige Kommunikation zu bewältigen. Ausserdem müssen Verfahren und Mechanismen zur Umsetzung dieser Zusammenarbeit festgelegt und entwickelt werden.

6

Geteilte Datensammlungen, Anwendungsbenchmarks und Open-Source-Software fördern

Die Verfügbarkeit einer Vielzahl von frei zugänglichen Infrastrukturen birgt das Potenzial, die Wertschöpfung aus Daten zu beschleunigen. Um mehr offene Daten zu ermöglichen, wird die Entwicklung einer verfeinerten Politik zur Veröffentlichung von Daten empfohlen. Ebenso ist eine bessere Unterstützung für die Erstellung von Benchmarks und Anwendungsfällen in verschiedenen Wissenschaftsbereichen gefordert. Darüber hinaus stellt Open-Source-Software oft eine attraktive Alternative zu kommerzieller Software mit teuren Lizenzen dar. Um neue Open-Source-Funktionen und -Kapazitäten zu ermöglichen, z. B. neue Recheninfrastrukturen und Toolkits für maschinelles Lernen, werden zusätzliche Mittel für die Entwicklung von Open-Source-Software empfohlen.

Big-Data-Anwendungen können durch die Verfügbarkeit von relevanten Daten beschleunigt werden. Allerdings kann die Sammlung von Daten mit erheblichen Kosten verbunden sein. Um die kontinuierliche Erhebung wertvoller Daten zu gewährleisten, ist es wichtig, ein angemessenes Kosten-Nutzen-Verhältnis sicherzustellen. Wenn also die Datenerhebung mit erheblichen Kosten verbunden ist, einschliesslich der Kosten für die Einhaltung der FAIR-Richtlinien¹⁸, ist es wichtig, Anreize für die Datenerhebung zu schaffen, z. B. durch die Möglichkeit einer verzögerten Veröffentlichung, um Datensammler*innen eine anfängliche Wertschöpfung zu ermöglichen, bevor sie die Daten zur Verfügung stellen müssen (ähnlich der Rolle der Patentierung als Mittel zur Förderung von Investitionen in Erfindungen), durch die Sicherstellung, dass Förderorganisationen die Kosten für die Einhaltung der FAIR-Richtlinien übernehmen, die durchaus über das Ende eines Projekts hinausgehen können, und durch die Forderung nach einer Datenveröffentlichung als Voraussetzung für die Veröffentlichung zugehöriger Forschungsergebnisse. Es braucht einen differenzierten Ansatz für die geteilte Nutzung von Daten, der anerkennt, dass nicht alle Daten gleich wertvoll sind, damit die Ressourcen sinnvoll eingesetzt werden.

Darüber hinaus können Anwendungen auch durch Benchmarks beschleunigt werden, die anonymisierte Datensätze und Anwendungsfälle umfassen, die gängige Szenarien in den angestrebten Anwendungsbereichen darstellen. Solche Benchmarks können als Referenz für die Entwicklung und das Testen von Anwendungen sowie für die Verbesserung der Genauigkeit und der Vorhersagefähigkeit von Algorithmen dienen. Sie können eine effektivere Einführung, Prüfung und Validierung von Big-Data-Anwendungen in grossem Massstab ermöglichen.

Schliesslich kann die beschleunigte Verfügbarkeit von Open-Source-Software die Wertschöpfung aus Daten beschleunigen. Es wird daher empfohlen, Anreize für die geteilte Nutzung von Tools als Open-Source-Software zu schaffen. Zum Beispiel sollte der Einfluss auf die Gesellschaft durch die Entwicklung von Open-Source-Software mit breiter Akzeptanz gleichrangig mit dem wissenschaftlichen Einfluss, der durch Zitate gemessen wird, bewertet werden. Dies würde nicht nur eine wichtige öffentliche Dienstleistung fördern, sondern birgt auch das Potenzial, internationale Talente in die Schweiz zu locken und insgesamt ein wichtiges Element für die Digitalisierung der Schweiz zu sein.

¹⁸ Die FAIR-Prinzipien sind eine Sammlung von Auflagen, um Forschungsdaten und andere digitale Objekte auffindbar, zugänglich, interoperabel und wiederverwendbar zu machen. Siehe <https://www.go-fair.org/fair-principles> sowie die Nationale Schweizer Strategie für Open Research Data, swissuniversities (2021)

Aktualisierung und Schaffung einer angemessenen Regulierung

7

Proaktivere Regulierungen von Big Data anstreben

Während Big-Data-Technologien in rasantem Tempo weiterentwickelt und eingesetzt werden, steckt die Regulierung von Big Data noch in den Kinderschuhen und hinkt der technologischen Entwicklung weit hinterher. Die fehlende Regulierung kann negative Auswirkungen haben, u. a. auf die Demokratie oder die psychische Gesundheit der Bevölkerung, z. B. aufgrund der Funktionsweise von Social-Media-Plattformen, auf den Wettbewerb und die Innovation, z. B. aufgrund von unlauteren Vorteilen und verzerrtem Wettbewerb. Da die Regulierung eine Schlüsselrolle bei der Vermeidung solcher negativen Auswirkungen spielt und das Potenzial hat, eine verbesserte Wertschöpfung durch Big Data zu ermöglichen, wird empfohlen, dass übergreifende Anstrengungen unternommen werden, um die Regulierungsprozesse zu beschleunigen.

Anwendungen von Big Data haben tiefgreifende und weitreichende Auswirkungen auf die Gesellschaft. Eine Regulierung kann die verantwortungsvolle Wertschöpfung beschleunigen und gleichzeitig die negativen Auswirkungen begrenzen. Ein proaktiverer Ansatz bei der Regulierung vermag daher eine verantwortungsvolle Wertschöpfung weiter zu fördern, Wettbewerb und Innovation zu erleichtern und der Demokratie besser zu dienen.

Die Big-Data-Kluft – das asymmetrische Verhältnis zwischen denjenigen, die grosse Datenmengen sammeln, speichern und analysieren, und denjenigen, die von der Datenerhebung betroffen sind – ist eine unvermeidliche Folge einer Gesellschaft, die Freiheit und Vielfalt schätzt. Anstatt zu versuchen, diese Kluft zu beseitigen, wird empfohlen, dass die Gesetzgebung realistische Schäden benennt, die aus der Big-Data-Kluft resultieren könnten, und rechtliche Schutzmassnahmen für diejenigen entwickelt, die benachteiligt werden.

Erfolgreiche Big-Data-Anwendungen setzen Vertrauen und Akzeptanz voraus. Bei der Schaffung von Rahmenbedingungen, unter denen Daten gesammelt, analysiert und genutzt werden können, sollte nicht nur auf die Schaffung von (selbstregulierenden) Standards gesetzt werden, die Interessen von Unternehmen und Konsument*innen ausgleichen, sondern die Konsument*innen sollten auch in die Lage versetzt werden, informierte Entscheidungen zu treffen.

Insgesamt ist es von grosser Bedeutung, rechtliche Schutzmassnahmen zu entwickeln, um den durch Big Data verursachten Schaden zu kompensieren, indem Standards für die Erhebung, die Weitergabe und die Analyse von Daten festgelegt werden, die den Schutz von vulnerablen Gruppen ermöglichen, die durch den Einsatz von Big-Data-Technologien potenziell benachteiligt sind.

8

Datenschutz und digitale Souveränität in Big-Data-Anwendungen fördern

Der Einsatz von Big-Data-basierten Anwendungen birgt Risiken für den Datenschutz und die damit verbundenen Rechte der Einzelnen. Auch wenn in der Schweiz (neues Datenschutzgesetz) und in der EU (Datenschutz-Grundverordnung DSGVO) ein grundlegender rechtlicher Rahmen vorhanden ist, ist die Einhaltung der geltenden Vorschriften oft eine schwierige Herausforderung. Es wird empfohlen, das Bewusstsein von Datenwissenschaftler*innen sowie von Dateningenieur*innen, Dateneigentümer*innen und Datenschutzbeauftragten für Fragen des Datenschutzes zu schärfen, umfassende Datenschutzstandards auszuarbeiten und der Sicherheit digitaler Infrastrukturen mehr Aufmerksamkeit zu schenken.

Regulierungen, die auf der Souveränität von Staaten über begrenzte physische Räume beruhen, greifen zu kurz, wenn es um Big Data geht. Stattdessen ist eine internationale Koordination und Zusammenarbeit erforderlich, um die Sicherheit digitaler Infrastrukturen sowie die Privatsphäre und andere datenbezogene Rechte der Bürger*innen zu schützen. Nichtsdestotrotz können und sollten auch auf nationaler Ebene Anstrengungen unternommen werden, um den Schutz der Daten zu gewährleisten. Nationale und kantonale politische Entscheidungsträger*innen und Verwaltungen sowie Universitäten und Wissenschaftler*innen sind gefordert, den nationalen Rechtsrahmen zu stärken und zu ergänzen.

Eine nationale Agenda für Datenschutz und Datenrechte umfasst zahlreiche Akteure und Themen. Daher ist es wichtig, die Schaffung starker Verbindungen zwischen den vielen Akteur*innen aus unterschiedlichen Disziplinen zu fördern, die an der Realisierung von Big-Data-Anwendungen beteiligt sind. Darüber hinaus sollte eine Methodik entwickelt werden, um bewährte Verfahren zur Erfassung und Anonymisierung von Daten, zur sicheren Speicherung von Daten und zur Gewährleistung einer datenschutzfreundlichen Wertschöpfung festzulegen. Bei hochsensiblen Daten scheint zum Beispiel die Entwicklung von Datenklaven eine wertvolle Option zu sein, um die Privatsphäre zu gewährleisten. Ausserdem sollte das bestehende Konzept der informierten Einwilligung durch spezifische Schutzmechanismen ergänzt werden.

Verschiedene Techniken zur Wahrung der Privatsphäre werden in der Forschung behandelt, aber ihr realer Einsatz steht noch aus. Einige von ihnen könnten für eine nationale Strategie zum Schutz der Privatsphäre und der datenbezogenen Rechte in Betracht gezogen werden, wie z. B. die Ernennung von Datentreuhänder*innen, die persönliche Daten von Einzelpersonen schützen, oder die Einführung von Fairnesskriterien im Zusammenhang mit Big-Data-Analysen, um Diskriminierung zu vermeiden. Zur Erleichterung des täglichen Umgangs mit Datenschutzfragen sollten weitere Datenschutzrichtlinien entwickelt werden, aber auch ein Kompetenzzentrum wie das Nationale Zentrum für Cybersicherheit NCSC oder eine Abteilung desselben mit einem öffentlichen Dienst zur Behandlung von Rechtsfragen rund um den Datenschutz beim Einsatz von Big Data könnte eingerichtet werden.

9

Verstärkte länderübergreifende Vereinheitlichung von Regulierungen

Daten fließen oft über Grenzen hinweg, der Datenzugriff aus dem Ausland und die internationale Nutzung von Big-Data-basierten Diensten sind weit verbreitet. Daher ist eine rein nationale Perspektive bezüglich Anwendung und Regulierung von Big-Data nicht ausreichend. Vielmehr ist es notwendig, international zu beobachten und zu handeln. Aufgrund der zahlreichen internationalen Organisationen, die ihren Sitz in der Schweiz haben, befindet sich die Schweiz in einer einzigartigen Position, um die Harmonisierungsaktivitäten der transnational ausgerichteten Institutionen zu unterstützen. Die Schweiz hat die Möglichkeit, ihr Engagement und ihre Expertise sowohl in internationalen Organisationen als auch in der nationalen Gesetzgebung unter Beweis zu stellen.

Die Globalisierung der Datenströme und der zunehmende Einsatz von Big-Data-basierten Anwendungen machen es notwendig, harmonisierte grenzüberschreitende Regelungsrahmen zu schaffen. Ein solcher Rahmen ist beispielsweise für den internationalen Handel von Bedeutung: Während die Verhandlungen im Rahmen der Welthandelsorganisation (WTO) noch andauern, regeln bilaterale und regionale (präferenzielle) Handelsabkommen zunehmend den Handel mit digitalen Waren und Dienstleistungen sowie den Datenfluss. Die jeweiligen neuen Regeln umfassen häufig Aspekte des Datenschutzes, der Cybersicherheit und des Geschäftsgeheimnisses.

Während sich die Schweiz aktiv an den Verhandlungen beteiligt, wird weitere Unterstützung empfohlen. Ebenso ist es sehr empfehlenswert, dass die Schweiz ihren Beitrag zur laufenden Entwicklung der OECD-Leitlinien für verantwortungsvolles unternehmerisches Handeln (RBC) leistet. Schliesslich hat die Schweiz eine wichtige Rolle als Förderin des UN Internet Governance Forum (Genf) gespielt. Angesichts der zunehmenden Spannungen in der digitalen Welt wird empfohlen, dass die Schweiz Anstrengungen unternimmt, die dazu beitragen können, eine Fragmentierung bei der Regulierung der daten-getriebenen Wirtschaft zu vermeiden.

Anhang:

Das Nationale Forschungsprogramm «Big Data» (NFP 75)

www.nfp75.ch

Grundlagen

Ablauf

2014

Vorschlag für ein Nationales Forschungsprogramm (NFP) zu Big Data

2015

Mandat des Bundesrates an den Schweizerischen Nationalfonds zur Durchführung des NFP 75

2015

Programmausschreibung und Selektion der Projekte

2017–2021

Forschungsphase

2022

Synthesearbeiten und Verbreitung der Ergebnisse

2023

Veröffentlichung des Résumés des NFP 75

Zahlen

Budget

CHF 25 Millionen

Projekte

34 Forschungsprojekte und 3 Querschnitt-Aktivitäten

Organisation

Leitungsgruppe des NFP 75

Prof. Dr. Christian S. Jensen

Fakultät für Informatik, Universität Aalborg (Präsident)

Prof. Dr. Sihem Amer-Yahia

CNRS, Laboratoire d'Informatique de Grenoble LIG, Université Grenoble Alpes UGA (ab 12.07.2016)

Prof. Dr. Sabrina de Capitani di Vimercati

Fakultät für Informatik, Universität Mailand

Prof. Dr. Friedrich Eisenbrand

Institut für Mathematik, EPFL (ab 01.01.2021)

Prof. Dr. Jörg Huelsken

Schweizerisches Institut für experimentelle Krebsforschung ISREC, EPFL

Prof. em. Dr. Erkki Oja

Fakultät für Informatik, Universität Aalto

Prof. Dr. Reinhard Riedl

Institut Digital Technology Management, Berner Fachhochschule

Prof. Dr. Caroline Sporleder

Institut für Digitale Geisteswissenschaften, Georg-August-Universität Göttingen (bis 31.12.2019)

Prof. Dr. Rolf H. Weber

Juristische Fakultät, Universität Zürich

Delegierte der Abteilung Programme des Nationalen Forschungsrats für das NFP 75

Prof. Dr. Bert Müller

Biomaterials Science Center, Universität Basel (ab 01.01.2021)

Prof. Dr. Friedrich Eisenbrand

Institut für Mathematik, EPFL (bis 31.12.2020)

Programm-Manager NFP 75

Boris Buzek

Schweizerischer Nationalfonds, Bern (ab 01.11.2022)

Dr. Stefan Husi

Schweizerischer Nationalfonds, Bern (von 01.11.2020 bis 31.10.2022)

Dr. Christian Mottas

Schweizerischer Nationalfonds, Bern (bis 31.10.2020)

Bundesvertreter im NFP 75

Dr. Uwe Heck

Bundeskanzlei, Bereich Digitale Transformation und IKT-Lenkung (seit 01.01.2019)

Willy Müller

Informatik-Lenkungsstelle des Bundes (bis 31.12.2018)

Leiterin Wissenstransfer

Beatrice Huber

Schweizerische Akademie der Technischen Wissenschaften (SATW), Zürich (ab 01.12.2018)

Dr. Béatrice Miller

Schweizerische Akademie der Technischen Wissenschaften (SATW), Zürich (bis 30.11.2018)

Die 34 Forschungsprojekte

Modul 1: Big-Data-Infrastrukturen

Datenströme: Überwachung in Echtzeit
David Basin, Dmytro Traytel, ETH Zürich
Big data monitoring

Datenstromanalytik: Tools zur schnellen und datenschutzgerechten Verarbeitung
Michael Böhlen, Universität Zürich
Privacy preserving, peta-scale stream analytics for domain-experts

Modelle maschinellen Lernens: Robustheit und Verallgemeinerung
Volkan Cevher, EPFL
Theory and methods for accurate and scalable learning machines

Rechenzentren: effiziente Leistungsüberwachung
Lydia Yiyu Chen, Technische Universität Delft, Niederlande (davor IBM Research Zurich)
Dapprox: dependency-ware approximate analytics and processing platforms

Wenig strukturierte Daten: neue Techniken für die Datenintegration
Philippe Cudré-Mauroux, Universität Freiburg
Tighten-it-all: big data integration for loosely-structured data

Sprachmodelle: neue Methoden für dialogfähige Agenten
Thomas Hofmann, ETH Zürich
Conversational agent for interactive access to information

Digitale Zwillingstädte: 3D-Modelle aus einem scannenden Auto
Frédéric Kaplan, EPFL
ScanVan – a distributed 3d digitalization platform for cities

Coresets: Big Data mit weniger Daten
Andreas Krause, ETH Zürich
Scaling up by scaling down: big ML via small coresets

Scala-Programmiersprache: Big-Data-Analytik ermöglichen
Martin Odersky, EPFL
Programming language abstractions for big data

In-Network-Computing: Lösungen für Graph Analytics
Robert Soulé, Università della Svizzera italiana
Exploratory visual analytics for interaction graphs

Schnelle Vorhersagealgorithmen
Marco Zaffalon, Istituto Dalle Molle di Studi sull'Intelligenza Artificiale USI-SUPSI
State space Gaussian processes for big data analytics

Graph Analytics und Mining
Willy Zwaenepoel, University of Sydney (davor EPFL)
Building flexible large-graph processing systems on commodity hardware

Modul 2: Gesellschaftliche und regulatorische Herausforderungen

Handelsabkommen: Auswirkungen auf das nationale Recht
Mira Burri, Universität Luzern
The governance of big data in trade agreements: design, diffusion and implications

Big Data im Versicherungswesen
Markus Christen, Universität Zürich
Between solidarity and personalization – Dealing with ethical and legal big data challenges in the insurance industry

Regulierung der Big-Data-gestützten Forschung
Bernice Simone Elger, Universität Basel
Ethical and legal regulation of big data research – Towards a sensible and efficient use of electronic health records and social media data

Rechtliche Herausforderungen von Big Data
Sabine Gless, Universität Basel
Legal challenges in big data. Allocating benefits. Averting risks

Unsicherheit in Big-Data-Anwendungen: Lehren aus Klimasimulationen
Reto Knutti, ETH Zürich
Combining theory with big data? The case of uncertainty in prediction of trends in extreme weather and impacts

Big Data in der Praxis: Soziologie, Datenwissenschaften und Datenjournalismus
Sophie Mützel, Universität Luzern
Facing big data: methods and skills needed for a 21st century sociology

Big Data im Gesundheitsbereich: ein ethisches Rahmenkonzept
Effy Vayena, ETH Zürich
BEHALF – Bigdata-ethics-health framework

Big Data im HR-Bereich
Antoinette Weibel, Universität St. Gallen
Big data or big brother? Big data HR control practices and employee trust

Modul 3: Big-Data-Anwendungen

Optimierung der Verkehrssysteme: anonymisierte individuelle Mobilitätsspuren
Kay W. Axhausen, ETH Zürich
Big data transport models: the example of road pricing

Pig Data: Analytik für die Schweizer Schweinehaltung
John Berezowski, Universität Bern
Pig data: health analytics for the Swiss swine industry

Hochwassererkennung: automatisches Geotagging von Crowdsourced-Videos
Susanne Bleisch, FHNW
EVAC – Employing video analytics for crisis management

Computergestützte Chemie: die Entdeckung neuer Moleküle
Helmut Harbrecht, Universität Basel
Big data for computational chemistry: unified machine learning and sparse grid combination technique for quantum based molecular design

Intensivstationen: ein automatisiertes Alarmsystem
Emanuela Keller, Universitätsspital Zürich
ICU-cockpit: IT platform for multimodal patient monitoring and therapy support in intensive care and emergency medicine

Evidenzbasierte Strategie: Kausalität aus Daten entdecken
Michael Lechner, Universität St. Gallen
Causal analysis with big data

Mapping globaler Innovation: Analyse von Patenten

Alessandro Lomi, Università della Svizzera italiana

The global structure of knowledge networks: data, models and empirical results

Big Data Genetik: leistungsstarke Indexierung

Gunnar Rätsch, ETH Zürich

Scalable genome graph data structures for metagenomics and genome annotation

Rückenschmerzen: eine personalisierte Smartphone-basierte Lösung

Robert Riener, ETH Zürich, und Walter Karlen, Universität Ulm (davor ETH Zürich)

Personalized management of low back pain with mHealth: big data opportunities, challenges and solutions

Bodenerosion: Quantifizierung durch Luftaufnahmen in der Schweiz

Volker Roth, Universität Basel

WeObserve: integrating citizen observers and high throughput sensing devices for big data collection, integration, and analysis

Genomvergleich: schnellere Analyse

Nicolas Salamin, Universität Lausanne

Efficient and accurate comparative genomics to make sense of high-volume low-quality data in biology

Potenzial erneuerbarer Energien: Bewertung für die Schweiz

Jean-Louis Scartezzini, EPFL

Hybrid renewable energy potential for the built environment using big data: forecasting and uncertainty estimation

Bioinformatikdatenbanken: Abfragen in natürlicher Sprache

Kurt Stockinger, ZHAW

BIO-SODA: enabling complex, semantic queries to bioinformatics databases through intuitive searching over data

Sonneneruptionen: geomagnetische Stürme vorhersagen

Svyatoslav Voloshynovskiy, Universität Genf

Machine learning based analytics for big data in astronomy

Die 3 Querschnitt-Aktivitäten

Big Data: Offene Daten und rechtliche Herausforderungen

Sabine Gless, Universität Basel

ELSI-Task-Force für das nationale Forschungsprogramm «Big Data»

Markus Christen, University of Zurich

Women in Big Data

Lydia Yiyu Chen, Technische Universität Delft, Niederlande (davor IBM Research Zurich)

Veröffentlichungen und Lehrmaterial

Ethical, legal and social issues of big data – A comprehensive overview

Eleonora Viganò (Ed.), NRP 75 (2022)

Big Data: Lehrmittel für die Sekundarstufen, NFP 75 und Museum für Kommunikation, Bern (2020)

Big data ethics recommendations for the insurance industry, NRP 75 (2019)

Impressum

Dieses Résumé des Nationalen Forschungsprogramms Big Data» (NFP 75) fasst die Ergebnisse der 37 Projekte des NFP 75 zusammen und integriert sie in einen Überblick über die Chancen und Herausforderungen von Big Data. Die Autor*innen haben diese Ergebnisse synthetisiert und dabei ihr wissenschaftliches Fachwissen und ihre Erfahrung eingebracht. Im Anschluss hat ein Wissenschaftsjournalist die Texte zusammengefasst und redigiert. Die Schlussfolgerungen, die am Ende des Dokuments präsentiert werden, sind das Ergebnis eines kollektiven mehrstufigen Prozesses: Sie wurden von den Mitgliedern der Leitungsgruppe verfasst, diskutiert und zusammengefasst. Sie stellen den kollektiv verabschiedeten Konsens der Leitungsgruppe dar.

Das Résumé ist als wissenschaftlicher Beitrag zum Prozess der Meinungsbildung, zur politischen und fachlichen Debatte sowie zur Planung von Strategien und Massnahmen für den politischen und gesellschaftlichen Wandel mit und durch Big Data zu verstehen. Der Text liegt in der kollektiven Verantwortung der Leitungsgruppe. Ihre Einschätzungen und Schlussfolgerungen spiegeln nicht unbedingt die der Forschungsteams oder des Schweizerischen Nationalfonds wider. Weitere Informationen zu allen im Résumé genannten Forschungsprojekten des NFP «Big Data» finden Sie auf der Website www.nfp75.ch.

Autor*innen

Prof. Dr. Sihem Amer-Yahia, Université Grenoble Alpes UGA

Prof. Dr. Jörg Huelsken, EPFL

Prof. Dr. Christian S. Jensen, Universität Aalborg

Prof. em. Dr. Erkki Oja, Universität Aalto

Prof. Dr. Reinhard Riedl, Berner Fachhochschule

Prof. Dr. Rolf H. Weber, Universität Zürich

Redaktion

Dr. Daniel Saraga, Saraga Communications, Basel

Koordination

Boris Buzek, Schweizerischer Nationalfonds, Bern

Dr. Stefan Husi, Schweizerischer Nationalfonds, Bern

Beatrice Huber, Leiterin Wissenstransfer NFP 75

Produktion

Übersetzung aus dem Englischen: STP Language Services, Stäfa

Layout und Illustrationen: Gabriel Alber, Alber Visuelle Kommunikation, Zürich

Druck: Druckerei Herzog AG, Langendorf (SO)

Herausgeber

Leitungsgruppe des Nationalen Forschungsprogramms «Big Data» (NFP 75)

Vorgeschlagene Zitierweise

Leitungsgruppe des NFP 75 (2023): «Big Data: Anwendungen, Technologien und gesellschaftliche Aspekte, Résumé des Nationalen Forschungsprogramms «Big Data» (NFP 75)», Schweizerischer Nationalfonds, Bern.

Nationale Forschungsprogramme (NFP) sind ein Förderinstrument des Schweizerischen Nationalfonds (SNF). Ihr Ziel ist es, Forschung zu ermöglichen, die zur Lösung von aktuellen Problemen von nationaler Bedeutung beiträgt. Das Nationale Forschungsprogramm «Big Data» (NFP 75) lief von 2015 bis 2022.

ISBN: 978-3-907087-63-3

© 2023, Schweizerischer Nationalfonds, Bern

