

16.11 Einzelhandel	532
16.11.1 Vision	532
16.11.2 Daten	533
16.11.3 Anwendungsfälle	533
16.11.4 Herausforderungen	534
16.12 Anbieter von Telekommunikation	534
16.12.1 Vision	535
16.12.2 Daten	535
16.12.3 Anwendungsfälle	535
16.12.4 Herausforderungen	537
16.13 Transport	538
16.13.1 Vision	538
16.13.2 Daten	539
16.13.3 Anwendungsfälle	539
16.13.4 Herausforderungen	539
16.14 Lehre und Ausbildung	540
16.14.1 Vision	540
16.14.2 Daten	541
16.14.3 Anwendungsfälle	542
16.14.4 Herausforderungen	542
16.15 Die digitale Gesellschaft	543
16.16 Kurz und bündig	545
17 Mindset und Community	546
17.1 Data Driven Mindset	546
17.2 Data-Science-Kultur	549
17.2.1 Start-up oder Beratungsunternehmen?	549
17.2.2 Labs statt Konzernpolitik	550
17.2.3 Keiretsu statt Einzelkämpfertum	551
17.2.4 Agile Softwareentwicklung	552
17.2.5 Firmen- und Arbeitskultur	553
17.3 Antipatterns	556
17.3.1 Abwertung von Fachwissen	556
17.3.2 Die IT wird es schon richten	557
17.3.3 Widerstand gegen Veränderungen	558
17.3.4 Besserwisser-Mentalität	558
17.3.5 Schwarzmalerei	559
17.3.6 Pfennigfuchseriei	560
17.3.7 Angstkultur	560
17.3.8 Kontrolle über die Ressourcen	561
17.3.9 Blindes Vertrauen in die Ressourcen	561
17.3.10 Das Schweizer Taschenmesser	562
17.3.11 Over-Engineering	563
17.4 Kurz und bündig	564

18	Vertrauenswürdige KI	565
18.1	Rechtlicher und Soft-Law-Rahmen	566
18.1.1	Normen	568
18.1.2	Verordnungen	569
18.2	KI-Stakeholder	571
18.3	Fairness in der KI	572
18.3.1	Bias	573
18.3.2	Fairness-Metriken	576
18.3.3	Unerwünschten Bias in KI-Systemen reduzieren	580
18.4	Transparenz von KI-Systemen	581
18.4.1	Dokumentieren der Daten	582
18.4.2	Dokumentieren des Modells	584
18.4.3	Explainability (Erklärbarkeit)	585
18.5	Schlussfolgerung	587
18.6	Kurz und bündig	587
19	Die Autor:innen	588
Index	593

Geleitwort

„Mathematical science shows what is. It is the language of unseen relations between things. But to use and apply that language, we must be able to fully appreciate, to feel, to seize the unseen, the unconscious.“

Ada Lovelace

So wie die Computerkompetenz vor einer Generation eine neue Reihe von grundlegenden Fähigkeiten darstellte, die es zu erwerben galt, so stellt die Kompetenz im Bereich der künstlichen Intelligenz (KI) für unsere heutigen Generationen und darüber hinaus das Gleiche dar. In den letzten zwei Jahrzehnten hat sich die Datenwissenschaft zur mathematischen Architektur und zu der entsprechenden Sprache entwickelt, mit der wir Systeme aufbauen und in der wir mit ihnen interagieren, die unsere Sinne und Entscheidungsfähigkeiten erweitern. Es reicht also nicht mehr aus, Befehle per Mausklick an Computer zu senden, sondern es ist von entscheidender Bedeutung, dass wir in der Lage sind, KI-gestützte Empfehlungen von Computern zu interpretieren und mit ihnen zu interagieren. Derzeit verarbeiten Maschinen, d. h. Computer, die mit Sensoren (im weitesten Sinne) gekoppelt sind, ein immer breiteres Spektrum an Daten, darunter Texte, Bilder, Videos, Audiodaten, Netzwerkdiagramme und eine Vielzahl von Informationen aus dem Internet, der Privatwirtschaft und dem öffentlichen Sektor. In Anbetracht der Datenvielfalt nähern sich die Autoren dieses Buches der Datenwissenschaft als einem grundlegenden Schlüsselthema für die Gesellschaft. Sie tun dies mit großer Einsicht, aus mehreren wichtigen Blickwinkeln und in einem unterhaltsamen Stil, der bei Anfängern und Experten gleichermaßen Anklang findet.

Aus Daten Nutzen zu ziehen ist wohl das verbindende Ziel des Wissensarbeiters des 21. Jahrhunderts. Sogar Berufsbereiche, die als klassisch datenfern galten, wie Verkauf und Kunst, haben jetzt datengetriebene Teilbereiche wie Marketingautomatisierung und Computergestaltung. Zum Nutzen der Leser bringen die Autoren Erfahrungen aus erster Hand und sorgfältige Recherchen ein, um überzeugend darzulegen, welche Rolle wir alle spielen müssen, wenn wir versuchen, Daten für bessere Ergebnisse zu nutzen. In der Tat ist die Bandbreite, die in diesem Werk vermittelt wird, beeindruckend. Sie reicht von Überlegungen zur Hardwareleistung (z. B. CPU, Netzwerk, Speicher, I/O, GPU) bis hin zu den verschiedenen Rollen von Teammitgliedern beim Aufbau von Maschinen, die Muster in Daten finden können. Darüber hinaus gehen die Autoren auf die Möglichkeiten ein, mit denen Maschinen heute sehen und lesen können, nämlich Computer Vision und Natural Language Processing, was tiefgreifende Auswirkungen auf fast alle Industriebereiche hat.

Bei der Lektüre dieses Buches möchte ich Sie ermutigen, neugierig zu sein und sich eine Reihe von Fragen zu stellen, wie Ihr beruflicher Werdegang und die Gesellschaft, wie Sie sie sehen, gegenwärtig von immer fortschrittlicheren Maschinen beeinflusst werden: von den Möglichkeiten, die Ihr Smartphone bietet, bis hin zu der Art und Weise, wie Arbeits-

plätze auf dem Markt durch Automatisierungstools umgestaltet werden. Hier sind einige Fragen, die Ihnen den Einstieg erleichtern sollen:

- Wie verschiebt sich das Verhältnis der Aufgaben, mit denen Sie Ihre Zeit verbringen, mit dem Aufkommen von immer fortschrittlicheren Maschinen in Ihrem Arbeitsbereich?
- Was bedeutet es, wenn Maschinen über Wahrnehmungsfähigkeiten verfügen, die denen des Menschen entsprechen, also sehen, hören, riechen, schmecken, tasten und mehr?
- Wie gehen wir als Gesellschaft mit der Voreingenommenheit und dem Vertrauen in Daten um?
- Wie können wir den Bau und die Nutzung von Maschinen, die lernen, inklusiver gestalten?
- Welche eindeutig menschlichen Fähigkeiten können Sie hervorheben, um Organisationen, die Ihnen am Herzen liegen, zu mehr Wettbewerbsfähigkeit und Nachhaltigkeit zu verhelfen?

Ich habe mich davor gehütet, den Begriff „denkende Maschinen“ oder „künstliche allgemeine Intelligenz“ zu verwenden, um eine Übertreibung zu vermeiden. Worauf ich Ihre Aufmerksamkeit lenken möchte, ist die breite Anwendbarkeit dessen, was wir in der Forschung rund um Maschinen mit Lernfähigkeiten sehen. Aus meiner Zeit in den Labors der Universitäten Columbia und Cornell, dem Princeton Plasma Physics Laboratory, der American University of Armenia und dem von der NASA unterstützten TRISH (Translational Research Institute for Space Health), das mit TrialX zusammenarbeitet, ist mir klar, dass Maschinen in einem enorm breiten Spektrum von Bereichen Muster in Daten finden und Menschen sowohl in normalen als auch in missionskritischen Kontexten alarmieren können. Die Auswirkungen auf die menschliche Erfahrung sind also vielschichtig, und Data Scientists spielen eine wichtige Rolle bei der Entwicklung von Systemen, bei denen die menschliche Interaktion mit dem maschinellen Output eine positive Summe ergibt. Ich kann nicht genug betonen, dass ein Nullsummen-Ansatz bei der Automatisierung suboptimal ist. Unternehmer neigen jedoch dazu, einen Weg zur maximalen Summe zu finden.

Gemeinsam mit Kollegen und durch meine Arbeit beim BAJ Accelerator und Covenant Venture Capital unterstützte ich Start-ups bei einer Art Tandem-Lernen: wie ein schnell wachsendes Unternehmen eine Branche umgestalten kann, indem es Marktlücken aufspürt, und wie die Erfindung eines Unternehmens lernen und neue Fähigkeiten für Kunden bereitstellen kann. In dem leistungsstarken Technologiebereich der Computer Vision, der eine tragende Säule der Datenwissenschaft ist, stechen beispielsweise drei Unternehmen hervor, die in drei sehr unterschiedlichen Industriebereichen bahnbrechend sind: Embodied, Scylla und cognaize in den Bereichen Gesundheitswesen, Sicherheit bzw. Finanzen.

- Das Vorzeigeprodukt von Embodied, Moxie, ist ein Roboter, der das emotionale Wohlbefinden und die soziale Entwicklung von Kindern unterstützt. Um dies zu erreichen, muss Moxie die Familienmitglieder auf überzeugende Weise sehen und mit ihnen kommunizieren. Er muss den emotionalen Zustand der Menschen, mit denen er interagiert, sowohl visuell als auch durch andere Hinweise verstehen, um einen sinnvollen Dialog führen zu können. Die Gesundheitsdienstleister haben also ein neues robotisches Teammitglied, mit dem sie zusammenarbeiten können. Embodied war auf der Titelseite des TIME Magazine zu sehen.

- Scylla ermöglicht es dem Sicherheitsteam eines Unternehmens, die Sicherheit proaktiv zu verbessern. Mit Echtzeit-Erkennungsfunktionen müssen Kameranetzwerke nicht mehr passiv sein, sondern können proaktiv eingesetzt werden. Die Anwendungsmöglichkeiten sind vielfältig und reichen von der Erkennung von Ausrutschern und Stürzen in Krankenhäusern und Stadien, um die Gesundheitsergebnisse zu verbessern, bis hin zur Alarmierung von Eindringlingen in Produktionsstätten und Bürogebäuden, um die Mitarbeiter besser zu schützen. Scylla wurde bereits in Forbes vorgestellt.
- cognaize unterstützt Finanzinstitute und Versicherungsunternehmen bei der Verarbeitung einer enormen Menge unstrukturierter Daten zur Risikobestimmung. Eine wichtige Erkenntnis ist, Dokumente nicht nur als Text zu betrachten, sondern auch visuelle Informationen zu berücksichtigen: Stil, Tabellen, Struktur. Darüber hinaus verfügt cognaize über einen „Human-in-the-Loop“, bei dem Kollegen und das System insgesamt kontinuierlich lernen. cognaize wurde auf dem NASDAQ-Bildschirm am Times Square vorgestellt.

In den drei oben genannten Beispielen für aufstrebende Unicorn-Start-ups arbeiten Data Scientists eng mit Ingenieuren, Analysten, Designern, Inhaltentwicklern, Fachleuten und Kunden zusammen, um Maschinen zu entwickeln, die lernen und auf nuancierte Weise mit Menschen interagieren. Das Ergebnis ist ein Wandel in der Art der Arbeit: Menschen werden auf die wichtigsten Dokumente oder Momente aufmerksam gemacht, und aus der menschlichen Erfahrung wird gelernt, um die Qualität zu verbessern. Dies steht stellvertretend für einen neuen Wandel, der KI-Kenntnisse voraussetzt: Arbeitsplätze in fast allen Bereichen der Wirtschaft werden Aspekte aufweisen, die eine maschinelle Interaktion erfordern: Menschen, die Korrekturen vornehmen, neue Fähigkeiten erlernen, auf Warnungen reagieren und diese interpretieren und eine schnellere Reaktionszeit haben, um anderen Menschen zu helfen, indem sie Maschinen zur Unterstützung einsetzen. In den kommenden Jahren bin ich gespannt auf die Rolle der Datenwissenschaft in der Schnittstellenforschung, auf neue Algorithmen und darauf, wie Menschen ihre Arbeit um ein Vielfaches verstärken können.

Als ich vor fast einem Jahrzehnt die erste Ausgabe von *The Field Guide to Data Science* mit verfasst habe, war es bemerkenswert, wie sehr sich die Disziplin weiterentwickelt hat, sowohl in Bezug auf das, was technisch erreicht wurde, als auch in Bezug auf das, was noch zu erreichen ist. Das Handbuch Data Science bringt die Disziplin in diesen beiden Dimensionen voran und trägt die Fackel weiter.

Lesen Sie weiter.

Herbst 2021

Armen R. Kherlopian, Ph.D.