



Blick in die Wissenschaft | 44 45

Forschungsmagazin der Universität Regensburg

INFORMATIK UND DATA SCIENCE

Lehre an der Fakultät für Informatik und Data Science

Automatisches Beweisen: Methoden und Anwendungen

Algorithmen und Komplexitätstheorie

Evolution in Datenbanken und Data Engineering Workflows

IoT-basiertes Prozessmanagement – Mobile Benutzerführung in der digitalen Fabrik

Cyber Threat Intelligence: Gemeinschaftliche IT-Sicherheit durch den Austausch von Informationen

Kann man den Entscheidungen Künstlicher Intelligenz trauen? Zu den Auswirkungen unsicherer Daten auf die Entscheidungen Neuronaler Netze

Mensch vs. Maschine: Wettbewerb und Kooperation mit künstlicher Intelligenz in digitalen Märkten

Notfallpläne für den Ernstfall testen

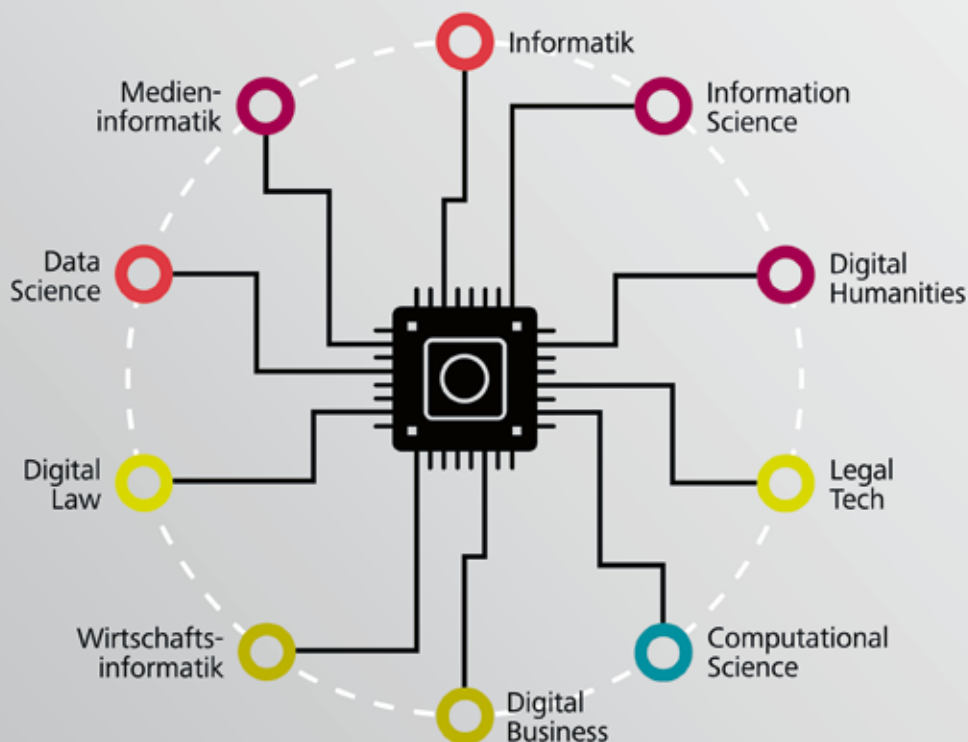
Maschinelles Lernen mit Anwendungen in den Naturwissenschaften

Automatisierte, KI-basierte Analyse von Bilddaten

Die Genome des Menschen – Forschungsschwerpunkte der Arbeitsgruppe für Algorithmische Bioinformatik

Heft 44/45 | 31. Jahrgang 2023 | € 14,00 | ISSN 0942-928-X

FIDS Bei uns wird Informatik buntgeschrieben.



© FIDS, Universität Regensburg



Die Spitzenregion
für exzellente
Wissenschaft
und innovative
Unternehmen

FORSCHUNG

2 Hochschulen, Zelltherapie
Immunologie, Onkologie

ENTWICKLUNG

80 Unternehmen im Bereich
BioTech, MedTech und Health IT

GESUNDHEITSWESEN

Über 20.000 Beschäftigte im
Gesundheitsbereich und mehr
als 2,7 Mrd. Euro Umsatz

SCHWERPUNKTE

Digitalisierung & KI
Pflegeinnovationen
Hygiene und Infektionsprävention

START-UP

BioPark Jump Accelerator Programm
und BioPark Brain & Network für
Gründer

Profitieren Sie vom
Netzwerk für Gesundheitswirtschaft
mit über 80 Unternehmen und Institutionen,
fünf Kliniken und zwei Hochschulen.

HEALTHCARE
REGENSBURG MANAGED BY BIOPARK



Liebe Leserinnen, liebe Leser,

wir freuen uns, Ihnen nach einer längeren, Corona-bedingten Pause eine neue Ausgabe des Forschungsmagazins ‚Blick in die Wissenschaft‘ in der Ausgabe 44/45 präsentieren zu können.

Die Corona-Pandemie hat auch die Universität Regensburg und alle ihre Mitglieder vor große Herausforderungen gestellt. Dennoch konnten zentrale Zukunftsprojekte weitergeführt und umgesetzt werden. So stellt vor allem die Gründung unserer neuen Fakultät für Informatik und Data Science (FIDS) einen wahren Meilenstein in der Geschichte und Entwicklung der Universität Regensburg dar. Als größtes Strukturprojekt seit der Gründung der Fakultät für Medizin vor 30 Jahren ist unsere Informatikfakultät ein Zukunftsprojekt von weitreichenden Dimensionen. Mit der neuen strategischen Schwerpunktsetzung im Bereich Informatik und Data Science und vor allem auch der Querschnittsorientierung der neuen Fakultät sieht sich die Universität Regensburg sehr gut gerüstet, ihre bisherigen Stärken in diesen Bereichen zu bündeln, weiter auszubauen und zu

vertiefen. Schließlich sind *Digital Transformations* als eines der vier Gestaltungsfelder und Zukunftsthemen in unserem *Hochschulentwicklungsplan 2025* fest verankert. Dieses Gestaltungsfeld adressiert die neue Fakultät ebenso wie den Bereich *Integrated Sciences in Life, Health, and Disease* als ein weiteres Schwerpunktgebiet unserer Universität.

Die Grundsatzbeschlüsse in den Gremien der Universität Regensburg im Sommer und Herbst 2019 zur Einrichtung der neuen Fakultät erfolgten nach einer vorhergehenden Phase intensiver Planungen dann letztlich fast zeitgleich mit der Regierungserklärung des Bayerischen Ministerpräsidenten Dr. Markus Söder am 10. Oktober 2019 und der Verkündung der Hightech Agenda Bayern. Unterstützt und beschleunigt durch die Mittel der Hightech Agenda Bayern konnte der Auf- und Ausbau der Fakultät für Informatik und Data Science zügiger umgesetzt werden, nachdem die neue Fakultät im März 2020 formal gegründet und im Laufe des WS 2021/22 aus sich heraus handlungs- und

funktionsfähig wurde. Im Mai 2022 konnten wir gemeinsam mit Ministerpräsident Dr. Markus Söder und Staatsminister für Wissenschaft und Kunst Markus Blume den offiziellen Kickoff für die Fakultät begehen. Dass dieser komplexe Prozess im Kontext der Herausforderungen der Corona-Pandemie vollzogen und abgeschlossen werden konnte, ist ein Zeichen für die Bedeutung dieser gesamtuniversitär-strategischen Maßnahme und für den Rückhalt für das Großprojekt in der universitären Gemeinschaft.

Im Laufe des Gründungsprozesses ist es gelungen, die verschiedenen Informatiknahen und -interessierten Kräfte der Universität an einen Tisch zu bringen und gemeinsam ein zukunftsorientiertes Konzept für die Fakultät zu entwickeln. Ein externes Gutachten mit hochrangiger Expertise skizzierte und evaluierte 2019 wesentliche inhaltliche Schwerpunkte und Strukturierungen für die neue Fakultät, an denen sich in den Jahren 2019-2021 die von Vizepräsident Prof. Dr. Nikolaus Korber geleitete Gründungskommission in der

konkreten Arbeit zum Aufbau der Fakultät orientierte. In insgesamt 15 Berufungsverfahren wurden die ersten neuen Professuren in der Fakultät zügig besetzt – ein Prozess, der in Kürze abgeschlossen sein wird. Im Besetzungsprozess hat sich vor allem auch gezeigt, wie attraktiv die Neugründung einer Fakultät und die Möglichkeiten zur Mitgestaltung und zum Aufbau neuer Strukturen für Wissenschaftlerinnen und Wissenschaftler sind und wie viel Zukunftspotential von unserer neuen Fakultät ausgeht. So konnten wir zum Wintersemester 2023/24 130 Studierende für den B.Sc. Informatik und den B.Sc. Data Science begrüßen.

Im vorliegenden Heft von ‚Blick in die Wissenschaft‘ möchten wir Ihnen nunmehr vor allem die Forschungsaktivitäten der Fakultät für Informatik und Data Science näher vorstellen. Dabei beglückwünsche ich die Fakultät, dass sie bereits eineinhalb Jahre nach ihrer vollständigen Handlungs- und Funktionsfähigkeit und während der weiteren Planungen zum Aufbau und der Ausarbeitung ihrer Studiengänge insbe-

sondere im Master-Bereich ein so vielfältiges Themenheft zu ihren aktuellen Forschungsarbeiten vorlegen konnte.

Das facettenreiche und vielfältige Themenspektrum dieses Sonderhefts illustriert, wie die Fakultät für Informatik und Data Science die an der Universität Regensburg bisher vorhandenen IT-Kompetenzen erfolgreich bündelt und in die Zukunft gerichtet erweitert. Sie ermöglicht die essentielle interdisziplinäre Vernetzung mit der gesamten Universität, von den Geistes- und Sozialwissenschaften bis zu den Natur- und Lebenswissenschaften. Die Beiträge verdeutlichen, wie interdisziplinäre Forschung das Fundament starker methodischer und fachlicher Grundlagen weiterentwickelt und wie die bisherigen Informatik-Schwerpunkte der Universität Regensburg (Computational Science, Informationswissenschaft, Medieninformatik, Wirtschaftsinformatik) erfolgreich in die neue Fakultät überführt werden konnten und die Querschnittsorientierung unterstützen.

Das vorliegende Heft mit seinem Schwerpunkt auf aktuellen Forschungsar-

beiten begleitet den im Wintersemester 2023/24 erfolgten Start der beiden grundständigen Bachelor-Studiengänge Informatik und Data Science. Ein kurzer Überblicksbeitrag zur Lehre in der neuen Fakultät zeigt anschaulich die bereits gewachsene Vielfalt der Informatikstudiengänge und die intensive und gelebte Verbindung von Forschung und Lehre auch an dieser neuen Fakultät.

Unsere neue Fakultät leistet hervorragende Arbeit und ich bin sicher, dass Ihnen die nachfolgenden Seiten einen spannenden Einblick in die verschiedenen Facetten der FIDS geben werden. Ein ganz besonderes Dankeschön möchte ich an dieser Stelle an das gesamte Dekanat der Fakultät für Informatik und Data Science und insbesondere an Forschungsdekanin Prof. in Dr. Meike Klettke richten, die für diese Sonderausgabe die Koordinationsarbeit der vorliegenden Ausgabe federführend übernommen hat.

Prof. Dr. Udo Hebel
Präsident der Universität Regensburg

**Blick in die Wissenschaft
Forschungsmagazin
der Universität Regensburg**

ISSN 0942-928-X

Heft 44/45

31. Jahrgang

Herausgeber

Prof. Dr. Udo Hebel
Präsident der Universität Regensburg

Redaktionsleitung für diese Ausgabe

Prof.in Dr. Meike Klettke / Fakultät für Informatik und Data Science

Redaktionsbeirat

Prof. Dr. jur. Christoph Althammer
Prof. Dr. rer. nat. Ferdinand Evers
Prof. Dr. rer. nat. Stefan Friedl
Prof. Dr. rer. nat. Mark W. Greenlee
Prof. Dr. theol. Andreas Merkt
Prof. Dr. phil. Omar W. Nasim
Prof. Dr. rer. nat. Klaus Richter
Prof. Dr. rer. pol. Daniel Rösch
Prof. Dr. med. Ernst Tamm
Prof. Dr. paed. Oliver Tepner
Prof. Dr. phil. Christiane Heibach

Universität Regensburg
93040 Regensburg
Telefon +49 941 9432300
Telefax +49 941 9433310

Verlag

Universitätsverlag Regensburg GmbH
Leibnizstraße 13, 93055 Regensburg

Telefon +49 941 78785-0
Telefax +49 941 78785-16

info@univerlag-regensburg.de
www.univerlag-regensburg.de
Geschäftsführer: Dr. Albrecht Weiland,
Felix Weiland M.A.

Abonnementsservice

bestellung@univerlag-regensburg.de

Anzeigenleitung

Larissa Nevecny
MME-Marquardt
info@mme-marquardt.de

Herstellung

Universitätsverlag Regensburg GmbH
info@univerlag-regensburg.de

**Einzelpreis € 7,00
Doppelheft € 14,00**

Jahresabonnement

bei zwei Ausgaben pro Jahr

€ 10,00 / ermäßigt € 9,00

Für Schüler, Studierende und Akademiker/innen im Vorbereitungsdienst (inkl. 7% MwSt.) zzgl. Versandkostenpauschale € 1,64 je Ausgabe. Bestellung beim Verlag. Für **Mitglieder des Vereins der Ehemaligen Studierenden der Universität Regensburg e.V.**, des **Vereins der Freunde der Universität Regensburg e.V.** und des **Vereins ehemaliger Zahnmedizinstudenten Regensburg e.V.** ist der Bezug des Forschungsmagazins im Mitgliedsbeitrag enthalten.

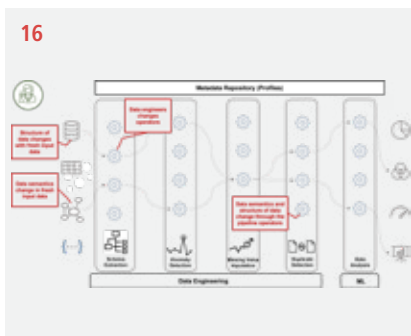
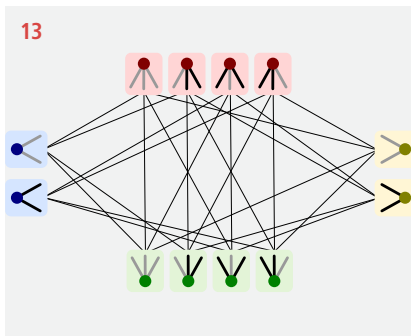
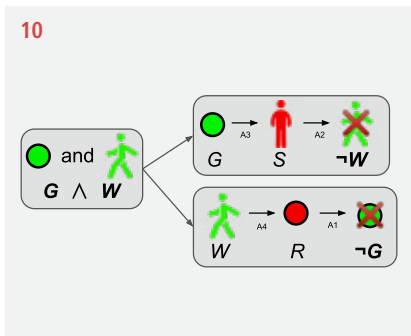
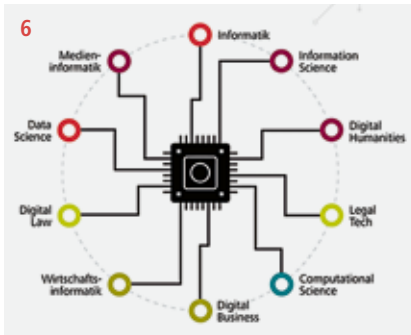


Rohstoffe
Transporte
Produktion

g CO₂e
492
Pro Produkt

CO₂-Emissionen
ausgeglichen

Inhalt



Einleitung 5
Florian Erhard, Bernd Heinrich, Meike Klettke, Christian Wolff

Lehre an der Fakultät für Informatik und Data Science 6
Florian Erhard, Udo Kruschwitz, Bernd Heinrich, Christian Wolff

Automatisches Beweisen: Methoden und Anwendungen 10
Julie Cailler, Philipp Rümmer

Algorithmen und Komplexitätstheorie 13
Radu Curticapean

Evolution in Datenbanken und Data Engineering Workflows 16
Meike Klettke

IoT-basiertes Prozessmanagement – Mobile Benutzerführung in der digitalen Fabrik 19
Stefan Schönig

Cyber Threat Intelligence: Gemeinschaftliche IT-Sicherheit durch den Austausch von Informationen 23
Johannes Grill, Daniel Schlette, Günther Pernul

Kann man den Entscheidungen Künstlicher Intelligenz trauen? Zu den Auswirkungen unsicherer Daten auf die Entscheidungen Neuronaler Netze 26
Thomas Krapf, Bernd Heinrich

Mensch vs. Maschine: Wettbewerb und Kooperation mit künstlicher Intelligenz in digitalen Märkten 30
Andreas Schauer, Daniel Schnurr

Notfallpläne für den Ernstfall testen 34
Maria Leitner

Maschinelles Lernen mit Anwendungen in den Naturwissenschaften 37
Merle Behr, Markus Schmitt

Automatisierte, KI-basierte Analyse von Bilddaten:

Der Lehrstuhl für Bildverarbeitung

Dorit Merhoff

40

Die Genome des Menschen – Forschungsschwerpunkte der Arbeitsgruppe für Algorithmische Bioinformatik

Birte Kehr

43

Algorithmen zum Entschlüsseln der Genregulation

Francisca Rojas Ringeling, Stefan Canzar

46

Mit Hilfe von Daten Immunprozesse entschlüsseln:

Der Lehrstuhl Computational Immunology

Florian Erhard

49

Maschinelles Lernen enthüllt den verborgenen Prozess der Tumorentstehung

Linda Hu, Andreas Lösch, Rainer Spang

52

Allgegenwärtige Mensch-Maschine-Interaktion: Entwicklung, Forschung und Infrastruktur der Medieninformatik

Raphael Wimmer, Johanna Bogon, Niels Henze, Christian Wolff

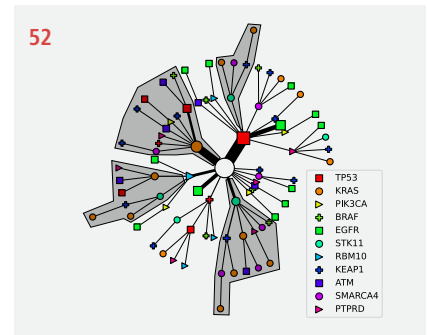
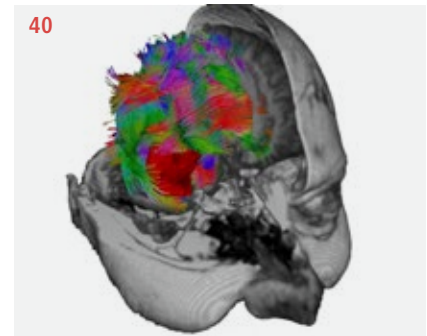
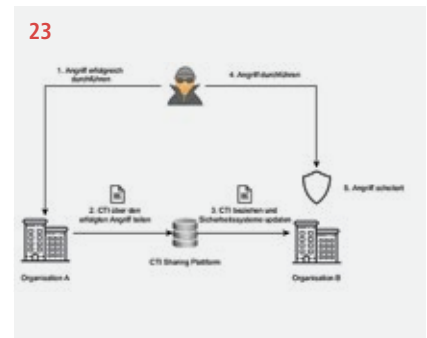
54

Wissen aus dem Internet – Genug, genau, geprüft und geeignet?

Informationswissenschaft in Regensburg

Udo Kruschwitz, Bernd Ludwig, David Elsweiler

60



SOFTWAREENTWICKLER (M/W/D) GESUCHT

Voll- oder Teilzeit (30 -40 Std.) | Alle Fachbereiche

Hallo, **wir sind mb Support.**

Seit 20 Jahren unterstützen wir die Versicherungswirtschaft mit unseren Lösungen, u.a. Konzerte und Kunstwerke weltweit zu versichern. Als inhabergeführtes **Familienunternehmen** arbeiten wir auf Augenhöhe in einer familiären Atmosphäre. Als stark wachsendes und profitables Unternehmen bieten wir gleichzeitig **attraktive Benefits**.

Wer bist du? Jedes unserer rund 60 Teammitglieder bringt seine eigene Persönlichkeit, Geschichte und Perspektive mit. Für uns zählt, was dich interessiert, was dich antreibt, wie du bist. Du entscheidest, wie du bei uns mitgestalten möchtest. **Wir freuen uns auf dich.**

Die gesamte Stellenausschreibung findest du auf www.mbsupport.de/karriere.

www.mbsupport.de | Friedenstraße 18 | 93053 Regensburg | +49 941 942 60 0 | mb Support GmbH

Einleitung

Prof. Dr. Florian Erhard, Prof. Dr. Bernd Heinrich, Prof. Dr. Meike Klettke, Prof. Dr. Christian Wolff

Liebe Leserinnen und Leser,

wir, die Mitglieder der Fakultät für Informatik und Data Science, freuen uns außerordentlich, in diesem Themenheft unser thematisches und methodisches Arbeitsspektrum darstellen zu dürfen.

Zunächst möchten wir Ihnen die neue Fakultät kurz vorstellen: Seit am 1. Januar 2022 die bereits an der Universität Regensburg tätigen Informatik-Kolleginnen und -Kollegen in die neue Fakultät gewechselt sind, neue Lehrstühle besetzt werden konnten, der Fakultätsrat und das Dekanat gewählt und viele Arbeitsgruppen initiiert wurden, hat das aktive Fakultätsleben begonnen.

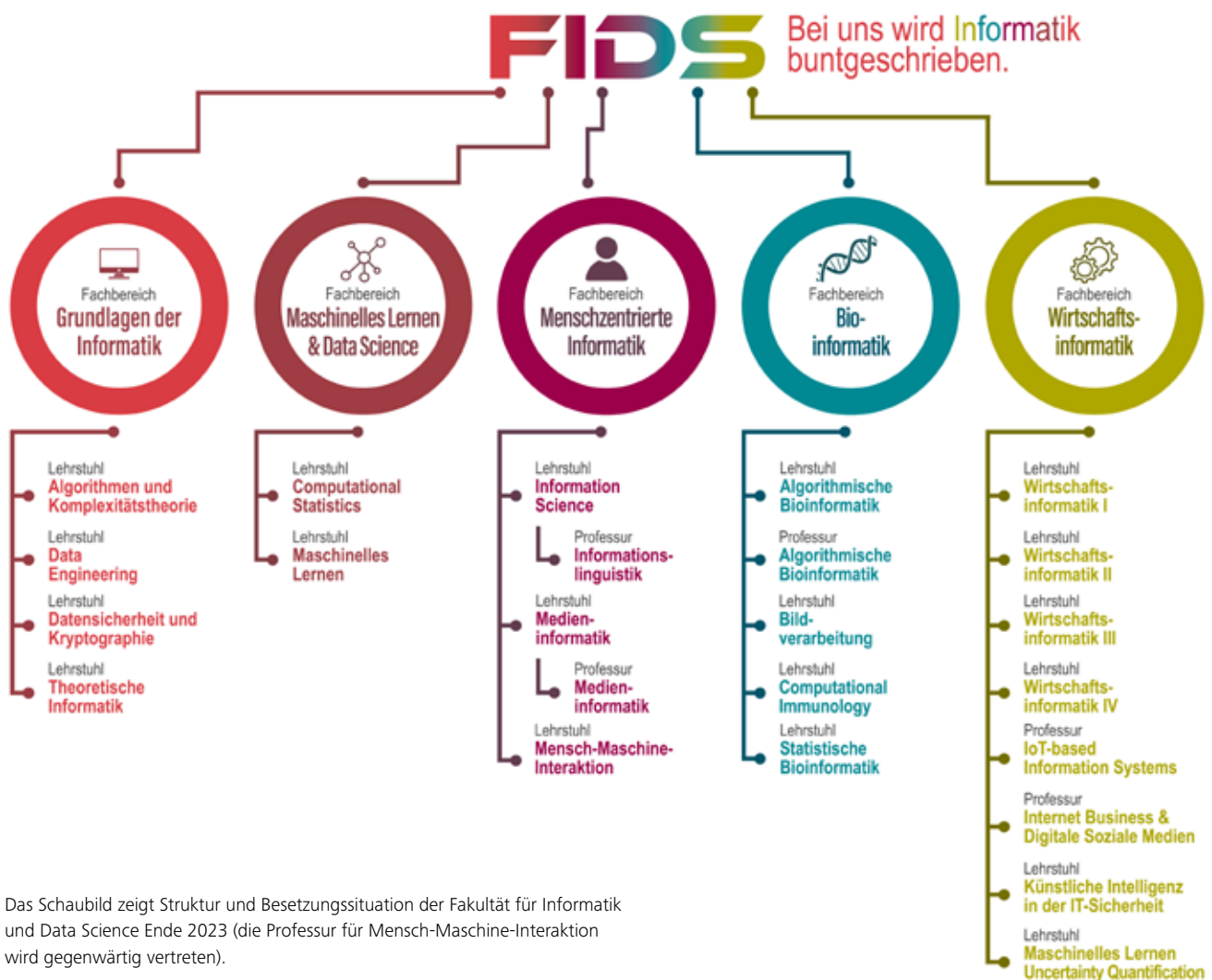
Nach Beendigung aller laufenden Berufungsverfahren wird die Fakultät 30 Professuren umfassen. Sie gliedert sich dabei in fünf Departments:

- Grundlagen der Informatik
- Maschinelles Lernen und Data Science
- Menschzentrierte Informatik
- Bioinformatik
- Wirtschaftsinformatik

Am Anfang des Themenheftes steht ein kurzer Überblick zu den Studiengängen der Fakultät. Ihre Vielfalt und ihre interdisziplinäre Vernetzung machen deutlich, dass wir nicht unbegründet davon sprechen,

dass bei uns „Informatik buntgeschrieben wird“. Neben den im Wintersemester 2023/2024 neu gestarteten grundständigen Bachelorstudiengängen Informatik sowie Data Science existiert bereits eine Vielzahl erfolgreicher, weiterer Informatik- und Informatiknaher Studiengänge an der Fakultät. Insgesamt werden derzeit bereits über 1.000 Studierende in diesen Studiengängen ausgebildet, wobei wir in den kommenden Semestern mit einem deutlichen Wachstum im Bereich der neuen Studiengänge rechnen.

Ausgehend von dem Kurzkonzept für die neue Fakultät, das die Gründungskommission erstellt hat, entwickelte die Fakultät



Das Schaubild zeigt Struktur und Besetzungssituation der Fakultät für Informatik und Data Science Ende 2023 (die Professur für Mensch-Maschine-Interaktion wird gegenwärtig vertreten).

in den letzten Monaten Department-übergreifende Forschungsschwerpunkte. Dabei handelt es sich um folgende Themenfelder:

- Informationssicherheit
- Explainable AI (XAI)
- Computational Methods in the Natural Sciences
- Human-Centered AI

Diese Schwerpunkte vereinen jeweils Grundlagenforschung in den Kernbereichen der Informatik und interdisziplinäre Forschung. Viele einzelne Forschungsthemen unserer Fakultät gestalten diese Schwerpunkte aus, einen kleinen Ausschnitt der thematischen Bandbreite möchten wir Ihnen in diesem Heft vorstellen. Die nachfolgenden 16 Beiträge geben einen Einblick, wie vielfältig, aber auch wie relevant für zentrale Fragen der modernen Gesellschaft unsere Forschungsthemen sind. Wir beginnen mit drei Beiträgen aus dem Grundlagenbereich der Informatik, die mit Fragen der automatisierten Beweisführung, Grundproblemen der Komplexität von Algorithmen und Fragen der Evolution im Bereich Datenbanken und darauf bezogener Workflows drei zentrale Problembereiche heutiger Informatik-Forschung adressieren.

Die folgenden sechs Beiträge aus der Wirtschaftsinformatik gehen nicht nur auf aktuelle Herausforderungen der Digitalisierung für Unternehmen ein, sondern thematisieren (kritisch) auch grundlegende Probleme Künstlicher Intelligenz und die Frage, welche Rolle die Gesellschaft und der Mensch künftig im Wechselspiel mit KI-basierten Anwendungen spielen wird. Weitere sechs Beiträge aus den Bereichen maschinelles Lernen und Anwendungen von KI und maschinellem Lernen in den Lebenswissenschaften zeigen die vielfältigen und grundsätzlichen Herausforderungen für Informatik und Data Science in den Natur- und Lebenswissenschaften auf. Gerade die derzeit untersuchten Themen aus der Medizin verdeutlichen, wie wichtig der Beitrag der Informationstechnologie zur Entwicklung neuer Tumorbehandlungskonzepte ist oder wie automatisierte Bildanalyse medizinische Diagnostik verbessern helfen kann. Schließlich zeigen zwei Beiträge aus dem Arbeitsbereich menschenzentrierte Informatik die buchstäbliche Allgegenwärtigkeit der Mensch-Maschine-Interaktion und die damit verbundenen Herausforderungen der Forschung auf und thematisieren den Umgang mit Wissen im Zeitalter von Internet, Social Media und KI-gestützter Textgenerierung.

Dass Informationstechnologie und Digitalisierung alle Lebensbereiche durchdringen, ist eine Binsenweisheit, die allzu oft zitiert wird. Wenn man aber wirklich die Vielfalt moderner Forschungsfragen in Informatik und Data Science anschaulich nachvollziehen möchte, dann hoffen wir, mit diesem Themenheft einen Beitrag dazu zu leisten und wünschen erkenntnisreiche Lektüren!



Foto © rivat

Prof. Dr.-Ing. habil. Meike Klettke leitet den Lehrstuhl für Data Engineering und ist seit Juni 2022 Forschungsdekanin der Fakultät für Informatik und Data Science.



Foto © Margit Scheid

Prof. Dr. Florian Erhard leitet den Lehrstuhl für Computational Immunology und ist seit Mai 2023 Studiendekan der Fakultät für Informatik und Data Science.




Foto © privat

Prof. Dr. Bernd Heinrich leitet den Lehrstuhl für Wirtschaftsinformatik II und ist seit Juni 2022 Prodekan der Fakultät für Informatik und Data Science.



Foto © Andreas Schmid

Prof. Dr. Christian Wolff leitet den Lehrstuhl für Medieninformatik und ist seit April 2022 Gründungsdekan der Fakultät für Informatik und Data Science.

Lehre an der Fakultät für Informatik und Data Science

Prof. Dr. Florian Erhard, Prof. Dr. Udo Kruschwitz, Prof. Dr. Bernd Heinrich, Prof. Dr. Christian Wolff

Die Gründung der Fakultät für Informatik und Data Science am 6. März 2020 markierte ein bemerkenswertes Kapitel in der Geschichte der Lehre an der Universität Regensburg. Zwar existierten schon vor dieser Gründung bereits eine breite Palette von Informatik- und informatiknahen Studiengängen an der Universität Regensburg, jedoch bündelt erst die FIDS alle an diesen Studiengängen maßgeblich beteiligten Lehrstühle in einer gemeinsamen Fakultät. Zusätzlich wurden 17 weitere Lehrstühle geschaffen, die nicht nur neue Schwerpunkte in der Forschung setzen, sondern auch das Lehrangebot sowohl um Informatik-Kernthemen als auch um informatiknahe und multidisziplinäre Bereiche erweitern.

Die Einführung der neuen Studiengänge in Informatik und Data Science ab dem Wintersemester 2023/24 erweitert das Lehrangebot und bietet den Studierenden die Möglichkeit, sich noch gezielter in den Grundlagen dieser zentralen Disziplinen ausbilden zu lassen, und sich gleichzeitig auf bestimmte Teilrichtungen zu spezialisieren. Dies ist von großer Bedeutung: Die Informatik und die Data Science spielen in der heutigen Zeit Schlüsselrollen in nahezu allen Aspekten der Gesellschaft und der Wirtschaft. Die Absolventinnen und Absolventen dieser Studiengänge werden gut vorbereitet sein, um die Herausforderungen der digitalen Ära anzunehmen und innovative Lösungen zu entwickeln.

Informatik

Der Studiengang B.Sc. Informatik wird seit dem Wintersemester 2023/24 als einer der beiden neuen, eigenen Studiengänge der FIDS angeboten. Ein konsekutiver Studiengang M.Sc. Informatik wird zum Wintersemester 2025/26 eingeführt.

Der B.Sc. Informatik bietet eine umfassende Ausbildung, die alle Kerngebiete der Informatik abdeckt, einschließlich Zukunftsthemen wie Künstliche Intelligenz, Mensch-Maschine-Interaktion und IT-Sicherheit. Wahlpflichtfächer in den Gebieten Data Science, Human Information Behavior, Medieninformatik oder Wirtschaftsinformatik ermöglichen die Vertiefung persönlicher Interessen und den Erwerb von Teamwork- und Rhetorikkompetenzen. Dieser Studiengang qualifiziert für eine erfolgreiche berufliche Karriere in der Informatik und bereitet gleichzeitig auf ein weiterführendes Masterstudium in der Kerninformatik oder einer Vielzahl von informatiknahen Fächern vor.

Data Science

Ebenso wie für die Kerninformatik wird der Studiengang B.Sc. Data Science seit dem Wintersemester 2023/24 neu an der FIDS angeboten. Zum Wintersemester 2025/26 schließt sich ein konsekutiver Studiengang M.Sc. Data Science an.

Das Data-Science-Studium an der Universität Regensburg vermittelt ein solides Fundament in Mathematik, Statistik und Informatik, das essenziell für erfolgreiche Data Scientists ist. Der breitgefächerte Wahlpflichtbereich ermöglicht die Anwendung dieser Werkzeuge in verschiedenen Themengebieten, von Tumorforschung über Quantenmechanik bis hin zu Fake-News-Erkennung und IT-Sicherheit. Zusätzlich besteht die Möglichkeit, das Werkzeugwissen individuell zu vertiefen und Kurse aus dem gesamten Lehrangebot der Universität zu wählen, um Kompetenzen in Bereichen wie Rhetorik oder Projektmanagement zu entwickeln. Dieses Studium bereitet optimal auf eine Karriere im Bereich Data Science vor und bietet große Flexibilität in der Studiengestaltung.



1 An der Universität Regensburg werden eine große Vielfalt an Informatik- und informatiknahen Studiengängen angeboten. Diese Studiengänge arbeiten eng mit anderen Fakultäten zusammen (Dunkelrot: Fakultät für Sprach-, Literatur- und Kulturwissenschaften, Gelb: Fakultät für Rechtswissenschaft, Ocker: Fakultät für Wirtschaftswissenschaften, Türkis: Fakultät für Biologie und Vorklinische Medizin und Fakultät für Physik).

Information Science

Digitalisierung, Künstliche Intelligenz und Big Data verändern unser Leben – vom Alltag über die Arbeitswelt bis hin zur Gesundheitsvorsorge. Zahlreiche Herausforderungen, die diese Trends an uns stellen, betreffen den menschlichen Umgang mit Informationen: Die unüberschaubare Informationsflut im Internet, die rasante Verbreitung von Fake News und die Begünstigung von Cybermobbing in anonymen digitalen Räumen sind nur einige Beispiele. Information Scientists stellen sich diesen und weiteren Herausforderungen, indem sie versuchen, den menschlichen Umgang mit Informationen zu verstehen und darüber hinaus Werkzeuge entwickeln, um Menschen in der modernen digitalen Welt bei ihrem Informationsumgang zu unterstützen und zu schützen (von Assistenz-

Chatbots bis hin zu zur automatischen Erkennung von Fake News).

Studierende der Information Science erhalten daher ein breit gefächertes Wissen und praxisorientierte Handlungskompetenz in den Bereichen Informatik (v. a. Programmieren), Data Science und Forschungsmethoden. Die erlernten Methoden und Werkzeuge aus diesen Bereichen bereiten die Studierenden sowohl auf ein Berufsleben in der Industrie als auch auf eine akademische Karriere vor. Studierende der Information Science finden Anstellung in der Automobilindustrie, der IT- und Softwareindustrie, in Banken und Versicherungen oder in der Web- und Medienbranche. Dort entwickeln sie zum Beispiel Bedienkonzepte für PCs, Smartphones und intelligente Geräte, sind verantwortlich für die nutzerfreundliche und zielgruppengerechte Strukturierung von Informationsangeboten oder sind mit dem Informations- und Wissensmanagement ihres Unternehmens betraut. Tätigkeiten im Bereich IT-Projektmanagement, Data-Engineering, Web-Engineering oder User-Experience-Design sind überdies möglich. Mit unserem Ansatz der forschungs-basierten Lehre bahnen wir vielen unserer Studierenden darüber hinaus den Weg ins akademische Leben.

Information Science (Informationswissenschaft) wird an der Universität Regensburg als kombinatorischer Bachelorstudiengang (B.A.) und als Masterstudiengang (M.A.) angeboten.

Medieninformatik

In den Studiengängen steht die Konzeption, Entwicklung und Bewertung der Mensch-Maschine-Interaktion im Mittelpunkt, ein mittlerweile allgegenwärtiges Thema, denkt man nur an die Intensität, mit der heute Smartphones als kleine Universalcomputer genutzt werden. Im Bachelorfach Medieninformatik, das im kombinatorischen Bachelorstudiengang der philosophischen Fakultäten gewählt werden kann, stehen in der Anfangsphase zunächst die Einführung in das Fach (und in die Informatik), mathematische Grundlagen sowie die Softwaretechnik-Grundausbildung im Mittelpunkt. Kernthemen der Medieninformatik wie Mensch-Maschine-Interaktion, Multimedia and Usability Engineering und Medientechnologie kommen hinzu. Im fortgeschrittenen Studium können Studierende aus einer Vielzahl von Ver-

tiefungsvorlesungen wählen, z. B. Computergraphik oder Entertainment Computing / Game Engineering. Außerdem wird ein fünfteiliger Vorlesungszyklus zu Digitalisierung und digitaler Gesellschaft angeboten.

Im Masterstudium (M. Sc. Medieninformatik) können die besonderen Schwerpunkte der Medieninformatik forschungsnah vertieft werden. Neben einer weiterführenden Ausbildung in Softwaretechnik und Projektmanagement können Studierende aus einer Reihe von thematischen Projektseminaren wählen (Digital Humanities, Advanced Usability Engineering, Interaktionstechniken und -technologien, interaktive Informations- und Assistenzsysteme sowie Game Engineering). Ein Praxisseminar vertieft die softwaretechnischen Erfahrungen, ein Forschungsseminar führt zusammen mit einer Vertiefungsveranstaltung zu wissenschaftlicher Methodik an die Forschung und die Masterarbeit heran. Insgesamt sind das Masterstudium und seine Prüfungsleistungen sehr stark team- und projektorientiert gestaltet. Neben weiteren hier beschriebenen Studiengängen (Digital Humanities, Digital Law, Legal Tech) ist die Medieninformatik auch am Master Public History und Kulturvermittlung beteiligt.

Digital Humanities

Der Masterstudiengang Digital Humanities richtet sich an Geisteswissenschaftler*innen und vermittelt Methoden aus der Informatik und Data Science. Das Studium ist untergliedert in ein Lehrjahr und ein Projektjahr. In ersterem werden Themen wie Einführung in Digital Humanities, Digitalisierung, Programmierung mit Python, Datenanalyse, maschinelles Lernen, empirische Methoden, automatische Sprachverarbeitung und Webtechnologien behandelt. Im zweiten Jahr wählen die Studierenden Projektseminare aus verschiedenen Bereichen (Digitalisierung, maschinelle Sprachverarbeitung, Informationsverhalten und Geographische Informationswissenschaft).

Wirtschaftsinformatik

Die Wirtschaftsinformatik als ein interdisziplinäres Fachgebiet beschäftigt sich mit innovativen und zukunftsgerichteten Themen und neuen Technologien z. B. in den Bereichen Digital Business, IT Security, Process Science, Data Analytics und Machine Learning oder Social Media. Diese führen

zu einer Vielzahl digitaler Dienste und Innovationen, auf deren Basis neue, innovative Geschäftsmodelle, Produkte und Dienstleistungen entwickelt werden. Unternehmen des Internetzeitalters (z. B. Google, Amazon, Celonis oder Netflix) aber auch die immer stärkere technologische Ausrichtung traditioneller Unternehmen (z. B. Automobilhersteller, Dienstleister) verdeutlichen diese Entwicklung. Das Bachelor- und Masterstudium Wirtschaftsinformatik greift diese Veränderungen auf und erlaubt – aufbauend auf methodischen, technischen und fachlichen Grundlagen – eine individuelle Schwerpunktsetzung in den vier Vertiefungen Digital Business, Data Science & AI Applications, Process Science und IT Security. Daneben besteht eine große Auswahl an Wahlkursen zur flexiblen Ausgestaltung des Studiums nach den eigenen Präferenzen. Für besonders leistungsbereite und begabte Studierende wird zusätzlich ein »Honors«-Elitestudiengang für Wirtschaftsinformatik angeboten. Studierende werden dadurch gezielt an die internationale Spitzenforschung, aber auch an die Aufgaben und Verantwortung von Führungskräften in der Industrie herangeführt.

Digital Business

Der Bachelorstudiengang B.Sc. Digital Business behandelt die Auswirkungen der Digitalisierung auf Unternehmen. Er kombiniert Data Analytics, Wirtschaftsinformatik und Betriebswirtschaft, um ein umfassendes Verständnis für datenbasierte Entscheidungsfindung und die Integration von Technologie und Strategie in der Wertschöpfungskette zu vermitteln. Das Studium ist in zwei Phasen unterteilt: In der ersten Phase werden grundlegende Kenntnisse in Betriebswirtschaft, Digitalisierung und Technologiemanagement erworben, während die zweite Phase eine fachliche Spezialisierung in den Bereichen Data Analytics, Wirtschaftsinformatik und Betriebswirtschaft ermöglicht. Studierende können zudem Schwerpunkte im Wertschöpfungsmanagement oder Finanzmanagement setzen und nehmen in beiden Phasen Module in den Bereichen Data Analytics und Wirtschaftsinformatik wahr.

Digital Law

Der Bachelorstudiengang LL.B. Digital Law befasst sich mit den Auswirkungen der Digitalisierung auf das Rechtswesen und



untersucht die rechtlichen Herausforderungen, die mit den technologischen Veränderungen und gesellschaftlichen Transformationsprozessen im digitalen Zeitalter einhergehen. Die Absolventinnen und Absolventen sind gut positioniert, um in einer Vielzahl von beruflichen Feldern tätig zu sein, in denen die Digitalisierung komplexe rechtliche Herausforderungen aufwirft. Dies umfasst Bereiche wie Softwareunterstützung rechtlicher Prozesse, technologische Innovationen im Rechtswesen und die Bewältigung der »digitalen Revolution« in verschiedenen Branchen. LL.B. Digital Law kann im Doppelstudium mit dem Staatsexamensstudiengang Jura studiert werden.

Legal Tech

LL.M. Legal Tech ist ein Weiterbildungsmaster, der sich auf die Auswirkungen der Digitalisierung auf das juristische Arbeiten konzentriert. Er ist als ein modernes Studienprogramm auf dem Gebiet der Rechtsinformatik konzipiert. Zu etwa gleichen Teilen werden Themen aus der Informatik und aus der Rechtswissenschaft behandelt: Themen wie Big Data, Künstliche Intelligenz, Information Retrieval, Blockchain / Smart Contracts und die Nutzung moderner IT-Tools

und Methoden wie Legal Design Thinking im Rechtswesen prägen den technischen Teil. Er befasst sich mit rechtlichen Aspekten von Algorithmen und Cybercrime sowie mit Digital Law in Anwendungsfeldern wie IT-Recht, Datenschutz, e-Commerce und KI-Regulierung. Besonderes Augenmerk erhalten die aktuellen Legal Tech-Angebote (z. B. Abwicklung von Massenfällen (Dieselskandal), alternative Streitbeilegung, Vertrags- und Dokumentengeneratoren).

Computational Science

Das Fach Computational Science wird als Bachelor- (B.Sc.) und Masterstudiengang (M.Sc.) angeboten und verbindet die Bereiche Physik, Mathematik und Genomik/Bioinformatik. Die Studiengänge vermitteln Fähigkeiten in der Entwicklung wissenschaftlicher Software, der Simulation komplexer Systeme, der Datenanalyse und der Anwendung mathematischer Methoden zur Optimierung von Prozessen. Computational Science bietet eine solide Grundausbildung in computerbasierter Forschung und ermöglicht Schwerpunktsetzungen in Bereichen wie Computational Physics, Numerische Mathematik oder Medizinische Bioinformatik.



Foto © Margrit Scheid

Prof. Dr. Florian Erhard ist seit März 2023 Inhaber des Lehrstuhls Computational Immunology. Im darauf folgenden Juni wurde er zum Studiendekan gewählt. Als Studiendekan leitet er unter anderem die Planungskommission für die zukünftigen Masterstudiengänge Informatik und Data Science.



Foto © privat

Prof. Dr. Bernd Heinrich ist seit 2011 Inhaber des Lehrstuhls für Wirtschaftsinformatik II an der Universität Regensburg. Er promovierte an der Universität St. Gallen/Schweiz und habilitierte sich an der Universität Augsburg. Seine Forschungsschwerpunkte liegen in den Bereichen Data Science, Data Quality & Data Uncertainty, Process Planning & Process Science sowie Service Systems.



Foto © UR, Antonia Prois

Prof. Dr. Udo Kruschwitz ist seit 2019 Professor für Informationswissenschaft an der Universität Regensburg. Vorher war er über 20 Jahre an der University of Essex tätig. Sein Hauptforschungsgebiet ist die Schnittstelle zwischen Information Retrieval (IR) und Natural Language Processing (NLP). Er ist derzeit Chair der Information Retrieval Specialist Group bei der British Computer Society.



Foto © Andreas Schmid

Prof. Dr. Christian Wolff (geb. 1966) ist seit 2003 Professor für Medieninformatik an der Universität Regensburg. Er ist promovierter Informationswissenschaftler (1994, Universität Regensburg) und habilitierter Informatiker (2000, Universität Leipzig).

Automatisches Beweisen: Methoden und Anwendungen

Julie Cailler, Prof. Dr. Philipp Rümmer

Einleitung

Mit der zunehmenden Verbreitung und Komplexität von Computersystemen ist Zuverlässigkeit zu einem wichtigen Thema geworden. Dies gilt insbesondere für Computersysteme, die sicherheitskritische Funktionen implementieren, z. B. Systeme, die in Fahrzeuge oder Flugzeuge eingebettet sind. *Formale Methoden* bieten eine Möglichkeit, Vertrauen in solche Systeme zu schaffen. Insbesondere die *Programmverifikation* ermöglicht es uns, die Abwesenheit von Fehlern in einem System oder seinen Komponenten nachzuweisen: mit stringenten logischen und mathematischen Methoden wird ein Beweis für die Korrektheit eines Systems erbracht, der bestätigt, dass das System genau so funktioniert wie beabsichtigt und keine Abweichungen aufweist.

Formale Methoden stützen sich auf Techniken für computergestütztes Ableiten logischer Formeln, ein Bereich der Informatik, der sich in den letzten Jahren rasch weiterentwickelt hat. Logisches Denken wird alltäglich von Menschen verwendet, z. B. wenn sie argumentieren oder schlussfolgern. Logisches Schließen kann aber auch von Maschinen ausgeführt werden und ermöglicht formale Schlussfolgerungen über Probleme, die viel zu komplex sind, um von Menschen allein bewältigt zu werden. Um ein System formal zu verifizieren, werden sowohl das System selbst als auch die gewünschten Eigenschaften des Systems (automatisch) in logische Formeln übersetzt, die dann von Theorembeweisern verarbeitet werden können, um das korrekte Verhalten des Systems nachzuweisen oder zu falsifizieren. Um dies zu erreichen, werden zwei Hauptansätze verwendet: *Interaktives Beweisen* und *Automatisches Beweisen*.

Interaktive Beweiser funktionieren wie Assistenten, die einen menschlichen Benutzer bei der Beweisführung unterstützen und gleichzeitig sicherstellen, dass die konstruierte Ableitung nicht fehlerhaft ist. Interaktive Beweiser werden insbesondere in der Mathematik eingesetzt, um Beweise von Sätzen zu formalisieren und zu überprüfen; derartige Beweiser werden aber auch in der Verifikation genutzt, um Komponenten in Systemen zu überprüfen, die für die Sicherheit von Menschen von kritischer Bedeutung sind. So wurden beispielsweise Sicherheitsaspekte der Pariser Metrolinie 14 mit Hilfe der sogenannten B-Methode verifiziert, das Betriebssystem sel4 wurde mit dem Beweisassistenten Isabelle/HOL verifiziert und der Beweisassistent HOL Light war ein wichtiges Werkzeug beim Beweis der Keplerschen Vermutung im Jahr 2017. Die neuesten Entwicklungen im Bereich des interaktiven Beweisens wurden kürzlich auf der Sommerschule *Interactions of Proof Assistants and Mathematics* vorgestellt, die gemeinsam von FIDS und der Fakultät für Mathematik in Regensburg organisiert wurde.

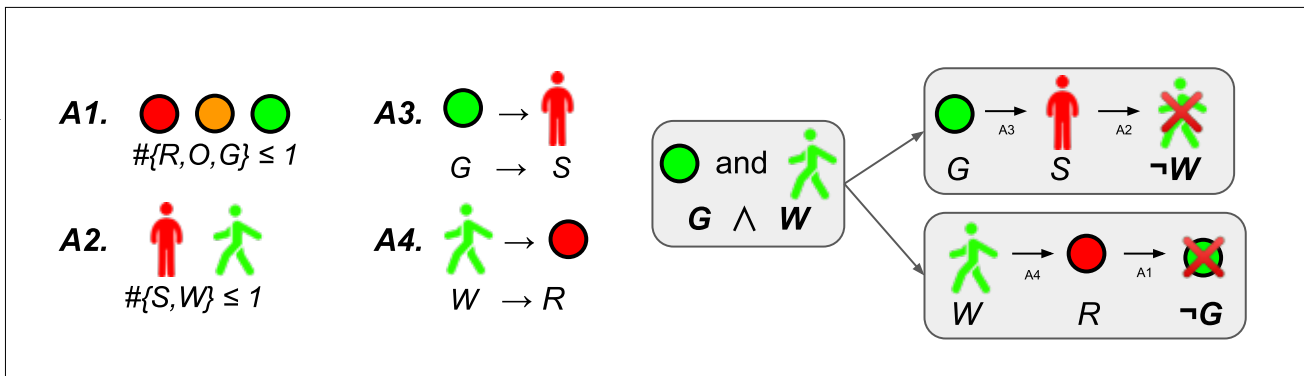
Automatische Beweiser sind in der Lage, logische Formeln ganz oder teilweise automatisch abzuleiten. Sie werden auch als *Theorembeweiser* bezeichnet und in vielen Bereichen eingesetzt, z. B. bei der Verifikation und dem Testen von Programmen, im Scheduling und auch bei der Lösung mathematischer Probleme. Interaktive und automatische Beweiser können in Kombination verwendet werden, indem beispielsweise einfachere Teile eines interaktiv erstellten Beweises an automatische Theorembeweiser ausgelagert werden. Es gibt ein breites Spektrum an automatischen Beweismethoden, die alle von einer mathematischen Formulierung eines Prob-

lems und vorgegebenen Annahmen ausgehen. Automatische Beweiser verarbeiten die formalisierte Aussage und untersuchen symbolisch den Raum der möglichen Interpretationen der Formel und versuchen dadurch, ein Modell zu finden (d. h. eine Interpretation, die die Formel erfüllt) oder andernfalls einen Beweis zu konstruieren, dass die Formel nicht erfüllbar ist.

Ein Verifikationsbeispiel

Als Beispiel betrachten wir die Beweisschritte, die zur Verifikation eines (drastisch vereinfachten) Systems angewandt werden. Eine Kreuzung ist mit zwei Ampeln ausgestattet: eine für die Autos und eine für die Fußgänger. Die Interaktion der beiden Ampeln kann wie folgt spezifiziert werden. Die Ampel für Autos kann Grün (G), Orange (O) oder Rot (R) sein, während die Fußgängerampel zwei Modi hat: Gehen (W) und Stehen (S). Um das System zu analysieren, müssen wir anschließend Annahmen (oder Axiome) über die möglichen Zustände des Systems formulieren. In realistischeren Fällen wird diese formale Beschreibung des Systems automatisch aus der Systemimplementierung abgeleitet:

- **A1:** Rot, Orange und Grün schließen sich gegenseitig aus: höchstens eines von R, O, G ist wahr ($\#\{R, O, G\} \leq 1$)
- **A2:** Stop und Walk schließen sich gegenseitig aus: höchstens eines von S und W ist wahr ($\#\{S, W\} \leq 1$)
- **A3:** Die Fußgängerampel zeigt nur dann Gehen an, wenn die Ampel Rot ist: $W \rightarrow R$
- **A4:** Die Ampel zeigt nur dann Grün, wenn die Fußgängerampel Stehen anzeigt: $G \rightarrow S$



1 Formalisierung und Verifikation einer vereinfachten Ampelschaltung.

Unser Ziel ist es, sicherzustellen, dass der Verkehr an der Kreuzung ordnungsgemäß geregelt ist, um Unfälle zu vermeiden. Eine interessante Eigenschaft in diesem Modell ist die Frage, ob es jemals vorkommen kann, dass die Ampel Grün zeigt, während die Fußgängerampel im Modus Gehen ist: Können G und W gleichzeitig wahr sein, oder, logisch ausgedrückt, ist die Formel $G \wedge W$ unter unseren Annahmen erfüllbar. Ein Theorembeweiser kann dieses Szenario auswerten und Schritt für Schritt beweisen, dass $G \wedge W$ unmöglich ist. Da dieser Beweis *maschinell* geführt und überprüft wird, ist er ein starkes Argument für die Korrektheit der Ampelimplementierung.

Praktisch gesehen simuliert ein Beweiser alle möglichen Szenarien und zeigt, dass die untersuchte Fehlersituation in jedem Fall aufgrund von Widersprüchen unerreichbar ist. Für ein und dieselbe Eigenschaft können mehrere Beweise vorliegen: Wenn beispielsweise die Ampel grün ist, befindet sich die Fußgängerampel im Modus Stehen ($G \rightarrow S$), wie im oberen Block dargestellt. Der Modus Stehen ist mit dem Modus Gehen unvereinbar ($\#\{S, W\} \leq 1$), was der ursprünglichen Aussage (W) widerspricht. Eine ähnliche Argumentation kann auch für den Modus Gehen gemacht werden, was uns erlaubt, auf zwei verschiedene Arten zu schließen, dass $G \wedge W$ unerfüllbar ist. Indem wir auf diese Weise die komplette Funktionalität der Ampelsteuerung analysieren, können wir bestätigen, dass jede Komponente wie erwartet funktioniert und somit das System als Ganzes verifiziert ist.

Methoden und Forschungsthemen in automatischen Beweisen

Manche Systeme lassen sich relativ leicht modellieren, während andere, komplexere Systeme auch ausdrucksstärkere Beschreibungssprachen erfordern. Glücklicherweise gibt es nicht nur eine einzige Art von Logik, sondern eine Vielzahl von verschiedenen Logiken. Diese unterscheiden sich bezüglich der verfügbaren logischen Konstrukte (*Syntax*) und der Art und Weise, wie man Termen und Formeln Bedeutung zuweist (*Semantik*). *Aussagenlogik*, die relativ einfache Logik, die in unserem Ampelbeispiel verwendet wird, hat den Vorteil, dass sie entscheidbar ist; das bedeutet, dass wir in der Lage sind, effektiv zu berechnen, ob eine Formel erfüllbar ist oder nicht. Erfüllbarkeit in der Aussagenlogik ist jedoch ein *NP-vollständiges* Problem, was bedeutet, dass alle bekannten Algorithmen im schlimmsten Fall exponentielle Laufzeit haben. Verbesserungen bei den Beweissystemen, den Algorithmen und den in den Implementierungen verwendeten Datenstrukturen haben in den letzten Jahren jedoch zu einer überraschenden Steigerung der Leistungsfähigkeit von Theorembeweisern für Aussagenlogik geführt. Solche Beweiser, die gemeinhin als *SAT-Solver* bekannt sind, können heute problemlos Formeln mit Millionen von Variablen verarbeiten, z. B. Formeln, die industrielle Systeme modellieren.

Ausdrucksstärkere Logiken, z. B. die *Prädikatenlogik erster Stufe* oder *Logiken höherer Stufe*, erlauben es uns, komplexere Aussagen (z. B. mathematische Probleme) zu formulieren, aber Beweise sind schwieriger zu finden. Auch hier wurde viel Forschungsarbeit in die Entwicklung effizienter automatischer Theorembeweiser

investiert, und die heutigen Systeme sind in der Lage, auch sehr komplexe Probleme zu lösen. In unserer Gruppe in Regensburg haben wir insbesondere an Methoden zur *Parallelisierung* von Theorembeweisern gearbeitet, d. h. an der Nutzung moderner paralleler Hardwarearchitekturen um die Beweissuche zu beschleunigen.

Weitere Logiken werden fortlaufend entwickelt und bringen eine Vielzahl konzeptioneller, theoretischer und praktischer Fragestellungen mit sich, wie z. B.:

- Was ist die *richtige Logik* für die Modellierung einer bestimmten Art von System?
- Welche *Zeit-* oder *Platzkomplexität* haben Inferenzprobleme in einer bestimmten Logik? Sind Probleme wie Erfüllbarkeit entscheidbar?
- Was sind *gute Beweissysteme* und effektive oder effiziente Algorithmen in einer gegebenen Logik?

Viele Anwendungen erfordern auch den Umgang mit domänenspezifischen *Axiomen*. In unserem Beispiel gibt es nur vier Axiome, A1-A4, während die Modellierung realer Systeme Tausende von Axiomen erfordern kann. Um diese Situation handhabbar zu machen, werden bestimmte standardisierte (Mengen von) Axiomen als *Theorien* zusammengestellt, die in der Programmierung verwendeten standardisierten Datentypen oder Bibliotheken entsprechen. Eine Theorie umfasst Hintergrundwissen aus einem bestimmten Bereich (z. B. verschiedene Arten von Arithmetik, Mengenlehre oder Arrays) und kann durch spezifische Beweisregeln und Prozeduren repräsentiert werden. Die Entwicklung spezieller Theorembeweiser, die mit Theorien umgehen können, auch bekannt als *Satis-*

fiability Modulo Theory Solver oder *SMT Solver*, hat sich in den letzten Jahren zu einem besonders erfolgreichen Forschungszweig entwickelt. SMT Solver haben den Bereich der formalen Methoden grundlegend verändert, indem sie leistungsstarke Bausteine zur Verfügung stellen, aus denen sich neue Verifikationswerkzeuge mit viel geringerem Aufwand konstruieren lassen.

Unsere Forschungsgruppe in Regensburg hat selbst mehrere SMT Solver entwickelt, die auf Probleme in der Programmverifikation zugeschnitten sind. Wir arbeiten dabei an einigen der spannendsten Theorien auf diesem Gebiet: unter anderem

an Theorembeweisern für verschiedene Varianten der Theorie der *Zeichenketten*, die sowohl mit Wortgleichungen über freien Monoiden als auch mit der Automatentheorie zusammenhängen; Theorembeweisern für verschiedene Arten von *Arithmetik*, einschließlich Problemen in nichtlinearer Ganzzaharithmetik und modularer Arithmetik; und Theorien von *Sequenzen* und *Arrays*. Die Vielfalt der in Frage kommenden Theorien und logischen Fragmente macht den Bereich SMT zu einem hochinteressanten Forschungsgebiet, sowohl aus theoretischer als auch aus praktischer Sicht.

Referenzen

Leonardo Mendonça de Moura, Nikolaj S. Bjørner: *Satisfiability modulo theories: introduction and applications*. Commun. ACM 54(9): 69–77 (2011)

John Harrison: *Handbook of Practical Logic and Automated Reasoning*. Cambridge University Press 2009, ISBN 978-0-521-89957-4, pp. I-XIX, 1-681

Interactions of Proof Assistants and Mathematics. International Summer School, Regensburg, Germany, 2023. <https://itp-school-2023.github.io/>

Raymond M. Smullyan. *What is the Name of this Book?: The Riddle of Dracula and Other Logical Puzzles*. Prentice-Hall, 1978.

DREI FRAGEN ...

Wie verändert KI Ihr Fachgebiet?

Automatisches Beweisen ist ein Teilbereich der künstlichen Intelligenz. Im Gegensatz zu statistischen Methoden der KI und insbesondere dem maschinellen Lernen, fokussiert das automatische Beweisen auf symbolische Methoden, mit denen logische oder mathematische Aussagen exakt und fehlerfrei von gegebenen Annahmen abgeleitet werden können. Automatisches Beweisen und maschinelles Lernen ergänzen sich daher ausgesprochen gut, weshalb viele Forschergruppen heute versuchen, Methoden des maschinellen Lernens in das automatische Beweisen zu integrieren oder umgekehrt.

Welche Entwicklung in der Informatik wünschen Sie sich für die nächsten 5 Jahre?

Formale Methoden werden heute hauptsächlich im Bereich von kritischen Computersystemen eingesetzt. Wir sind überzeugt, dass es an der Zeit ist, dass formale Methoden diese Nische verlassen und zu einem Mainstream-Werkzeug in der Softwareentwicklung werden. Neben den im Artikel genannten Gründen ist ein weiterer Punkt, dass in Zukunft immer mehr Computerprogramme maschinell generiert werden, was zu dem neuen Problem führt, sicherzustellen, dass sich diese generierten Programme tatsächlich wie erwartet verhalten. Genauso wie wir einem von ChatGPT generierten Text nicht blind vertrauen sollten, bedürfen auch maschinell generierte Computerprogramme einer sorgfältigen Überprüfung. Im Gegensatz zu Texten in natürlicher Sprache sind Menschen jedoch eher schlecht darin,

Programmcode zu überprüfen: maschinell erstellte Programme sollten maschinell überprüft werden!

Welches Fachbuch lesen Sie gerade?

»What Is the Name of This Book?« von Raymond M. Smullyan – Eine unterhaltsame Einführung in die Deduktion und logische Beweisführung, die Grundlage des automatischen Beweisens. Anhand von Rätseln gibt das Buch einen Einblick in die Art und Weise, wie Theorembeweiser Probleme lösen. Dies ist wahrscheinlich das unterhaltsamste Logikbuch, das je geschrieben wurde, und wir empfehlen es Lesern jeden Alters! Mit den Worten von Melvin Fitting: »Ich stelle Ihnen jetzt Professor Smullyan vor, der Ihnen beweisen wird, dass entweder er nicht existiert oder Sie nicht existieren (aber welches davon werden Sie nicht wissen!).«



Foto © Universität Montpellier

Julie Cailler

ist seit 2023 Post-Doktor am Lehrstuhl für Theoretische Informatik. In ihrer Forschung beschäftigt sich Julie Cailler primär mit Techniken und Werkzeugen des automatischen Beweisens, sowie mit der Kombination von interaktiven und automatischen Beweismethoden. Sie ist die Hauptautorin des Beweissystems Goéland, eines nebenläufigen Tableau-basierten Theorembeweislers für die Prädikatenlogik ersten Stufe.



Foto © Isabella Rümmer

Prof. Dr. Philipp Rümmer

ist seit 2022 Professor für Theoretische Informatik an der Universität Regensburg. Die Schwerpunkte seiner Forschung sind das automatische Beweisen, insbesondere Methoden aus dem Bereich SMT-Solving und der Constraint-Programmierung, sowie Methoden für die Verifikation von Software-Programmen mit Hilfe von Deduktion und Model Checking. Neben theoretischen Resultaten hat diese Forschung auch zu mehreren Software-Werkzeugen für das automatische Beweisen und der Programmverifikation geführt.



Algorithmen und Komplexitätstheorie

Prof. Dr. Radu Curticapean

Eine fiktive Situation

Ihr Herz pocht, aber alles läuft nach Plan. Die Countdowns an der Decke des *Technik-Tempels* zählen in großen Ziffern die verbleibende Zeit im Gewinnspiel herunter. Nur noch 28 Sekunden!

Sie eilen mit Ihrem Einkaufswagen über den dunkelgrauen Teppichboden, während Sie durch die übersteuernde Durchsageanlage von einem Prominenten angefeuert werden. Eine große Anzeige beziffert den Gesamtpreis der Waren in Ihrem Einkaufswagen auf 8.562,11 €. Wenn Sie diesen Wert in der verbleibenden Zeit auf exakt 10.000,00 € bringen, werden Sie den gesamten Inhalt des Einkaufswagens abräumen.

Sie laufen zu einem Fernseher einer namhaften koreanischen Marke, dessen Preis natürlich exakt 1.437,89 € beträgt. Der Weg dorthin wird 6 Sekunden dauern, das zuvor geübte Einladen des Fernsehers 7 Sekunden, und in 8 weiteren Sekunden werden Sie das Ziel erreichen, wo Ruhm, Ehre und elektronische Konsumgüter auf Sie warten.

Eine nützliche Vorlesung

Für die Auswahl der Waren und die Planung Ihrer *TechnikTempel*-Tour haben Sie ein Computerprogramm genutzt, das Sie mit Wissen aus Ihrem Informatikstudium selber entwickelt haben. Insbesondere die Vorlesung »Algorithmen und Datenstrukturen« hat sich (vielleicht wider Erwarten) hierfür als nützlich erwiesen: In dieser Vorlesung ging es zwar nicht um dieses fiktive Gewinnspiel, aber dafür um abstrakte **Berechnungsprobleme** und **Algorithmen**, die Ihnen letztendlich auch in dieser kon-

kreten Anwendung die Planung abnehmen konnten.

Die Vorlesung wurde vom Lehrstuhl für Algorithmen und Komplexitätstheorie an der neuen Fakultät für Informatik und Data Science gehalten. Das Forschungsgebiet dieses Lehrstuhls werden wir in diesem Artikel nun etwas eingehender beleuchten.

Was sind Berechnungsprobleme und Algorithmen?

Ganz allgemein gilt es in Berechnungsproblemen, aus einer gegebenen und klar definierten Eingabe von Informationen eine erwünschte Ausgabe von Informationen zu erzeugen. Einige Beispiele: Beim *Sortieren* ist die Eingabe eine Liste von Zahlen, und die erwünschte Ausgabe ist eine aufsteigend sortierte Version dieser Liste. Auch das Testen auf *Primzahlen* ist ein Berechnungsproblem: Die Eingabe ist eine Zahl, und die erwünschte Ausgabe ist JA, falls es sich bei der Eingabe um eine Primzahl handelt, andernfalls NEIN.

Berechnungsprobleme werden durch Algorithmen gelöst, wie sie auch am Lehrstuhl für Algorithmen und Komplexität entwickelt werden. Algorithmen lassen sich als »Kochrezepte« auffassen, die präzise beschreiben, wie sich die erwünschte Ausgabe für ein Berechnungsproblem durch einfache Schritte aus der eingegebenen Information herstellen lässt. Je nach erwünschtem Abstraktionsniveau können solche Algorithmen als natürlichsprachliche Anweisungen gegeben sein, als expliziter Programmcode, oder als sogenannter *Pseudo-Code* zwischen diesen beiden Extrema. Algorithmen können dann schrittweise mit Rechengereäten auf konkreten Eingaben ausgeführt werden.

Algorithmen aus der Grundschule

Schon in der Grundschule haben wir, mit Stift und Papier gewappnet, als kleine Rechengereäte Algorithmen ausgeführt. Beispielsweise lernten wir alle schon früh, wie sich große Zahlen addieren lassen. Das ist ein ganz klassisches Berechnungsproblem: Die Eingabe besteht hier aus Zahlen a und b in ihren Dezimaldarstellungen, und die erwünschte Ausgabe ist die Dezimaldarstellung der Summe $a+b$.

Sie haben bereits unzählige Male einen Algorithmus für dieses Problem ausgeführt: Zur Berechnung der Summe schreiben Sie die beiden Zahlen rechtsbündig ausgerichtet untereinander und addieren dann die Ziffern unter Berücksichtigung von Überträgen schrittweise von rechts nach links.

Auch ein Algorithmus zur Multiplikation von Zahlen ist Ihnen bekannt, der allerdings erfahrungsgemäß mit viel mehr Rechenaufwand verbunden ist als die Addition. Mit deutlich fortgeschritteneren Algorithmen, die Sie im Informatikstudium kennenlernen können, lässt sich selbst die Multiplikation fast so schnell gestalten wie die Addition, zumindest für sehr große Zahlen.

Korrektheit und Effizienz von Algorithmen

Bei der Entwicklung von Algorithmen gibt es viel Spielraum für mathematische Kreativität. Dabei müssen allerdings stets zwei fundamentale Qualitätsstandards beachtet werden: Korrektheit und Effizienz. Sinnvolle Algorithmen müssen nämlich *korrekt* sein, also auf gültigen Eingaben verlässlich korrekte Ausgaben erzeugen. In der For-

schung können wir die Korrektheit von Algorithmen mit mathematischen Methoden oft zweifelsfrei nachweisen. Anders als empirische Methoden ermöglichen solche mathematischen Beweise, die Korrektheit eines Algorithmus für alle möglichen Eingaben nachzuweisen.

Zudem sollten Algorithmen auch *effizient* sein, also in ihrer Ausführung sparsam mit Ressourcen wie Rechenzeit, Arbeitsspeicher oder Energie umgehen. Idealerweise sollte sich etwa die Rechenzeit im Verhältnis zur Eingabelänge annähernd proportional verhalten, d.h. wird der Umfang der Eingabe verdoppelt, dann sollte der Algorithmus auch nur doppelt so lange rechnen.

Effizienz kann empirisch oder mathematisch nachgewiesen werden; hier werden interessante Diskrepanzen deutlich, da einige wichtige Algorithmen in der Praxis viel effizienter rechnen als durch gegenwärtige mathematische Modelle erklärbar. Es gibt in der Forschung zu Algorithmen also noch eine Kluft zwischen der theoretischen und praktischen Analyse für mehrere fundamentale Berechnungsprobleme.

Übrigens: Selbst innerhalb rein theoretischer Modelle kann die Analyse mancher Algorithmen schier unmöglich sein. Betrachten Sie etwa den folgenden absurd einfachen Algorithmus: Beginnend von einer Eingabezahl n , ersetzen Sie diese durch $n/2$ (falls sie gerade ist)

oder durch $3n+1$ (falls sie ungerade ist), und wiederholen Sie diesen Prozess mit dem neuen Wert von n in einer Schleife. Sollte dabei die Zahl 1 erreicht werden, beenden Sie den Vorgang.

Dieser Algorithmus durchläuft beispielsweise von $n=5$ aus die Zahlen 5, 16, 8, 4, 2, 1. Von $n=6$ aus erreicht er 6, 3, 10, 5, 16, 8, 4, 2, 1.

Seit knapp 90 Jahren ist trotz ernsthafter Anstrengungen kein Beweis dafür bekannt, dass dieser Algorithmus tatsächlich für jede Eingabe die Zahl 1 erreicht. (Stichwort: Collatz-Vermutung.)

Komplexitätstheorie gegen Algorithmen

Komplementär zur Entwicklung und Analyse von Algorithmen beschäftigen wir uns am Lehrstuhl auch mit der *Komplexitätstheorie*. Man könnte dieses Gebiet als eine negative Perspektive auf Algorithmen beschreiben: Aus pragmatischer Sicht geht es nämlich darum, korrekte und effiziente Algorithmen für manche Berechnungsprobleme *auszuschließen*. Allgemeiner verfolgt die Komplexitätstheorie das Ziel, abstrakte Zusammenhänge zwischen Berechnungsproblemen zu verstehen und so Rückschlüsse auf ihre inhärente Komplexität zu ermöglichen. Sind manche konkrete Berechnungsprobleme etwa grundsätzlich

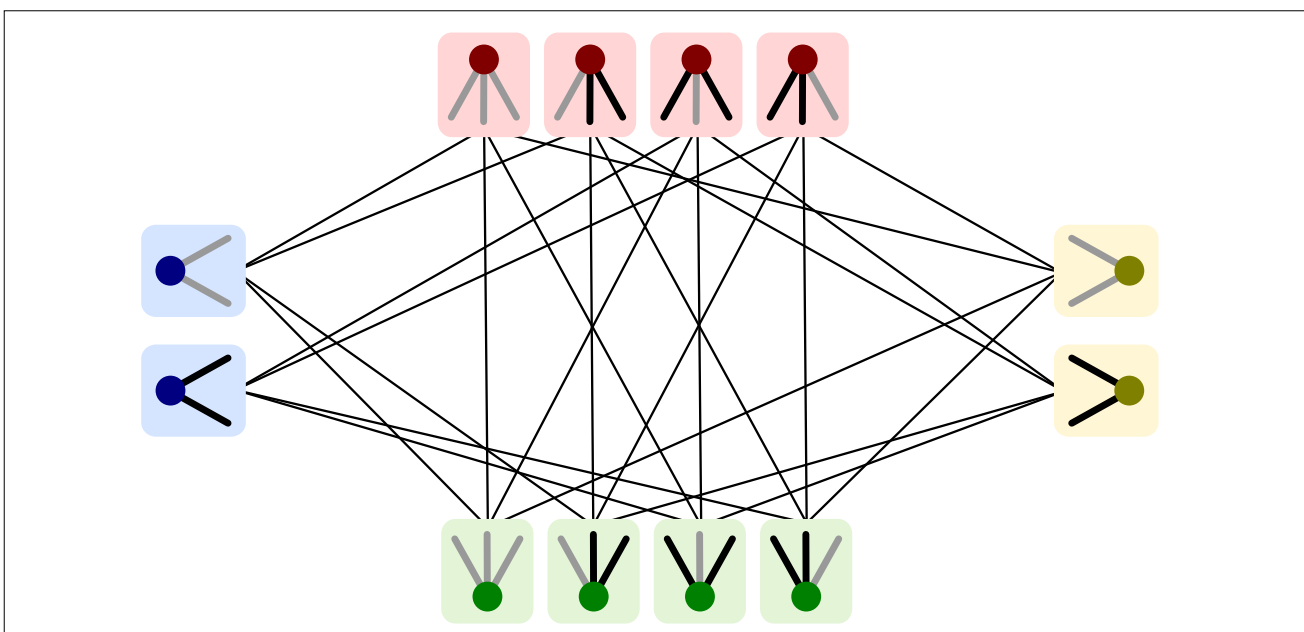
schwerer als die Addition von Zahlen? Wie Sie wohl erwarten: Ja.

P versus NP

Die mutmaßlich wichtigste ungelöste Frage in der Komplexitätstheorie ist die sogenannte »P versus NP«-Frage, eine von sieben »Millennium-Fragen« der Mathematik, deren Lösung mit jeweils einer Million Dollar Preisgeld dotiert ist.

Im Kern geht es hier um die Frage, ob eine bestimmte Klasse besonders interessanter Berechnungsprobleme durch effiziente Algorithmen lösbar ist. Diese Klasse wird mit *NP* bezeichnet und umfasst beispielsweise das *Teilsommen-Problem*, bei dem die Eingabe eine Liste von Zahlen und ein zusätzlicher Zielwert ist, und die erwünschte Ausgabe eine Auswahl von Eingabezahlen ist, die exakt auf den Zielwert summieren, so wie in der *TechnikTempel-Eskapade*. Auch abstrakte Puzzles wie verallgemeinerte Varianten von Sudoku und praktische relevante Probleme sind in der Klasse *NP* enthalten.

All den Problemen in *NP* ist gemein, dass sie nach Lösungen fragen, die sich schnell *verifizieren*, aber nicht notwendigerweise auch schnell *finden* lassen: Manche Sudoku mögen sehr schwer zu lösen sein, etwa nur durch viel brachiales Durchprobieren, es lässt sich aber immer schnell überprüfen, ob eine angebliche Lösung korrekt ist.



1 Ein Beispiel für einen sogenannten CFI-Graphen, benannt nach Jin-Yi Cai, Martin Fürer und Neil Immerman. Die Kästchen entsprechen den Teillösungen eines bestimmten Problems aus der Graphentheorie, und zwei Kästchen sind miteinander durch eine Kante verbunden, wenn sie kompatible Teillösungen enthalten. Über eine unerwartete Verbindung stellte sich Anfang dieses Jahres heraus, dass sich CFI-Graphen für komplexitätstheoretische Begründungen der Schwierigkeit von Berechnungsproblemen nutzen lassen. Auch unser Lehrstuhl beschäftigt sich in mehreren aktuellen Projekten mit diesen faszinierenden Objekten aus der diskreten Mathematik.



Kurzgesagt geht es in der »P versus NP«-Frage also darum, ob das Lösen von Sudoku-artigen Berechnungsproblemen durch Algorithmen schwerer als das Überprüfen einer Lösung ist. Wie Sie vielleicht nicht erwarten, kann dies seit über 50 Jahren niemand beantworten. Wir widmen uns am Lehrstuhl Varianten und Spezialfällen dieser Fragestellung, die sich erfahrungsgemäß als zugänglicher herausgestellt haben.

Der Algorithmen-Begriff in der Gegenwart

Dieser Artikel hat Algorithmen aus der wissenschaftlich-theoretischen Sicht dargestellt, die am Lehrstuhl vertreten ist. An dieser Stelle ist vielleicht eine Abgrenzung zur neueren Allgemeinsprache angebracht, wo der Algorithmen-Begriff oft wahlweise als Synekdoche für spezifische Konzepte des maschinellen Lernens (etwa

für sogenannte *Empfehlungssysteme* von YouTube, Spotify, oder Amazon) oder für Informationstechnologie im Allgemeinen benutzt wird, gelegentlich unter Andeutung einer gewissen Autonomie oder Macht. Am Lehrstuhl beschäftigen wir uns mit Algorithmen im Sinne von Lösungsmethoden für klar formalisierte und abstrakte Berechnungsprobleme, wie in diesem Artikel beschrieben.

Über den Lehrstuhl

Nach dieser allgemeinen Einführung in das Thema folgen noch einige Worte zum Lehrstuhl für Algorithmen und Komplexitätstheorie, der im Oktober 2023 an der Universität Regensburg gegründet wurde. Unsere Forschung widmet sich vorwiegend Berechnungsproblemen auf abstrakten Netzwerken (sogenannten *Graphen*), die Relationen zwischen Objekten abbilden,

wie sie etwa in sozialen Netzwerken oder Straßennetzen auftreten. Für solche Probleme entwickeln wir Algorithmen mittels mathematischer Methoden aus der Algebra, indem wir beispielsweise Daten in Polynome über bestimmten Körpern übersetzen und diese Daten durch algebraische Manipulationen der assoziierten Polynome weiterverarbeiten. Dieser Ansatz geht dann nahtlos in die sogenannte *algebraische Komplexitätstheorie* über.

Gegenwärtig arbeitet neben Prof. Dr. Radu Curticapean auch Dr. Cornelius Brand als wissenschaftlicher Mitarbeiter am Lehrstuhl, unterstützt von Sekretärin Annett Reisinger. Weitere Mitarbeiter werden folgen. Neben der Grundvorlesung zu Algorithmen und Datenstrukturen für Studierende im zweiten Semester werden wir auch weiterführende Vorlesungen und Seminare anbieten. Wir stehen Ihnen sehr gerne zur Verfügung, wenn Sie mehr erfahren möchten!

DREI FRAGEN ...

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

In meinem Forschungsprojekt *CountHom*, das von der Europäischen Union im Rahmen eines *ERC Starting Grants* mit 1,5 Mio. Euro gefördert wird, werden spannende Verbindungen zwischen verschiedenen kombinatorischen Berechnungsproblemen untersucht. Einige fundamentale Berechnungsprobleme, die das Testen und Zählen kleiner Muster in Netzwerken betreffen, wurden nämlich früher unabhängig voneinander untersucht, können aber aus der richtigen Perspektive als ein und dasselbe Problem aufgefasst werden! Diese verallgemeinernde Perspektive wird ermöglicht durch

sogenannte *Homomorphismen*, strukturerhaltende Abbildungen aus der Mathematik. In den nächsten Jahren möchte ich am Lehrstuhl mittels Homomorphismen eine allgemeinere Theorie und dadurch auch optimale Algorithmen für solche Berechnungsprobleme entwickeln.

Was ist Ihre größte fachliche Herausforderung?

In der Komplexitätstheorie dürfte das Millennium-Problem »P versus NP« eine der größten Herausforderungen sein, wie im Text detaillierter beschrieben. Mir ist allerdings kein Forschungs-Team in meinem Gebiet bekannt, das sich ernsthafte Hoff-

nungen macht, dieses Problem in absehbarer Zeit zu lösen.

Welchen Beitrag liefert Ihre Forschung für die gesellschaftliche Entwicklung?

Konkrete Anwendungen finden sich in den Bereichen der Netzwerkanalyse und Bioinformatik, wo das Testen und Zählen von Mustern in Netzwerken relevant ist. Allgemeiner hoffe ich, dass die Arbeit des Lehrstuhls auch andere Forschende und Studierende innerhalb der Informatik dazu inspiriert, weitere Zusammenhänge zwischen Berechnungsproblemen und mathematischen Theorien zu suchen.



Foto © privat

Prof. Dr. Radu Curticapean

ist Inhaber des neu gegründeten Lehrstuhls für Algorithmen und Komplexitätstheorie und hat den Dienst an der Universität Regensburg am 1. Oktober angetreten. Curticapean hat 2015 an der Universität des Saarlandes promoviert. Seine Dissertation zur Komplexität von Zählproblemen mit dem Titel "The simple, little and slow things count" wurde von der Gesellschaft für Informatik und der European Association for Theoretical Computer Science ausgezeichnet. Seit 2015 bis Dienstantritt in Regensburg hat Curticapean im Ausland gelebt und gearbeitet, davon zwei Jahre als Post-Doc am Forschungsinstitut für Informatik der Ungarischen Akademie der Wissenschaften (Budapest, Un-

garn), zwei Semester als Research Fellow am Simons Institute for the Theory of Computing an der Universität Berkeley (Berkeley, USA) und fünf Jahre an der IT University of Copenhagen (Kopenhagen, Dänemark), dort als Post-Doc, Assistant Professor und schließlich als Associate Professor.

Evolution in Datenbanken und Data Engineering Workflows

Prof. Dr. Meike Klettke

Wenn Sie den Begriff Evolution hören, denken Sie sicher an Biologie. Wir alle haben in der Schule Darwins Evolutionstheorie kennengelernt, die unter anderem beinhaltet, dass Veränderungen ständig passieren und eine Quelle der weiteren Entwicklung sind.

Den Begriff der Evolution verwenden wir auch in der Informatik. Auch hier sind Änderungen ständig auf der Tagesordnung. Programme und Daten, die nicht aktualisiert werden, altern. Das passiert, obwohl digitale Objekte wie Software und elektronisch gespeicherte Daten nicht physisch verschleißen; sie werden dennoch über die Zeit immer weniger verwendbar, wenn sie nicht kontinuierlich aktualisiert werden. Am deutlichsten wird es sicher sichtbar, wenn man einmal eine Webseite oder ein Computerprogramm findet, die seit mehreren Jahren nicht verändert wurden, sie erscheinen uns heute hoffnungslos veraltet.

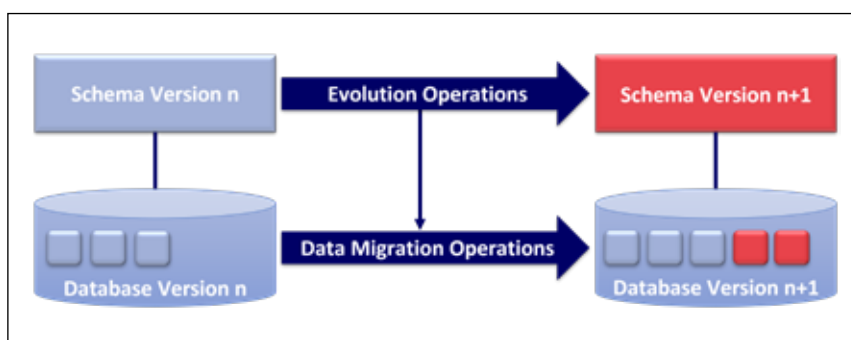
Sowohl Programme als auch Daten und deren Speicherung müssen sich also ständig weiterentwickeln, das nennen wir in der Informatik Evolution von Software und Daten. Im Mittelpunkt der Forschungsthemen des Lehrstuhls Data Engineering steht die kontinuierliche Evolution der Datenbestände. Diese ist notwendig, damit die

Daten aus der Vergangenheit, die wir auch als legacy Daten oder Altdatenbestände bezeichnen, heute noch auswertbar bleiben. Oft sind die Daten sogar wesentlich länger im Einsatz als die Software, die diese verwendet. In vielen Bereichen müssen die Daten über lange Zeit aufbewahrt werden, wie zum Beispiel bei der gesetzlich vorgeschriebenen Aufbewahrungspflicht für Rechnungsdaten. Hier ist es gefordert, dass diese Daten mindestens 10 Jahre verfügbar sein müssen. In anderen Anwendungen werden Daten noch über deutlich längere Zeiträume verwendet, auch weil deren Erstellung mit Kosten verbunden war. Vor allem aber erhofft man sich durch die Auswertung der Daten einen Erkenntnisgewinn, der sich auch daraus ergeben kann, dass man in Firmen und in der Wissenschaft Auswertungen über längere Zeiträume durchführen kann. In ökonomischen oder gesellschaftlichen Fragen möchten wir bestimmte Daten zusammen mit den Vergleichswerten des Vorjahres auswerten können; in wissenschaftlichen Experimenten, in denen wir zum Beispiel Umweltdaten erfassen, möchten wir diese über sehr lange Vergleichszeiträume auswerten können.

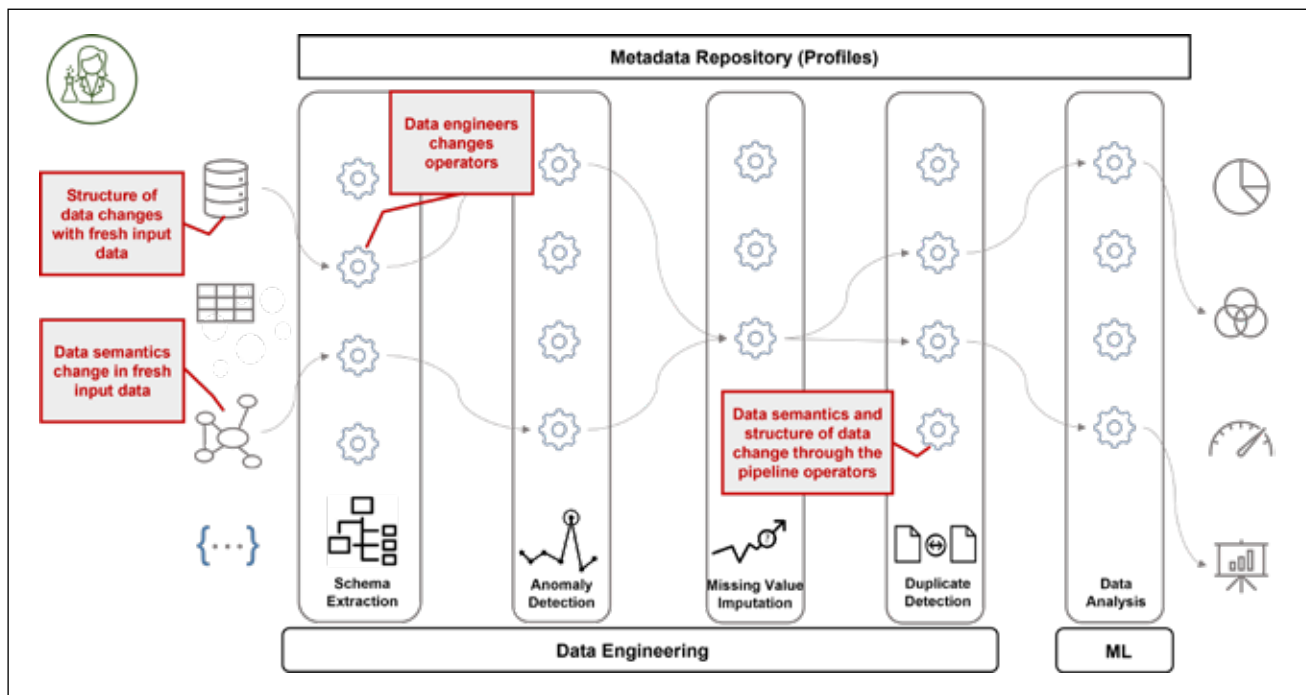
Teilweise sind gespeicherte Firmendaten wie auch Mess- und Experimentdaten viele Jahre alt. In einigen Forschungsinstituten reicht die Erhebung sogar etliche Jahrzehnte zurück. Nun kann man sich natürlich leicht vorstellen, dass sich seitdem einiges in der Informatik geändert hat, nicht nur die Hardware, sondern auch die Speicherformate, die Datenmodelle und die Schemata der Daten haben sich weiterentwickelt.

Um solche legacy Daten zu aktualisieren, entwickelt unsere Gruppe Verfahren, die Daten zwischen verschiedenen Datenmodellen transformieren und die Evolution von Daten in die aktuellen Schemata durchführen. Schematisch ist das in Abbildung 1 dargestellt. Dafür haben wir eine Evolutionssprache entwickelt, die sowohl einfache Erweiterungen und Änderungen mit den Operationen add, delete und rename durchführen kann, als auch komplexe Umstrukturierungen von Daten durch die Operationen copy, move, split und merge. Mit diesen Operationen ist ein komplettes Refactoring von Datenbanken möglich.

Die Umsetzung dieser Evolutionssprache für verschiedenen Datenbankmanagementsysteme ist eine weitere Forschungsaufgabe. So wurde die Umsetzung der Evolutionsoperationen für JSON-Daten entwickelt. JSON steht für JavaScript Object Notation und ist ein gut lesbares Datenmodell zur Darstellung von Objekten, damit ist es also sehr flexibel einsetzbar. JSON Datenbanksysteme sind Datenbankmanagementsysteme, die solche JSON Objekte performant speichern können. Das bekannteste System dieser Klasse ist MongoDB [2]. Da das Datenmodell komplexer ist als in den relationalen Datenbanken, die alle Daten tabellarisch speichern



1 Schemaevolution und Datenmigration für NoSQL-Datenbanken und Graphdatenbanken



2 Evolution von Data Engineering Workflows

und SQL als Anfragesprache dazu einsetzen, gehören JSON Datenbanken zu den sogenannten NoSQL Datenbanken. NoSQL steht dabei für »not only SQL«. Diese Datenbanksysteme sind gegenwärtig häufig im Einsatz, insbesondere Data Science Anwendungen verwenden das Speicherformat. Für die darin gespeicherten Daten entwickeln wir Verfahren zur Evolution, die eingesetzt werden können, um die Daten in neue Versionen zu migrieren und so für die Auswertung bereitzustellen.

Darüber hinaus werden bei uns am Lehrstuhl Evolutionsoperationen für Graphdatenbanken entwickelt [3]; hier erfolgt die Umsetzung für Neo4J [4]. Graphdatenbanksysteme sind eine andere Technologie zur Datenspeicherung. Ihre Bedeutung hat in den letzten Jahren stark zugenommen, Graphen werden als die natürliche Form der Datenspeicherung in der Logistik, zur Darstellung von Rechnernetzen und anderen Kommunikationswegen sowie in sozialen Netzwerken angesehen. Da immer mehr Daten als Graphen gespeichert werden, ist auch hier die Evolution, also die Überführung in die aktuellen Strukturierungen erforderlich. Wenn man solche Daten aktualisiert, dann soll das Verfahren dazu so weit wie möglich automatisiert werden. Wir arbeiten gegenwärtig an Methoden, damit die einmal für einen Datenbestand beschriebenen Evolutionsoperationen, also eine Folge aus add, rename, copy, merge

usw., auf alle Daten in einem Schritt angewendet werden können.

Das ist in Abbildung 1 schematisch für beide Klassen von Datenbanksystemen zu sehen. Entwickler:innen sollen mit dem geringstmöglichen Aufwand Daten kontinuierlich weiterentwickeln können, indem sie die gewünschten Änderungen mit der formalen Sprache (Schemaevolution) beschreiben. Der sich anschließende Schritt der Datenmigration erfolgt dann automatisch durch das Verfahren.

Ebenso ist eine ressourcenschonende Evolution möglich, die nur die benötigten Daten »on demand« aktualisiert. Das heißt also, dass erst in dem Moment, in dem Daten benötigt werden, geprüft wird, ob zur Beantwortung einer aktuellen Anfrage oder Lösen einer Aufgabe auch legacy Datenbestände existieren. Dabei wird nur der Anteil der Daten ausgewählt und aktualisiert, der für die Beantwortung der Frage benötigt wird. Es ist leicht vorstellbar, dass das technologisch etwas aufwändiger ist als das vollständige Aktualisieren des Gesamtdatenbestandes. Die dazu von uns entwickelte Methode ist ein Query Rewriting [5], bei dem eine Anfrage auf viele verschiedene (frühere) Versionen der Datenspeicherung verteilt wird und die Teilergebnisse zusammengesetzt werden. Beide Verfahren (vollständige Evolution und Evolution on demand) wurden in zwei DFG-Projekten in enger Zusammenarbeit mit den Arbeitsgruppen von Prof. Uta Störl

(FernUniversität in Hagen) und Prof. Stefanie Scherzinger (Universität Passau) für Daten in verschiedenen Datenmodellen entwickelt.

Die Weiterentwicklung der Daten ist nur ein Aspekt, warum in Informatiksystemen eine Aktualisierung notwendig sein kann. In anderen Aufgaben am Lehrstuhl entwickeln wir Verfahren für sogenannte Data Engineering Workflows. Das sind Folgen von Programmen, die Rohdaten, wie zum Beispiel vorhandene Datenbestände, Messdaten, Daten wissenschaftlicher Experimente oder empirische Befragungen aufbereiten. Dabei werden diese Daten kontrolliert, korrigiert, vervollständigt, transformiert, in einheitliche Formate integriert und dabei validiert, sodass sie auswertbar werden. Abbildung 2 zeigt schematisch, wie aus den verschiedenen verfügbaren Data Engineering Algorithmen komplette Workflows zusammengestellt werden können.

Die Auswertungen können auf verschiedene Weise erfolgen, möglich sind Datenbankabfragen, OLAP-Auswertungen, (OLAP steht für On-Line Analytical Processing und beschreibt Analysen auf großen Datenbeständen), Report-Generierungen und Analyseverfahren auf Basis von Machine Learning Methoden. Ein wichtiges Ziel ist es dabei, dass alle durchgeführten Schema- und Datenänderungen nachvollziehbar, reproduzierbar und replizierbar sind. Hierfür kommen in den aktuellen

Arbeiten am Lehrstuhl sogenannte Provenance-Techniken zum Einsatz, die sich sowohl auf die Daten als auch den Workflow selbst beziehen. Mit diesen kann für Anfrageergebnisse bestimmt werden, aus welchen Daten und durch welche Verarbeitungsschritte im Workflow diese entstanden sind. Damit werden die Ergebnisse eines Data Engineering Workflows erklärbar, ebenso werden die Ergebnisse der Workflow Evolution nachvollziehbar.

Unsere momentan laufenden Forschungsarbeiten im Bereich Data Engineering Workflows umfassen mehrere Themen: i) die Entwicklung einer Monitoringkomponente soll Änderungen in den Daten sichtbar machen und das Erkennen von Datenfehlern ermöglichen, ii) im Fall einer Evolution der Daten oder der Workflows wollen wir bestimmen können, inwieweit die Ergebnisse der Auswertung noch mit den Ergebnissen früherer Auswertungen vergleichbar sind, iii) Data Engineering Workflows können auch in der Cloud durchgeführt werden, hierbei ist die Opti-

mierung der Data Engineering Workflows für die jeweiligen Cloud Provider ein aktuelles Thema und iv) nicht zuletzt wenden wir die Data Engineering Techniken auch für spezielle Datenarten wie Eye-Tracking-Daten an.

Auch Data Engineering Workflows benötigen Verfahren zur Evolution, um im Fall von Änderungen Data Engineering Workflows zu aktualisieren und so die Workflows über lange Zeiträume und unter veränderten Rahmenbedingungen einsatzbereit zu halten. In Abbildung 2 sind die verschiedenen Arten von Änderungen, die dabei zu berücksichtigen sind, farblich hervorgehoben. Für die großen Themenbereiche der Optimierung und Evolution von Datenbeständen, sowie der Erstellung, Verwendung, Kontrolle, Protokollierung und Evolution von Data Engineering Workflows entwickeln wir in unseren gegenwärtigen Forschungsarbeiten sowohl die formalen Methoden, die zur Anwendung kommen, als auch die Tools für deren Verwendung [6].

DREI FRAGEN ...

Welches Fachbuch lesen Sie gerade?

Momentan lese ich mehrere Grundlagenbücher über die Programmiersprache Python – immer auf der Suche nach guten Erklärungen und interessanten Beispielen für die Vorlesung Programmierung I für das 1. Fachsemester im Bachelor Informatik und im Bachelor Data Science.

Die Lösung von welchem fachlichen Thema haben Sie sich leichter vorgestellt?

Die Speicherung und Auswertung heterogener Daten. Datenbanksysteme können sehr gut und effizient mit tabellarischen Daten umgehen, viele Informationen sind aber entweder komplexer aufgebaut oder unregelmäßig strukturiert oder beides. Die vollständige Zuverlässigkeit von Datenbanksystemen dennoch zu garantieren, ist eine große fachliche Herausforderung.

Was möchten Sie in den nächsten Jahren entwickeln?

Meine Wunschvorstellung ist es, Werkzeuge für das Data Engineering zu bauen, mit denen Anwendungsexpert:innen ihre Daten selbst bereitstellen können. Bisher erfordern diese Verfahren noch einen großen manuellen Aufwand und speziell für eine Aufgabe programmierte Skripte. Das Ziel und gleichzeitig auch der aktuelle Forschungsgegenstand unserer Gruppe ist es, Werkzeuge zu entwickeln, mit denen diese Schritte leichter ausführbar und intuitiv verständlich werden.

Literatur

- [1] Uta Störl, Meike Klettke: Darwin: A Data Platform for NoSQLSchema Evolution Management and Data Migration, DataPlat@EDBT, 2022
- [2] MongoDB, <https://www.mongodb.com/>
- [3] Dominique Hausler, Meike Klettke and Uta Störl: A Language for Graph Database Evolution and its Implementation in Neo4j, 42nd International Conference on Conceptual Modeling (ER), Forum, 2023
- [4] Neo4j Graph Database & Analytic, Graph Database Management System, <https://neo4j.com/>
- [5] Mark Lukas Möller, Meike Klettke, Andrea Hillenbrand, Uta Störl: Query Rewriting for Continuously Evolving NoSQL Databases. Article, International Conference on Conceptual Modeling (ER conference), 2019
- [6] Meike Klettke, Uta Störl: Four Generations in Data Engineering for Data Science - The Past, Presence and Future of a Field of Science, Datenbank-Spektrum, 2022



Foto © rivat

Prof. Dr.-Ing. habil. Meike Klettke ist Professorin für Data Engineering an der Universität Regensburg. Sie studierte Informatik an der Universität Rostock, promovierte dort 1997 zum Thema „Akquisition von Integritätsbedingungen in Datenbanken“ und habilitierte 2007 mit ihren Arbeiten zu „Modellierung, Bewertung und Evolution von XML-Dokumentkolektionen“. Seit April 2022 leitet sie die Data Engineering Gruppe an der Fakultät für Informatik und Data Science.

IoT-basiertes Prozessmanagement

Mobile Benutzerführung in der digitalen Fabrik

Prof. Dr. Stefan Schöning

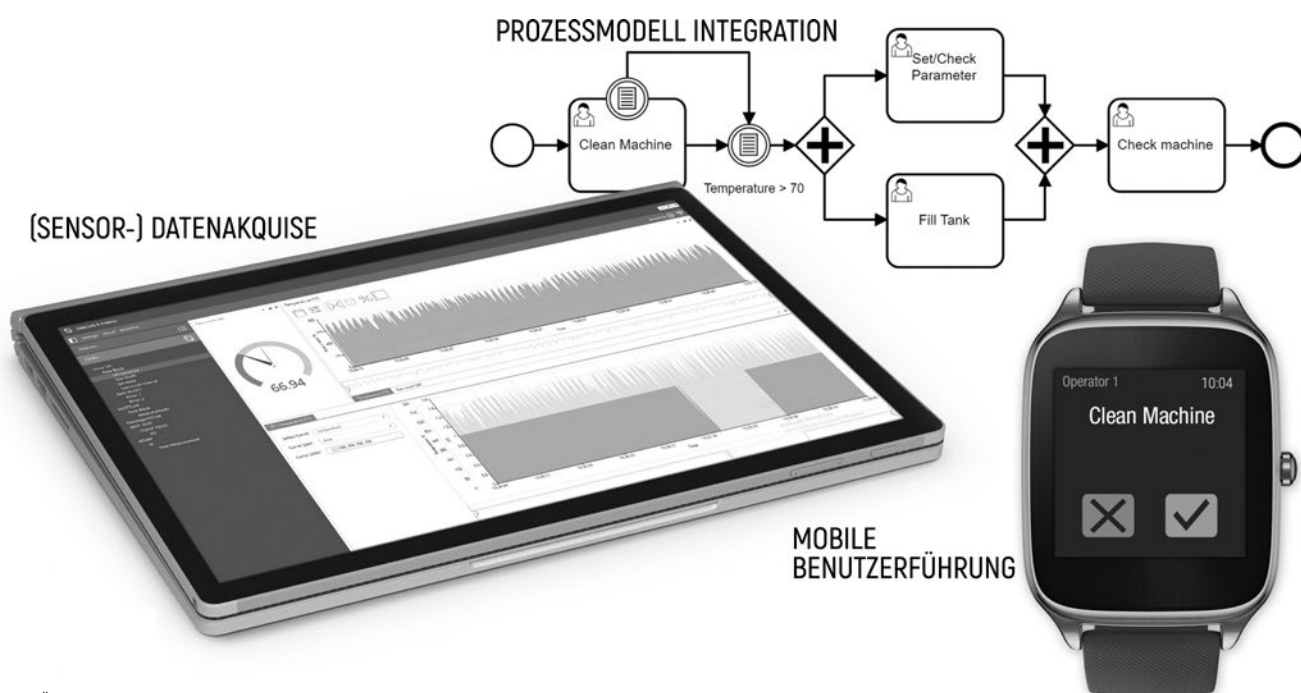
Prozessmanagement in der Produktion

Die durch die »Industrie 4.0«-Initiative geförderte Verschmelzung von Produktion und Informationstechnologie führt im zukünftigen Produktionsumfeld zu einer Allgegenwärtigkeit von rechnergestützten Informationsverarbeitungssystemen, die unter dem Begriff »Ubiquitous Computing« zusammengefasst werden. In der Produktion wird darunter auch die Vernetzung von Produktionsmaschinen mit übergeordneten Informationssystemen verstanden, so dass sämtliche produktionsbeteiligten Objekte über eine mobile Infrastruktur untereinander kommunizieren und interagieren können [6]. Prozessrelevante Informa-

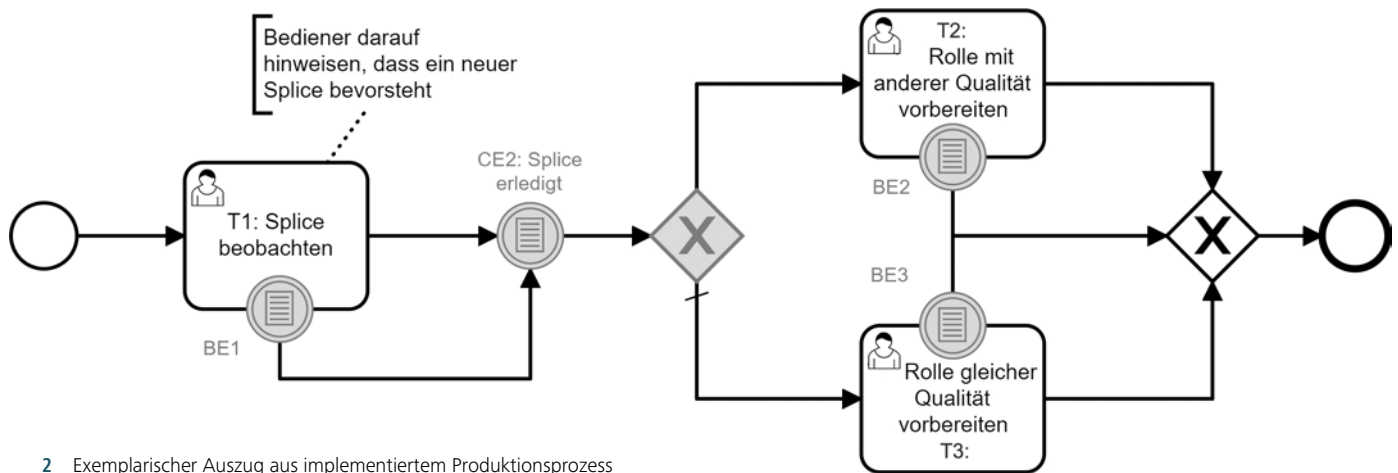
tionen lassen sich so in Echtzeit erfassen, wodurch Prozesse entlang der gesamten Wertschöpfungskette effizient gestaltet werden können [3].

Durch die Möglichkeiten, die mit der digitalen Transformation einhergehen, lässt sich die Produktionsplanung und -steuerung unterstützen. Dabei ermöglichen Prozessführungs-, Prozessüberwachungs-, sowie Analysedaten auf Basis von Internet of Things (IoT) Anwendungen eine umfassende Sicht auf Abläufe in der Produktion. Als ein mögliches Szenario wird ein Produktionsprozess betrachtet, der Material in einer Maschine bearbeitet. Für die Einhaltung der Produktqualität, welche durch Sensoren überwacht wird, sind manuelle Eingriffe und die ständige

Bereitschaft von Bedienern erforderlich. Basierend auf Sensordaten und seiner Erfahrung muss eine Person Entscheidungen treffen, um die Produktqualität durch einen Eingriff sicherzustellen. Ein solches Szenario wird besser beherrschbar, wenn auf digitale Produktions- und Maschinendaten zeitnah von menschlichen Bedienern zugegriffen und verarbeitet werden können. In der Kombination aus IoT-Anwendungen und Prozess- bzw. Produktionssteuerung ergeben sich sowohl Reaktionszeit- und Kosteneinsparungen für Unternehmen als auch Effizienz- und Qualitätssteigerungen, indem die gezielte Vergabe und Abwicklung von Arbeitsaufträgen durchgeführt und überwacht werden kann.



1 Übersicht des Ansatzes



2 Exemplarischer Auszug aus implementiertem Produktionsprozess

Der vorgestellte Ansatz (vgl. Abb. 1) verbindet die »BigData«-Welt der Erfassung und Analyse von IoT- bzw. Sensordaten [4] mit der Technologie des Prozessmanagements [1]. Dabei werden individuelle Arbeitsabläufe von Anlagenbedienern, d. h. Workflows bzw. Prozesse, in einer übersichtlichen Grafik mit Hilfe der Prozess-Standardnotation BPMN (Business Process Modeling and Notation) [5] modelliert. Ein Prozessmodell enthält also Aufgaben, sog. Tasks, die bestimmten Personen bzw. Gruppen von Personen zugeordnet sind. Durch die Anbindung jeglicher Art von Sensor- und Anlagendaten wird die selektive Aktivierung von Tasks auf Basis benutzerdefinierter Bedingungen ermöglicht. Der Ansatz ermöglicht somit die Verknüpfung von Sensoren und Steuerungen mit menschlichen Aufgaben. Modellierte Prozesse können über eine mobile Administrationsoberfläche gestartet und verwaltet werden. Der Status laufender Prozesse kann stets überwacht und nachvollzogen werden. Bereits abgeschlossene Prozesse können nachträglich analysiert und Flaschenhalse sichtbar gemacht werden. Bediener werden über aktuelle Aufgaben auf mobilen Endgeräten wie z. B. Smartwatches benachrichtigt. Hierdurch wird der Bediener bei seinen Tätigkeiten unterstützt, ohne bei der Durchführung seiner Kernaufgaben durch zusätzlichen Aufwand abgelenkt zu werden.

Eine Verknüpfung von Industrie 4.0-Konzepten und Prozessmanagement ermöglicht es, Mitarbeiter:innen bei ihren operationalen Aufgaben zu unterstützen. Es wird dafür gesorgt, dass Daten in Echtzeit zusammenlaufen und in eine interpretierbare Darstellung übersetzt werden. Diese abstrahiert zum einen von den

vielen, proprietären Darstellungen von Maschinen- und Sensordaten und ist daher von der gesamten Belegschaft eines Produktionsbetriebs leicht zu verstehen und zu interpretieren. Die Darstellung der aktuellen Produktionsdaten erfolgt in unserem Anwendungsfall auf Smartwatches. Prozessbeteiligte erkennen auf dem Display ihrer Smartwatch, welche Aufgaben anstehen, wie beispielsweise Wartungsarbeiten, Materialnachlieferung oder Behebung einer Maschinenstörung. Das Prozessmanagement-System löst Aufgaben aus und teilt sie einer geeigneten Person zu, unabhängig davon, wo sich Maschinenbediener, Wartungs- oder Servicemitarbeiter gerade – innerhalb der Produktionshalle – aufhalten. So wissen sie stets über den Status aller Anlagen und ihrer Aufgaben Bescheid. Dies verkürzt die Reaktionszeiten und reduziert Maschinenstillstände aufgrund verspäteter Reaktionen des Bedienerpersonals.

Schritt für Schritt zur IoT-basierten Ablauforganisation

Im ersten Schritt werden Produktionsprozesse, Arbeitsabläufe von beteiligten Bediener:innen sowie notwendige Daten und Informationen aus Maschinen, Steuerungen und Sensoren in einer standardisierten, graphischen Spezifikationsprache für Abläufe (z. B. BPMN) [5] modelliert und digitalisiert. Auf Basis einer Prozess-Automatisierungskomponente können die resultierenden Prozessmodelle schließlich ausgeführt werden und konkrete Aufgaben an Mitarbeiter:innen auf Basis aktueller Anlagen- und Sensordaten verteilt werden. Die mobile Zuteilung von

Aufgaben an Mitarbeiter:innen setzt das Vorhandensein einer mobilen, verteilten Netzwerkinfrastruktur voraus, d. h. beispielsweise in Form einer drahtlosen Client-Server Architektur. Auf diese Weise können gerade notwendige Aufgaben vom Planungssystem, z. B. via WLAN an diverse mobile Geräte (z. B. Smartwatches) gesendet werden, welche jeweils bestimmten Anlagenbediener:innen zugewiesen sind.

Überschreitet beispielsweise der aktuelle Wert eines Temperatursensors eine bestimmte Warngrenze, wird – wie vorher im Prozessmodell definiert sein muss – eine bestimmte Arbeitsaufgabe aktiviert und an die mobilen Geräte von Bediener:innen einer bestimmten, vordefinierten Gruppe gesendet. Das System sorgt automatisch dafür, dass Mitarbeiter:innen mit der passenden Qualifikation den Arbeitsauftrag erhalten. Damit die unterschiedlichen Aufgaben zielgerichtet zugewiesen werden können, muss ein Unternehmen Mitarbeiter:innen bestimmten Gruppen auf Basis von Qualifikationen oder Schichtgruppen zuteilen.

Mitarbeiter:innen werden auf diese Weise mittels haptischer, akustischer oder visueller Signale auf dem dedizierten mobilen Gerät auf neue Aufgaben hingewiesen, die sie annehmen oder ablehnen können. Ist eine konkrete Aufgabe von einer Person angenommen, wird diese allen anderen Personen der Gruppe entzogen. Falls eine Person eine angenommene Aufgabe nicht mehr erfüllen kann, kann sie diese auch eigenständig wieder für alle zuvor festgelegten Gruppen freigeben und entsprechend selbst wieder neue Aufgaben empfangen. Auf diese Weise lassen sich alle Arbeiter:innen mit mobilen Geräten gruppenspezifisch kontaktieren, um Aufgaben und Arbeitsanweisungen zu verteilen.

Anbindung von IoT-Variablen an Prozessmanagement-Systeme

Aktuelle Daten von IoT-Objekten werden an das Prozessmanagement-System gesendet. Hierbei ist zu berücksichtigen, dass bestimmte IoT-Daten nicht an alle laufenden Prozesse gesendet werden, sondern nur an diejenigen, welche auch eine bestimmte IoT-Variable referenzieren. Aus diesem Grund muss eine Abbildung zwischen IoT-Variablen und verwendeten Prozessmodellen etabliert werden. Nachdem eine derartige semantische Abbildung zwischen Daten und Modell definiert worden ist, werden jeweils aktuelle Werte der referenzierten Variablen an das Prozessmanagement-System gesendet. Die technische Umsetzung der Anbindung von Daten an laufende Prozesse wird dabei von der IoT-Infrastruktur übernommen. Zu beachten ist hierbei, dass die Abstraktionslücke zwischen teilweise hochfrequenten IoT- und Sensordaten und der Welt der menschlichen Aufgaben eines Geschäftsprozesses überbrückt werden muss.

Basierend auf der etablierten Echtzeit-Verbindung von IoT-Objekten und eines Prozessmanagement-Systems können IoT-Variablen auf verschiedene Weise in Prozessmodellen referenziert werden. Damit das Prozessmanagement-System Aktivitäten auf Basis der aktuellen Werte von IoT-

Variablen aktivieren oder beenden kann, muss das Modell bestimmte Informationen beinhalten. Dieser Schritt erfordert daher, gegebene Prozessmodelle von Arbeitsabläufen zu erweitern. Die Sprache BPMN enthält verschiedene Konstrukte, die ein Referenzieren von IoT-Variablen in Prozessmodellen erlauben. Trigger Events können beispielsweise angewendet werden, um Aufgaben oder ganze Prozesse zu aktivieren, wenn eine bestimmte Bedingung auf den aktuellen Werten von IoT-Variablen erfüllt ist. Beim sog. *Intermediate Conditional Catch Event* wird die Ausführung an einem Punkt angehalten, bis eine bestimmte Bedingung erfüllt ist. Anschließend wird der Prozess mit der nächsten Aufgabe fortgesetzt. *Boundary Events* können dazu verwendet werden, Aufgaben abzubrechen bzw. zu entziehen, wenn referenzierte IoT-Variablen eine hinterlegte Bedingung erfüllen. Ein *Conditional Start Event* startet gesamte Prozesse, z. B. das Auffüllen eines Teilleagers, auf Basis der Werte von verknüpften IoT-Werten.

Anwendung in Industrie

Die vorgestellten Konzepte wurden unter anderem in Produktionsanlagen für Wellpappe umgesetzt. Aufgrund zunehmender Automatisierung und Betriebspersonal-

reduktion wird eine derartige Produktionsstraße durch immer weniger Personal abgedeckt. Interaktionen mit der bis zu 140 m langen Anlage erfordern deshalb längere Wege zu Bedienterminals und führen zu verzögertem Informationsfluss. Die verlängerten Reaktionszeiten sind häufig Auslöser erhöhter Ausschussmengen.

Sämtliche Produktionsabläufe und notwendige Tätigkeiten von Bedienern werden vorab in Form von BPMN-Prozessmodellen abgebildet. Die modellierten Prozesse referenzieren dabei zahlreiche verschiedene Sensorvariablen in Form von bedingten Ereignissen (*Intermediate Conditional Catch Events*), wodurch bei bestimmten Sensorwerten neue manuelle Aufgaben ausgelöst werden oder laufende Aufgaben abgebrochen (*Boundary Events*) werden. Ein exemplarischer Auszug aus den modellierten Prozessen ist in Abb. 2 dargestellt. Das Modell enthält sowohl menschliche Aufgaben als auch IoT-basierte Ereignisse und Entscheidungspunkte, welche die Abarbeitung des Prozesses während der Ausführung steuern. Bei der Umsetzung des Konzepts wurde ein Rollenmodell implementiert, in dem jeder Bediener den von ihm zu überwachenden Bereich der Anlage definiert. Im Fall des abgebildeten Verklebungsprozesses (»Splice«) erhalten Bediener im Bereich, in dem die Papierrollen in die Anlage ein-



3 Mobile Prozesssteuerung auf der Smartwatch

gespeist werden, durch die kontinuierliche Überwachung Aufgaben zum nächsten anstehenden Rollenwechsel oder über auftretende Defekte in der Wellpappenbahn. Um Bediener in Echtzeit auf anstehende Aufgaben hinzuweisen, wurde das gesamte Personal mit Smartwatches (siehe Abb. 3) ausgestattet und Gruppen entsprechend dem jeweiligen Tätigkeitsbereich zugeteilt. Durch die beschriebene Umsetzung konnten Wege- und Reaktionszeiten signifikant verkürzt, der Ausschuss reduziert und die Qualität gesteigert werden. Auch auf die Anzahl der Anlagenstillstände wirkte sich der Ansatz positiv aus. Diese konnten deut-

lich früher erkannt und proaktiv behoben werden. Somit konnte eine nachhaltige Steigerung der Gesamtanlageneffektivität erzielt werden.

Referenzen

Dumas, M., La Rosa, M., Mendling, J., Reijers, H. A. (2013). *Fundamentals of business process management* (Vol. 1, p. 2). Heidelberg: Springer.

Eckert, M., Bry, F. *Informatik Spektrum* (2009) 32: 163. <https://doi.org/10.1007/s00287-009-0329-6>

Günthner W.; Wöfle, M.; Fischer, R.: *Wearable Computing und RFID in Produktion und Logis-*

tik – Ansätze zur bereichsübergreifenden Nutzung digitaler Informationen. *Logistics Journal: Nicht referierte Veröffentlichungen*, 2011.

Meroni, G., Baresi, L., Montali, M., Plebani, P. (2018). Multi-party business process compliance monitoring through IoT-enabled artifacts. *Information Systems*, 73, 61-78.

Object Management Group (OMG), *BPMN Spezifikation* (2011). Version 2.0. Final Adopted Specification.

Schönig, S.; Jablonski, S.; Ermer, A.; Aires Silva, A. P.: *Digital Connected Production: Wearable Manufacturing Information Systems*, In: Debruyne, C.; Panetto, H.; Weichhart, G.; Bollen, P.; Ciuciu, I.; Vidal, M.; Meersman, R.(ed.): *On the Move to Meaningful Internet Systems: OTM 2017 Workshops - Rhodes, Greece: Springer (Lecture Notes in Computer Science; 10697)*, S. 56-65. 2017.

Schönig, S., Ackermann, L., Jablonski, S., Ermer, A.: *An Integrated Architecture for IoT-Aware Business Process Execution*, In: Gulden, J., Reinhartz-Berger, I., Schmidt, R., Guerreiro, S., Guédria, W., Bera, P. (Eds.) *Enterprise, Business-Process and Information Systems Modeling: 19-34*, 2018.

Schönig S., Aires A.P., Ermer A., Jablonski S. (2018) *Workflow Support in Wearable Production Information Systems*. In: Mendling J., Mouratidis H. (eds) *Information Systems in the Big Data Era. CAiSE 2018. Lecture Notes in Business Information Processing*, vol 317. Springer, Cham

Singh, M., Rajan, M. A., Shivraj, V. L., Balamuralidhar, P. (2015). *Secure MQTT for Internet of Things (IoT)*. In *Communication systems and network technologies (CSNT), 2015 fifth international conference on* (pp. 746-751). IEEE.

Van der Aalst, W. M. (2016). *Process mining: data science in action*. Springer.



Foto © privat

Prof. Dr. Stefan Schönig

ist Professor am Department für Information Systems an der Universität Regensburg. Zuvor war er Post-Doktorand am Institut für Management Information Systems an der Wirtschaftsuniversität Wien. Er erhielt sowohl den Master-Abschluss (mit Auszeichnung) in Angewandter Informatik (Ingenieurwissenschaften/Informatik) als auch den Dokortitel in Informatik an der Universität Bayreuth.

Er hat einen fundierten Hintergrund im Bereich BPM/Process Mining und arbeitet seit über 12 Jahren in diesem Bereich. Stefan Schönig hat eine Vielzahl an Industrie- und Forschungsprojekten durchgeführt, die sich mit Process Mining, Prozessanalyse, Pro-

zessüberwachung und Prozessintelligenz im Kontext industrieller IoT-Umgebungen befassen. Er hat zahlreiche Publikationen im Bereich Geschäftsprozessmanagement und Informationssysteme in internationalen Konferenzen und Fachzeitschriften veröffentlicht.

Cyber Threat Intelligence

Gemeinschaftliche IT-Sicherheit durch den Austausch von Informationen

Johannes Grill, Dr. Daniel Schlette, Prof. Dr. Günther Pernul

Cybersecurity ist ein Thema von gesellschaftlicher Relevanz

Die tägliche Verwendung von Informationssystemen und Daten im privaten und beruflichen Kontext ist elementarer Bestandteil unserer Gesellschaft. Dadurch hat die Sicherheit dieser Systeme und der Schutz persönlicher Daten sowohl für Unternehmen als auch für gesellschaftliche Institutionen und Privatpersonen eine hohe Relevanz. In den letzten Jahren haben erfolgreiche Angriffe in immer größerem Ausmaß gezeigt, dass Cyberkriminelle zum Beispiel mittels Phishing E-Mails oder Ransomware vorhandene Schwachstellen zu ihrem Vorteil ausnutzen. Neben Spionage und Datendiebstahl hat sich ein profitables, arbeitsteiliges Ökosystem um die Verschlüsselung von Daten und die Lösegelderpressung entwickelt (Verizon. Data Breach Investigations Report 2023). Ins Visier sind dabei auch kritische Netzwerke wie Industrieanlagen und staatliche Behörden gerückt. Aktuelle Zahlen belegen, dass weltweit mehr als jedes zweite Unternehmen Opfer eines Angriffes mit erheblichen Schäden wurde (Hiscox. Cyber Readiness Report 2023). Auch viele kleinere Unternehmen verwenden Anwendungen (z. B. E-Mail, Dateitransfer-Dienste, oder Webanwendungen), die ein Eintrittstor für Cyberkriminelle darstellen können. Das weltweite Phänomen lässt sich leider auch in Deutschland beobachten und hat durch geopolitische Konflikte an Bedeutung gewonnen (BSI. Die Lage der IT-Sicherheit in Deutschland 2022). Dabei finden vermehrt auch Angriffe statt, die darauf abzielen, Systeme und Dienste lahmzulegen (De-

nial of Service) und erst in einem zweiten Schritt auf die Kompromittierung aus sind.

Um diesen Bedrohungen in einer vernetzten Welt entgegenzutreten, wird unter den Begriffen Cybersecurity, IT-Sicherheit, oder Informationssicherheit das Wissen rund um die Gewährleistung von Vertraulichkeit, Integrität, und Verfügbarkeit der Systeme (bestehend aus Hardware und Software) und Daten zusammengefasst. Da heutzutage Industrieanlagen (bezeichnet als Operation Technology, OT) und klassische Informationssysteme (IT) zunehmend vernetzt sind, existieren Sicherheitsinformationen über heterogene Systeme und vielfältige Arten von Bedrohungen. Trotz verstärkter Bemühungen im Bereich der Cybersecurity nimmt jedoch die Anzahl der erfolgreichen Angriffe weiter zu. Aus diesem Grund rücken nun vermehrt kollaborative Ansätze basierend auf Sicherheitsinformationen in den Fokus.

Cyber Threat Intelligence beschreibt vielfältige Sicherheitsinformationen

Daten-basierte Ansätze sind in der IT allgegenwärtig und dienen der Optimierung von Systemen, darauf aufbauenden Anwendungsfällen und der Gewinnung von neuen Erkenntnissen. Auch im Bereich Cybersecurity besteht durch die Gesetzgebung sowie die konstante Gefahrenlage die Notwendigkeit sicherheitsrelevante Daten und Informationen zu erfassen, auszutauschen und einzusetzen. Sicherheitsinformationen über aktuell vorhandene sowie neu auftretende Gefahren werden

unter den Begriff *Cyber Threat Intelligence (CTI)* zusammengefasst. Dabei kann CTI hinsichtlich ihrer Verwendung unterteilt werden (Tounsi and Rais 2018). Eine grundlegende Form stellt die *Technical CTI* dar. Diese beschreibt Informationen, welche direkt in den Sicherheitssystemen (z. B. als Blocklist der Firewall) einer Organisation verarbeitet werden. Bekannte Beispiele hierfür sind IP-Adressen von bösartigen Servern oder Signaturen von Malware, welche auch Indicators of Compromise (IoCs) genannt werden. Ein Sicherheitssystem kann diese IoCs nutzen, um Verbindungen zu bösartigen Servern zu blockieren oder anhand einer Malware-Signatur einen Download von Schadsoftware zu erkennen. Die *Tactical CTI* hingegen beschreibt Methoden und wiederkehrende Taktiken, welche Angreifer verwenden, um Systeme erfolgreich zu infiltrieren. Damit sind umfangreichere Modellierungen von komplexen Angriffen und Gegenmaßnahmen möglich. Eine umfassende und strukturierte Beschreibung von Angriffsverhalten und Indikatoren sowie Informationen über angegriffene Systeme wird auch als *Threat Report* bezeichnet. Diese Threat Reports werden von Analysten genutzt, um geeignete Abwehrmaßnahmen auszuwählen und auf das aktuelle Verhalten der Angreifer auszurichten. Threat Reports ähneln in gewisser Weise Dossiers, welche zum Austausch von relevanten Informationen verwendet werden. Zuletzt stellt *Strategic CTI* Sicherheitsinformationen auf hoher Abstraktionsebene dar, wie beispielsweise finanzielle Schäden durch bestimmte Cyberangriffe und Vorhersagen über zukünftige Entwicklungen. Diese Informationen rich-

CTI-Beispiele für die MOVEit Schwachstelle (CVE-2023-34362)

Technical CTI (IP-Adresse und Malware Hashwert): 91[.]222[.]174[.]95 und 0b3220b11698b1436d1d866ac07cc90018e59884e91a8cb71ef8924309f1e0e9 (SHA256)

Tactical CTI (Angriffsverhalten nach MITRE ATT&CK): Exploit Public-Facing Application (SQL Injection), Server Software Component: Web Shell

Strategic CTI (Angriffskampagne): CL0P Ransomware Gang, Januar/Februar und Mai/Juni Angriffswelle 2023, vermuteter Schaden 75–100 Mio USD

ten sich an Personen mit Leitungsfunktion und Führungsverantwortung, welche auf dieser Grundlage langfristige Entscheidungen bezüglich der Ausrichtung der organisationseigenen IT-Sicherheit treffen können. Zusammengefasst lässt sich über die verschiedenen Arten von Sicherheitsinformationen sagen, dass Technical CTI gerade für kurzfristige Zeiträume wertvoll ist, wohingegen Tactical CTI und Strategic CTI auch auf längere Sicht einen Mehrwert für die Wahrung der IT-Sicherheit bieten. Dies liegt daran, dass Angreifer sehr leicht die bössartige IP-Adresse ihres Malware-Servers umstellen können, eine Änderung von grundlegenden Angriffsmustern und Methoden aber wesentlich aufwändiger ist (Bianco. The Pyramid of Pain. 2014).

Der Austausch von Sicherheitsinformationen spielt eine wesentliche Rolle bei der Verteidigung

Für reibungslose und effiziente Prozesse setzen Organisationen eine Vielzahl an digitalen Informationssystemen ein. Alle diese Systeme müssen jedoch auch ausreichend gegen Cyberangriffe geschützt werden. Dabei ist es nötig, sämtliche Schwachstellen und Angriffsvektoren der Systeme zu kennen und bewältigen zu können, was im Allgemeinen einen großen Aufwand darstellt. Dahingegen reicht den Angreifern oftmals eine einzige Schwachstelle, um einen erfolgreichen Angriff gegen eine Organisation durchzuführen. Dieser grundlegende Nachteil gegenüber den Angreifern verschlimmert sich dahin-

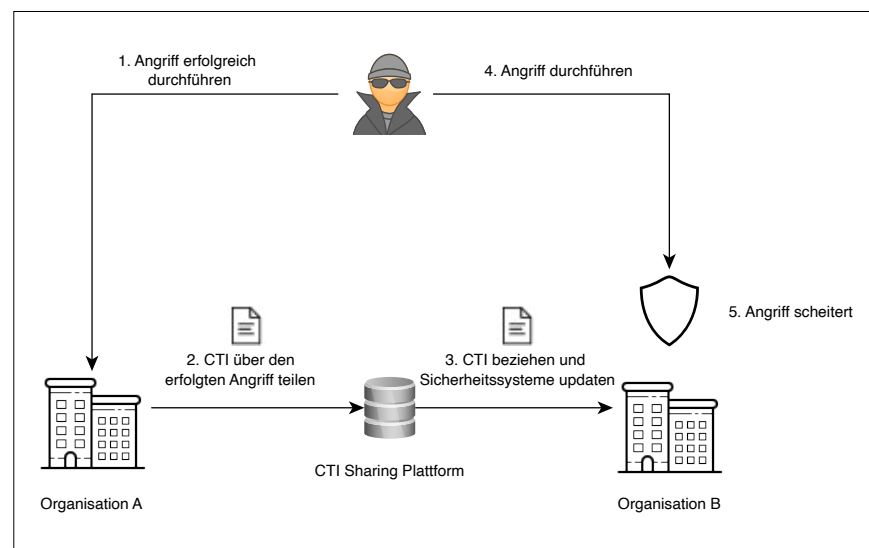
gehend, dass diese untereinander sehr gut vernetzt sind und arbeitsteilig agieren. In bestimmten Bereichen des Darknets können Cyberkriminelle aktuelle und hochspezialisierte Schadsoftware wie Ransomware gegen Bezahlung erwerben.

Vor diesem Hintergrund ist die Idee der kollaborativen Sicherheit vielversprechend. Die Cyberabwehr wird dafür nicht mehr ausschließlich siloartig (ausschließlich im Kontext einer Organisation), sondern gemeinschaftlich gedacht und durchgeführt. Kollaborative Cybersecurity beschreibt im Wesentlichen das Zusammenwirken von mehreren Akteuren (Unternehmen, Organisationen, Privatpersonen) mit dem gemeinsamen Ziel, die Sicherheit von Informationssystemen und Daten zu verbessern. Ein konkreter und vielversprechender Ansatz dafür ist der Austausch von CTI unter den verschiedenen Akteuren. Damit besteht die Möglichkeit, von vergangenen Erfahrungen und Erkenntnissen anderer zu lernen und seine eigene IT-Sicherheit anzupassen und zu verbessern. Neben einer rein internen Perspektive, bei der eine Organisation ihre internen Systeme und Abläufe kennt, wird somit eine zweite externe Perspektive im Rahmen der Cybersecurity hinzugenommen.

Für einen Austausch solcher Sicherheitsinformationen werden in der Praxis/Industrie verschiedene Kommunikationskanäle genutzt. Zum einen wird CTI informell in persönlichem Kontakt oder über E-Mail ausgetauscht. Ein vielversprechenderer Ansatz hingegen ist der systematische Austausch über dedizierte Plattformen. CTI sollte hierbei strukturiert erfasst und auf-

bereitet werden, um den Empfänger:innen der Sicherheitsinformationen einen wirklichen Mehrwert zu liefern. Dafür werden geeignete Datenformate bzw. Standards wie Structured Threat Information Expression (STIX) genutzt. Die Funktionsweise des Austausches über eine CTI-Plattform ist in der folgenden Abbildung beschrieben:

Ein bössartiger Akteur greift zunächst Organisation A an und nutzt beispielsweise ein fehlerhaft konfiguriertes System in der IT-Architektur von Organisation A aus. Organisation A muss nun den Angriff unter Kontrolle bringen und den Angreifer aus den internen Netzwerken der Organisation entfernen. Im Rahmen dieser Vorfallsbewältigung (Incident Response) wird auch analysiert, wie der Angreifer sich Zugriff auf die Infrastruktur verschaffen konnte. Die Erkenntnisse bereitet Organisation A als strukturierte CTI im STIX-Format auf und lädt die Daten anschließend auf einer Austauschplattform für CTI hoch. Organisation B, welche auch Teilnehmerin an dieser Plattform ist, bemerkt den neuen Eintrag. Da Organisation B eine ähnliche IT-Architektur wie Organisation A hat, kann der dokumentierte Angriff durch den bössartigen Akteur auch für Organisation B eine Gefahr darstellen. Daher bezieht Organisation B die CTI aus der Plattform und updatet seine Systeme entsprechend, um die bisherige Schwachstelle in der Konfiguration und somit den Angriffsvektor zu beseitigen. Startet der bössartige Akteur nun einen Angriff auf Organisation B, so scheitert dieser wegen der Anpassung basierend auf der zuvor bezogenen CTI.



1 Der systematische Austausch von CTI über Sharing Plattformen ermöglicht kollaborative Sicherheit und kann Angriffe verhindern.



Der Umgang mit Herausforderungen beschäftigt Forschung und Industrie.

Der Austausch von CTI mittels CTI Sharing Plattformen bietet Vorteile. Sowohl präventiv als auch reaktiv kann CTI zu einer Verbesserung der Cybersecurity einzelner Organisationen beitragen. Nichtsdestotrotz bestehen Herausforderungen im Austausch von CTI. Dem Austausch vorgelagert ist die Erfassung und Strukturierung von CTI. Auf Seite der teilenden Organisation sind dafür Ressourcen in Form von Personal und Zeit nötig. Ohne strukturierte CTI jedoch ist ein Austausch zwischen verschiedenen Akteuren schlicht nicht möglich oder liefert nur begrenzten Mehrwert wegen deutlicher Auswirkungen auf die Datenqualität. Es lässt sich bereits vermehrt feststellen, dass Organisationen bereit sind personelle Ressourcen für CTI zu Verfügung zu stellen. Auf kleine und mittlere Unternehmen trifft dies jedoch nur bedingt zu. Auch die Thematik der Datenqualität von CTI ist ein Thema der aktuellen Forschung und wurde teilweise in CTI Sharing Plattformen integriert. So können basierend auf semi-strukturierten Daten verschiedene Datenqualitätsmetriken angewendet werden und damit eine Einordnung der zugrundeliegenden CTI liefern. Die Thematik der Datenqualität ist

Bestandteil eines aktuellen Forschungsprojekts am Lehrstuhl. Genauer entsteht im Rahmen des DEVISE Projekts ein Reifegradmodell zur Messung und Verbesserung der Datenqualität von CTI.

Es stellen sich noch weitere organisatorische Fragen hinsichtlich des Austausches, dem Zusammenwirken und der Verwendung von CTI. Hierzu zählen: Welcher Akteur ist vertrauenswürdig? Welche Informationen sind relevant? Wie können Informationen in einem anderen Kontext verwendet werden? Neben der Ressourcenproblematik ist dabei der Reputationsverlust eine Herausforderung beim Austausch von CTI, da das Teilen der CTI einen (erfolgreichen) Angriff impliziert. Mittels Pseudonymisierung/Anonymisierung lässt sich hier nur teilweise Abhilfe schaffen, da Kontextinformationen häufig den Mehrwert von CTI deutlich verbessern. Erkenntnisse über die ideale Konstellation von Sharing Communities liegen außerdem nur begrenzt vor. Auch im Hinblick auf neuartige Technologien (insbesondere Blockchain zum Austausch und generative KI zur Strukturierung von CTI) ergeben sich weitere Forschungsfragen, deren Beantwortung den Austausch von CTI in der Praxis unterstützen kann.

Aktuelle Forschung am Lehrstuhl adressiert den Austausch von CTI mittels Blockchain. Im Rahmen eines weiteren For-

schungsprojekts (BMBF DEFENSIVE) wird um die Technologie herum ein Sharing System entwickelt. Dabei werden wesentliche Anforderungen von Teilnehmer:innen erfasst. Auch dienen Erkenntnisse aus langjährigen Industriekooperationen dazu, relevante Ergebnisse für aktuelle Herausforderungen zu erzielen.

Kollaborative Cybersecurity basierend auf dem Austausch von CTI verspricht einen schnelleren und treffsicheren Umgang mit Bedrohungen. Die Entwicklungen der letzten Jahre machen deutlich, dass dieser Ansatz vielversprechend ist und die Bereitschaft zum Austausch besteht. Mittels aktueller Forschung können bestehende Herausforderungen adressiert und die nächsten Schritte hin zu mehr Cybersecurity unternommen werden.

Quellen

- Verizon. Data Breach Investigations Report (DBIR) (2023).
- Hiscox. Cyber Readiness Report (2023).
- BSI. Die Lage der IT-Sicherheit in Deutschland (2022).
- Tounsi, W., and Rais, H. A survey on technical threat intelligence in the age of sophisticated cyber attacks. *Comput. Secur.* 72 (2018), 212–233.
- Bianco, D. The Pyramid of Pain (2014).



Foto © Johannes Grill

Johannes Grill, M. Sc., studierte Wirtschaftsinformatik an der Universität Regensburg. Seit 2023 promoviert er am Lehrstuhl Wirtschaftsinformatik I (Prof. Dr. Pernul) und ist an dem BMBF-Forschungsprojekt DEFENSIVE beteiligt. Zu seinen Forschungsschwerpunkten zählen Cyber Threat Intelligence (CTI) sowie Datentreuhändersysteme, über die CTI Daten zur Beschreibung von Sicherheitsvorfällen vertrauenswürdig ausgetauscht werden können.



Foto © Daniel Schlette

Dr. Daniel Schlette hat seinen Bachelor und Master in Wirtschaftsinformatik an der Universität Regensburg im Rahmen der Honors Elitestudiengänge des Elitenetzwerks Bayern erlangt. In seiner Promotion am Lehrstuhl Wirtschaftsinformatik I (Prof. Dr. Pernul) beschäftigte er sich mit Cyber Threat Intelligence und Incident Response. Er war an Forschungsprojekten mit Siemens und dem BMBF-Projekt DEVISE beteiligt.



Foto © Günther Pernul

Prof. Dr. Günther Pernul studierte Betriebs- und Wirtschaftsinformatik an der Universität Wien/TU Wien und schloss daran seine Promotion und Habilitation für Informatik an. Neben längeren Forschungsaufenthalten an der University of Florida und am Georgia Institute of Technology, USA war er Lehrstuhlinhaber an der Universität Duisburg-Essen. Seit 2002 hat er den Lehrstuhl für Wirtschaftsinformatik I an der Universität Regensburg und forscht schwerpunktmäßig im Bereich IT-Sicherheit.

Kann man den Entscheidungen Künstlicher Intelligenz trauen?

Zu den Auswirkungen unsicherer Daten auf die Entscheidungen Neuronaler Netze

Thomas Krapf, Prof. Dr. Bernd Heinrich

Im Laufe der letzten Jahre haben sich maschinelle Lernverfahren stetig weiterentwickelt und sind zunehmend in der Lage, immer anspruchsvollere und komplexere, reale Aufgaben zu bewältigen. So basieren beispielsweise neue Verfahren beim automatisierten Fahren, in Industrieprozessen, in der medizinischen Diagnostik oder in der maschinellen Textverarbeitung wesentlich auf Neuronalen Netzen. Ein geläufiges Beispiel mag hier das Sprachmodell GPT und darauf basierend ChatGPT sein, das mit interaktiv generierten Texten von menschen-ähnlicher Qualität auf sich aufmerksam machte. Insgesamt sind Neuronale Netze nicht nur ein zentrales Thema für Forschung und Industrie geworden, sondern erhalten zunehmend Einzug in unseren persönlichen Alltag.

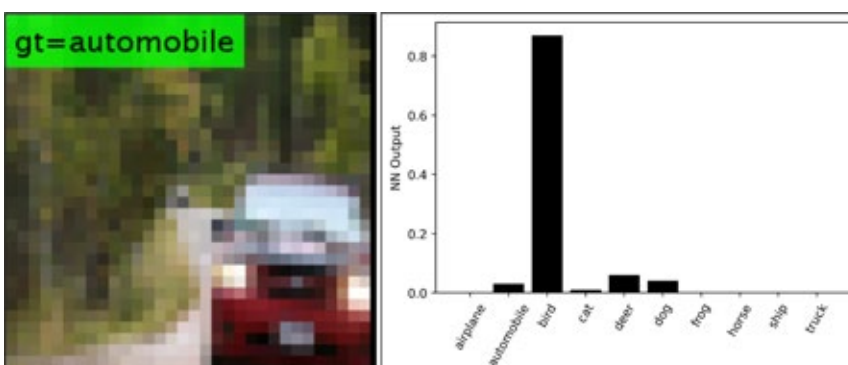
Soll ein Neuronales Netz für eine Klassifikationsaufgabe trainiert werden, so lernt es auf Basis einer größeren Datenmenge Muster und Zusammenhänge zwischen den Attributwerten der Dateninstanzen und dem Auftreten der jeweiligen tat-

sächlichen Klasse. Die für das Training und die Nutzung maschineller Lernverfahren verwendeten Daten liegen in der Realität jedoch oft mit Datenqualitätsdefekten vor. Beispielsweise sind Daten durch Mess-, Übertragungs- oder Erfassungsfehler inkorrekt, sie veralten im Zeitverlauf, sie sind unvollständig oder inkonsistent. Schlechte Datenqualität kann sich auf die Güte der Klassifikation oder Regression maschineller Lernverfahren stark auswirken (»garbage in, garbage out«), wie in experimentellen Analysen bereits mehrfach gezeigt wurde (z. B. Budach et al. 2022). Die Folge sind inkorrekte und wenig robuste Entscheidungen der Lernverfahren. Eine Illustration dieser Problematik aus dem Kontext der Bilderkennung (z. B. beim automatisierten Fahren) wird in Abbildung 1 dargestellt, welche die Klassifikation eines Objektes auf der unscharfen Aufnahme einer Fahrzeugkamera durch ein klassisches Neuronales Netz veranschaulicht.

Die Dateninstanz (hier die Kameraaufnahme) ist aufgrund schlechter Licht- oder

Wetterverhältnisse verpixelt und damit von schlechter Datenqualität. Dadurch klassifiziert das Neuronale Netz (z. B. LeNet) das Objekt in der Aufnahme (eindeutig) als »Vogel« anstatt als »Automobil«. Diese Prognose beeinflusst unmittelbar die nächsten Aktionen des Fahrzeugs und kann damit schwere Unfälle mit Personenschäden zur Folge haben (vgl. auch den tödlichen Unfall mit einem Tesla-Fahrzeug (NHTSA 2017) und die mangelhafte Bilddatenqualität). Die Problematik schlechter Datenqualität und ihr Einfluss auf die Güte und Robustheit von Entscheidungen Neuronaler Netze betrifft jedoch nicht nur Bilddaten, sondern auch Daten anderer Form. So sind beispielsweise in strukturierten Daten (wie medizinischen Patientendaten oder Sensordaten in Industrieprozessen) sehr oft fehlende oder inkorrekte Werte enthalten.

Um schlechte Datenqualität strukturiert zu berücksichtigen, kann die resultierende (sogenannte aleatorische) Unsicherheit über die wahre (aber unbekannt) Ausprägung eines Wertes modelliert und in den Lernprozess oder die Entscheidungsfindung einbezogen werden. Diese Modellierung erfolgt durch Wahrscheinlichkeitsverteilungen, welche Informationen über mögliche Ausprägungen des unsicheren Wertes sowie über deren jeweiligen Wahrscheinlichkeiten enthält. Klassische Neuronale Netze können allerdings mit wahrscheinlichkeitsbasierten Dateninstanzen nicht umgehen. Sie verarbeiten sichere Werte im Input zu einer punktweisen Entscheidung (z. B. für eine Klasse), meist basierend auf Werten einer Softmax-Funktion. D. h. für jede Klasse wird ein Wert ausgegeben (vgl. Y-

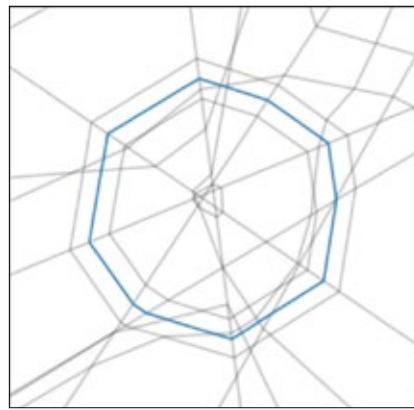


1 Prognose eines Neuronalen Netzes für ein unscharfes Bild (in Anlehnung an Gast & Roth 2018)

Achse in Abbildung 1) und die Klasse mit dem höchsten Softmax-Wert wird als Entscheidung gewählt. Die Softmax-Funktion neigt jedoch bekannterweise zur Überkonfidenz (vgl. auch Abbildung 1). Im Gegensatz dazu erzeugt die Berücksichtigung von datenqualitäts-induzierter Unsicherheit in den Inputdaten eines Neuronalen Netzes ebenso eine Wahrscheinlichkeitsverteilung im Outputraum des Netzes. Mit Hilfe dieser können für die jeweiligen Klassen nunmehr fundierte Wahrscheinlichkeiten ausgegeben werden. Dies hat Vorteile: Wird vereinfacht ein Neuronales Netz z. B. zur Unterstützung der Diagnose einer Krankheit herangezogen, so wäre es bei einer Verteilung von 40% für die Klasse »Patient besitzt Krankheit« und 60% für die Klasse »Patient besitzt Krankheit nicht« dringend geboten, nicht einfach die Klasse mit der höheren Wahrscheinlichkeit als korrekt anzunehmen. Vielmehr sind aufgrund der nicht unerheblichen Wahrscheinlichkeit von 40% einer Krankheit weitere Untersuchungen anzuraten. Dies zeigt insbesondere auch die Wichtigkeit einer fundierten Konfidenz von Prognosen, vor allem in risikoreichen Bereichen wie der medizinischen Diagnostik oder dem automatisierten Fahren, in denen die Gesundheit von Personen betroffen ist. Daher sind Methoden nötig, um die Unsicherheit in den Inputdaten eines Neuronalen Netzes möglichst exakt auf dessen Outputraum abzubilden.

Die exakte Propagation von Unsicherheit in Inputdaten, repräsentiert mittels Wahrscheinlichkeitsdichten durch ein Neuronales Netz, welches mathematisch eine Verkettung von linearen Transformationen und nicht-linearen Aktivierungsfunktionen ist, bringt einige methodische Herausforderungen mit sich. Für probabilistischen oder teilweise probabilistischen Input ist eine solche Propagation ungleich schwieriger und sogar analytisch nicht möglich, da für die Abbildung beliebiger Dichten durch nicht-lineare Funktionen im Allgemeinen keine geschlossene Lösung existiert. Daher müssen andere Methoden entwickelt werden, um Wahrscheinlichkeitsdichten exakt durch Neuronale Netze zu propagieren.

Ein Ansatz mit der komplexen, nicht-linearen Struktur von Neuronalen Netzen umzugehen, ist eine alternative Betrachtungsweise des Netzes als mathematische Funktion. So kann ein Neuronales Netz, welches grundsätzlich als iterative Verkettung von einzelnen Schichten (jeweils bestehend aus linearen Transformationen und

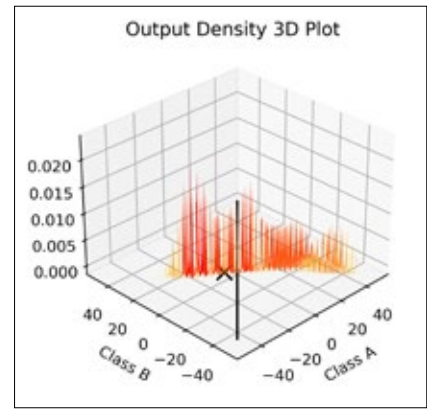


2 Lineare Regionen im Inputraum eines Neuronalen Netzes (schematische Abbildung, Sattlerberg et al. 2020)

nicht-linearen Aktivierungsfunktionen) definiert ist, auch als eine stückweise lineare Funktion auf dessen Inputraum aufgefasst werden (vgl. Sattlerberg et al. 2020). Dies bedeutet, dass das Netz auf polytop-förmigen Teilmengen des Inputraums durch eine (affin) lineare Abbildung gegeben ist. Diese sog. linearen Regionen teilen somit den Inputraum disjunkt auf, wie in Abbildung 2 schematisch zu sehen ist.

Diese Repräsentation des Netzes ist in unserer Methode für eine exakte Propagation nutzbar, indem zunächst die linearen Regionen des Netzes identifiziert werden und darauf basierend die Inputdichte (in Teilmassen) zerlegt wird. Die linearen Regionen und die darin enthaltenen Teilmassen der Inputdichte können nun exakt durch das Netz propagiert werden, da auf jeder linearen Region das Netz durch eine lineare Abbildung gegeben ist. Da sich die propagierten linearen Regionen im Output überschneiden können, müssen zuletzt noch deren Schnittmengen identifiziert und die propagierten Teilmassen der Inputdichte geeignet aggregiert werden. Als Ergebnis unserer Methode ergibt sich damit eine exakt propagierte Dichte im Outputraum des Netzes, welche als fundierte Basis zur Einschätzung der Unsicherheit einer Entscheidung des Neuronalen Netzes dient.

Neben der exakten Propagation von Wahrscheinlichkeitsdichten durch ein gegebenes Netz kann in diesem Kontext auch die Frage nach dem Training eines Neuronalen Netzes bei unsicheren Daten untersucht werden. Auch diese Fragestellung umfasst mehrere methodische Herausforderungen, insbesondere, dass sich Dichten oft über mehrere Bereiche des Inputraums erstrecken, die mit verschiedenen, auch inkorrekten Klassen im Output



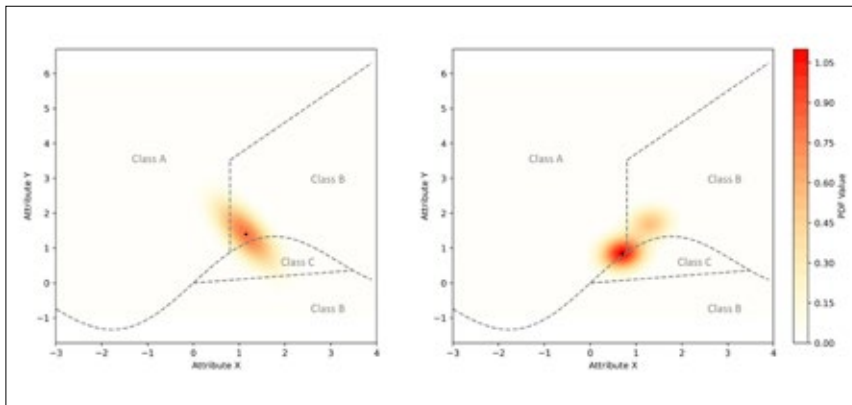
3 Dichte der Output-Verteilung

korrespondieren. Um zu verhindern, dass dadurch falsche Zusammenhänge erlernt werden, kann auch hier eine Aufteilung des Inputraumes und des Netzes in sogenannte Mixture of Experts-Netze, welche ausschließlich auf Teilmengen des Inputraums trainiert werden, hilfreich sein.

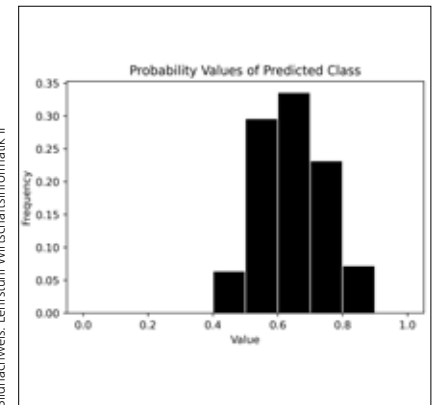
Aufgrund der exakten Propagation unsicherer Inputdaten mittels unserer Methode ergibt sich eine Reihe von Vorteilen, die hier auszugsweise diskutiert werden. Zur Illustration wird ein einfach gehaltenes Klassifikationsproblem zugrunde gelegt, das basierend auf nur zwei stetigen Datenattributen X und Y eine Zuordnung von Dateninstanzen durch ein Neuronales Netz in drei Klassen A, B und C betrachtet (z. B. die Klassifikation einer Patientin in »Hat Diabetesrisiko«, »Hat kein Diabetesrisiko« und »Hat Diabetes«, auf Basis der Attribute »Alter« und »Körpergewicht in Kg«).

1) Transparenz über die tatsächlichen Klassenwahrscheinlichkeiten: Die durch die exakte Propagation erhaltene Dichte im Output des Netzes ermöglicht es, die Wahrscheinlichkeiten für alle Klassen fundiert abzuleiten. In Abbildung 3 ist die propagierte Dichte für eine einzelne Dateninstanz unseres einfachen Klassifikationsproblems dargestellt. Die Dichte erstreckt sich über die Bereiche der beiden Klassen A und B im Outputraum (die Klasse C hat für die Instanz eine Wahrscheinlichkeit von 0 und bleibt daher aus Darstellungsgründen unberücksichtigt). Zudem ist die Entscheidungsgrenze (schwarze Linie) eingezeichnet, welche diese Bereiche trennt. Es ist zu erkennen, dass der Peak der Verteilung zwar in Klasse B ist, jedoch die größere, dichtere Masse den Bereich von Klasse A einnimmt. Folglich besitzt die Klasse B ca. 38% Wahrscheinlichkeit, die Klasse A ca.

Bildnachweis: Lehrstuhl Wirtschaftsinformatik II



4 Vergleich zweier Dateninstanzen mit identischen Klassenwahrscheinlichkeiten



5 Wahrscheinlichkeiten für die Entscheidungen der Testinstanzen

62 %. Die Entscheidung muss demnach offensichtlich als unsicher eingestuft werden. Diese Unsicherheit bliebe durch klassische, nicht probabilistische Analysen verborgen bzw. es können sogar falsche Entscheidungen resultieren: Würde nämlich – wie üblich – nur ein als sicher angenommener einzelner Wert (vgl. in Abbildung 3 z. B. den Modalwert als schwarzes Kreuz) als Input für das Netz verwendet werden, so würde dieser zur falschen Entscheidung »Klasse B« führen (die Instanz ist tatsächlich Klasse A). Erschwerend signalisiert das klassische Netz eine eindeutige Entscheidung, da die sich ergebenden Softmax-Werte der Klassen (0, 1, 0) lauten. D. h. es wird eine hohe Konfidenz angezeigt. Da keine weitere Information über die Unsicherheit dieser Entscheidung vorliegt, würde infolgedessen z. B. das tatsächliche Diabetesrisiko der Patientin nach dieser Analyse unerkannt bleiben.

2) Lage der Wahrscheinlichkeitsdichte und robuste Entscheidungen: Neben den reinen Klassenwahrscheinlichkeiten können mehrere inhärente Eigenschaften der Dichte, wie die generelle Verteilung der Wahrscheinlichkeitsmasse, deren Streuung oder deren Entfernung zu Entscheidungsgrenzen, analysiert werden. Die Abbildung 4 zeigt dies für zwei weitere Dateninstanzen unseres einfachen Klassifikationsproblems. Da wir nun den Inputraum betrachten, sind an den Achsen nicht mehr die Klassen (wie in Abbildung 3) abgetragen, sondern die Datenattribute X und Y, deren Unsicherheit im Inputraum sowie die Entscheidungsgrenzen zwischen den Klassen A, B, und C (grau gestrichelte Linien). Beide Dateninstanzen weisen gerundet mit (0.28, 0.40, 0.32) identische Klassenwahrscheinlichkeiten auf und wären damit nach Punkt 1) als vergleichbar zu interpretieren.

Allerdings liegen die Wahrscheinlichkeitsdichten der beiden Dateninstanzen sehr unterschiedlich im Raum. Während die gesamte Wahrscheinlichkeitsmasse der linken Dateninstanz sehr nahe bei den oder sogar auf den Entscheidungsgrenzen zwischen den drei Klassen liegt, befindet sich ein größerer Teil der Wahrscheinlichkeitsmasse der rechten Dateninstanz tiefer im Bereich der Klasse B. Eine Erhöhung oder Reduzierung der Werte der beiden Datenattribute in diesem Bereich führen demnach zu keiner Änderung der Klassenentscheidung. Die rechte Instanz weist dahingehend eine höhere Robustheit auf als die linke Dateninstanz. Robustheit stellt in diesem Zusammenhang eine zentrale Güteeigenschaft Neuronaler Netze dar. Daneben wird eine weitere potenzielle Problematik evident: Klassische Neuronale Netze entscheiden i. d. R. nach dem Prinzip der relativen Mehrheit, d. h. die Klasse mit dem höchsten Softmax-Wert wird gewählt (s. o.). Hier sind die berechneten Softmax-Werte auf Basis der Modalwerte der beiden Dateninstanzen mit gerundet (0, 1, 0) eindeutig. Dies signalisiert eine hohe Robustheit der Entscheidung, welche sowohl durch die tatsächlichen Klassenwahrscheinlichkeiten der beiden Instanzen von (0.28, 0.40, 0.32) als auch durch die Lage der Wahrscheinlichkeitsdichte nicht gerechtfertigt ist. Allein schon durch die Betrachtung der Klassenwahrscheinlichkeiten wird klar, dass die beiden nicht-gewählten Klassen einen Großteil der Wahrscheinlichkeitsmasse (ca. 60%) ausmachen, d. h. ihr Eintreten ist insgesamt wahrscheinlicher als der von Klasse B.

3) Globale Robustheit der Entscheidungen eines Neuronalen Netzes: Die bisher auszugsweise vorgestellten Vorteile beziehen sich auf die Analyse von

Entscheidungen für einzelne Instanzen. Die Wahrscheinlichkeiten für die Klassen lassen sich aber auch auswerten, um die Entscheidungen des Neuronalen Netzes für alle Dateninstanzen (»global«) hinsichtlich ihrer Robustheit zu beurteilen. Ordnet man für alle Testinstanzen im betrachteten Klassifikationsproblem die Wahrscheinlichkeiten der getroffenen Entscheidungen (d. h., das Maximum der Klassenwahrscheinlichkeiten) Intervallen zu, so ergibt sich Abbildung 5. Hier sieht man, dass z. B. ca. ein Drittel aller Entscheidungen nur mit einer Klassenwahrscheinlichkeit knapp über 50 % gefällt wurden. Insgesamt ist die Robustheit der Entscheidungen des Neuronalen Netzes basierend auf der gegebenen hohen Datenunsicherheit (schlechte Datenqualität) wenig überzeugend. Hier werden Schwächen sichtbar, die infolge der reinen Betrachtung der Softmax-Werte (hier werden über 95 % aller Entscheidungen mit einem Softmax-Wert nahe eins getroffen) nicht transparent werden.

In den letzten Jahren hat die Frage nach der Robustheit und Nachvollziehbarkeit KI-basierter Prognosen stark an Bedeutung gewonnen. Da KI immer mehr auch bei risikoreichen Entscheidungen zur Unterstützung oder als Entscheidungsinstanz herangezogen wird, muss vor allem in diesen Einsatzgebieten neben der Prognosegüte auch ein Maß für die Robustheit berücksichtigt und in ein verantwortungsvolles Handeln einbezogen werden. Ein solches Maß ist Grundvoraussetzung, dass das Vertrauen in die Prognosen maschineller Lernverfahren wächst und damit ihr Potenzial voll ausgeschöpft werden kann. So können Personen (z. B. Ärzte) Neuronale Netze zur Ergänzung und Entlastung auch in kritischen Domänen unterstützend heranziehen, aber auch die Unsicherheit der

Entscheidung berücksichtigen, um individuell auf dieser Basis das weitere Vorgehen zu bestimmen.

Die grundsätzliche Bedeutung dieser Thematik für Individuen, Gesellschaft und Industrie wurde auch bereits erkannt. So berät die EU im Rahmen des »AI-Act«-Gesetzes eine gemeinsame Position und Strategie zur Regulierung des Einsatzes von KI-Technologien. Dieser Rechtsrahmen soll Regeln festlegen, um sicherzustellen, dass KI-Systeme transparent und verantwortungsvoll eingesetzt werden. Herausforderungen wie Datenqualität und maschinelles Lernen basierend auf unsicheren Daten

sind explizit genannt. Gleichzeitig fördert der AI-Act weitere innovative Entwicklungen, da Leitlinien für die Entwicklung und Nutzung von KI-Systemen bereitgestellt werden sollen, um eine positive Integration von KI in verschiedene Lebensbereiche grundsätzlich zu ermöglichen.

Referenzen

Budach, L.; Feuerpfeil, M.; Ihde, N.; Nathansen, A. Noack, N.; Patzlaff, H., Naumann, F.; and Harmouch, H. 2022. The Effects of Data Quality on Machine Learning Performance. arXiv preprint arXiv:2207.14529.

Gast, J. and Roth, S. 2018. Lightweight Probabilistic Deep Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*: 3369–3378. IEEE Computer Society.

NHTSA. 2017. PE 16-007. Tesla Crash Preliminary Evaluation Report. U.S. Department of Transportation, National Highway Traffic Safety Administration. Technical report.

Sattelberg, B.; Cavalieri, R.; Kirby, M.; Peterson, C.; and Beveridge, R. 2020. Locally Linear Attributes of ReLU Neural Networks. arXiv preprint arXiv:2012.01940.



Foto © privat

Prof. Dr. Bernd Heinrich

ist seit 2011 Inhaber des Lehrstuhls für Wirtschaftsinformatik II an der Universität Regensburg. Er promovierte an der Universität St. Gallen/Schweiz und habilitierte sich an der Universität Augsburg. Seine Forschungsschwerpunkte liegen in den Bereichen Data Science, Data Quality & Data Uncertainty, Process Planning & Process Science sowie Service Systems

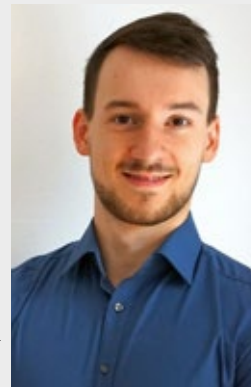


Foto © privat

Thomas Krapf, M.Sc.,

Thomas Krapf, geboren 1997 in Amberg. 2015–2019 Studium der Mathematik mit Nebenfach Betriebswirtschaftslehre und Aktuarwissenschaften an der Universität Regensburg mit Vertiefung in den Bereichen der partiellen Differentialgleichungen und angewandten Analysis. Seit Januar 2021 Tätigkeit als wissenschaftlicher Mitarbeiter und Doktorand am Lehrstuhl für Wirtschaftsinformatik II der Universität Regensburg.

Seine Forschungsschwerpunkte umfassen die Bereiche Data Quality und Data Uncertainty in Machine Learning.

Mensch vs. Maschine

Wettbewerb und Kooperation mit künstlicher Intelligenz in digitalen Märkten

Andreas Schauer, Prof. Dr. Daniel Schnurr

Algorithmische Preissetzung und künstliche Intelligenz in digitalen Märkten

In digitalen Märkten werden Unternehmensentscheidungen zunehmend automatisiert mit Hilfe algorithmischer Verfahren und künstlicher Intelligenz (KI) getroffen. Dies gilt beispielsweise heute schon für die Preissetzung auf elektronischen Online-Handelsplätzen. So verwenden zwei Drittel der Händler:innen auf der Plattform »Amazon Marketplace« automatisierte Preissetzungsverfahren oder Entscheidungsunterstützungssysteme, die Preise der Wettbewerber überwachen. Infolgedessen entstehen neue hybride Marktumgebungen, in denen menschliche und algorithmische Akteure miteinander konkurrieren.

In Folge des zunehmenden Einsatzes KI-basierter Verfahren haben internationale Organisationen wie die OECD sowie nationale Wettbewerbsbehörden Bedenken geäußert, dass Unternehmen Algorithmen und maschinelle Lernverfahren nutzen könnten, um sich implizit zu koordinieren und höhere Marktpreise zum Nachteil der Konsument:innen durchzusetzen. Als Beispiele werden häufig die automatisierte Preissetzung bei Flugtickets, der maschinelle Handel auf Finanzmarktplätzen und der Wettbewerb auf eCommerce-Plattformen angeführt.

Simulationsstudien zeigen zudem, dass autonome Computeragenten mit Hilfe von maschinellen Lernverfahren tatsächlich in der Lage sein können, sich strategisch auf Preise über dem Wettbewerbsniveau zu

koordinieren. Unklar ist jedoch, wie stark das Marktergebnis wirklich vom Wettbewerb zwischen menschlichen Entscheidern abweicht. Zudem ist bisher wenig erforscht, wie sich Wettbewerb in hybriden Marktumgebungen mit menschlichen und maschinellen Entscheidern gestaltet. Besonders praxisrelevant ist dabei die Frage, welchen Einfluss bereits automatisierte Entscheidungsunterstützungssysteme, wie sie beispielsweise auf der Handelsplattform eBay oder der Unterkunftsplattform Airbnb angeboten werden, auf den Wettbewerb haben können.

Analyse des Wettbewerbs zwischen menschlicher und künstlicher Intelligenz mittels Computersimulationen und ökonomischer Laborexperimente

Im Rahmen des Forschungsprojekts »Mensch vs. Maschine: Wettbewerb und Kooperation mit künstlicher Intelligenz in digitalen Märkten«, das durch das Dr. Theo und Friedl Schöller Forschungszentrum für Wirtschaft und Gesellschaft gefördert wird, erforscht der Lehrstuhl für Maschinelles Lernen, insbesondere Uncertainty Quantification die Wettbewerbsdynamik zwischen menschlicher und künstlicher Intelligenz und untersucht, welche ökonomischen Auswirkungen der vermehrte Einsatz von maschinellen Lernen in digitalen Märkten mit sich bringt. Mit Hilfe von Computersimulationen und einer Serie von ökonomischen Laborex-

perimenten wird analysiert, welchen Einfluss Algorithmen auf das Marktergebnis haben, wenn Preise automatisiert von maschinellen Lernverfahren bestimmt werden. Durch die kontrollierte Variation (i) der handelnden Akteure und (ii) des Grades KI-basierter Entscheidungsunterstützung für menschliche Entscheider werden Einflussfaktoren und Implikationen der Mensch-Maschine-Interaktion im strategischen Wettbewerb systematisch untersucht.

Als Referenzszenario für die Studie wurde eine digitale Marktumgebung entwickelt, die es sowohl menschlichen als auch KI-basierten Akteuren erlaubt, in Echtzeit Preise für Güter auf einer Handelsplattform zu setzen (siehe Abbildung 1). Die Kundennachfrage sowie die Unternehmensgewinne werden auf Grundlage der gesetzten Preise und einem parametrisierbaren Wettbewerbsmodell berechnet. Über die Dauer eines Szenarios erzielen die Unternehmen in Abhängigkeit von Preis und Nachfrage einen Gesamtertrag. Dieser stellt für menschliche Entscheider zudem die Gewinnauszahlung dar, die sie für die Teilnahme am Experiment erhalten. Um einem Hauptmerkmal digitaler Märkte gerecht zu werden, erfolgt die Wettbewerbsinteraktion in Echtzeit, das heißt Entscheider können zu jedem Zeitpunkt interagieren und auf Preise von Wettbewerbern reagieren.

Bereits im Vorfeld der experimentellen Studie wurden KI-basierte Agenten in Computersimulationen auf der gleichen Handelsplattform trainiert. Dabei wurde systematisch untersucht, welche KI-De-



sign-Parameter für das von den Agenten verwendete Reinforcement Learning zu den profitabelsten Ergebnissen führt. Aus 12.000 Simulationsläufen zu je 50 Millionen Trainingsperioden wurde die erfolgreichste Klasse von KI-basierten Agenten ermittelt, die im darauffolgenden Experiment sowohl als eigenständige Wettbewerber als auch für die Entscheidungsunterstützung von menschlichen Entscheidern eingesetzt wurde.

Im Rahmen des Laborexperiments wurde mit Blick auf die Auswirkungen der handelnden Akteure zunächst der Wettbewerb zwischen menschlichen Entscheidern untersucht und mit den Ergebnissen der Simulationen der KI-basierten Agenten verglichen. In der dritten Versuchsanordnung traten menschliche Akteure dann in den direkten Wettbewerb mit den KI-basierten Agenten.

In weiteren Experimentsitzungen wurde systematisch der Einfluss KI-basierter Entscheidungsunterstützung auf das strategische Verhalten von menschlichen Entscheidern sowie auf das Marktergebnis untersucht. Hierfür wurde menschlichen Entscheidern über die gesamte Experimentdauer eine dynamische Preisempfehlung präsentiert, die auf den gesetzten Preisen der Wettbewerber und dem trainierten KI-Modell basierte. Der Einfluss von Entscheidungsunterstützungssystemen (DSS) wurde sowohl im Wettbewerb zwischen menschlichen Entscheidern als auch im Wettbewerb zwischen Menschen und Algorithmen untersucht.

Ergebnisse: Der Einfluss KI-basierter Systeme auf Preise und Unternehmensgewinne

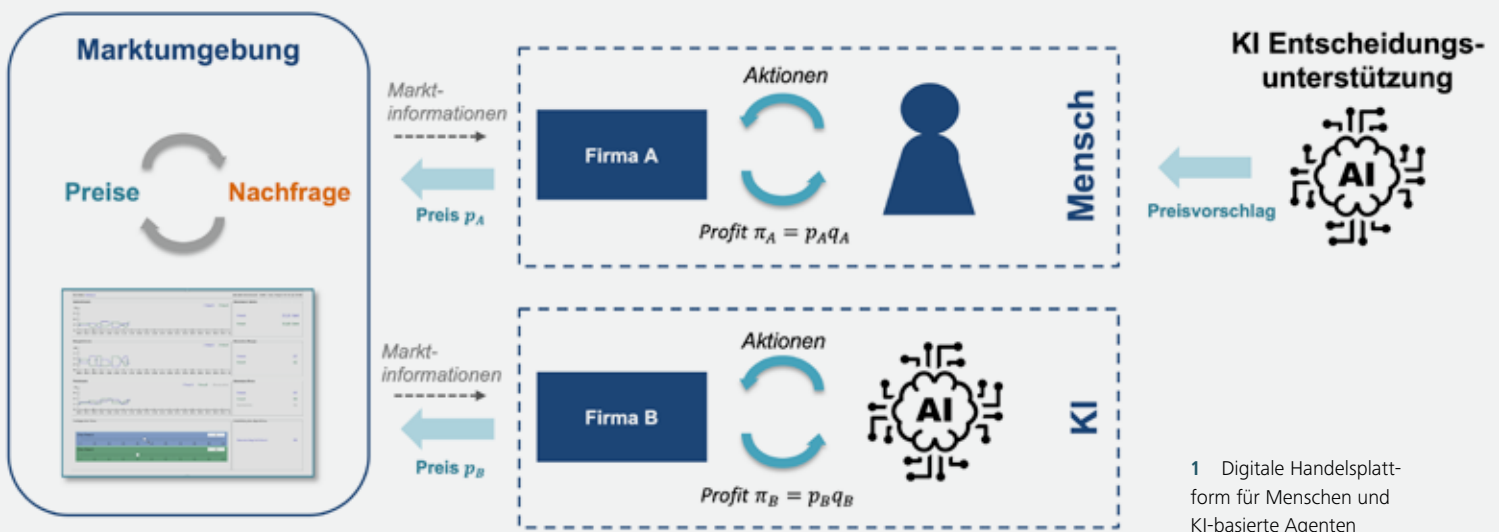
Für die Auswertung der Experimentergebnisse ist insbesondere das empirisch gemessene Preisniveau und dessen Relation zum Gleichgewichtspreis bedeutsam, welcher sich gemäß der spieltheoretischen Vorhersage bei effektivem Wettbewerb einstellen sollte. Wenn Akteure in der Lage sind, sich ohne explizite Absprachen auf Preise über dem Wettbewerbsniveau zu koordinieren, erfolgt dies zum unmittelbaren Schaden der Verbraucher:innen.

Über alle Versuchsanordnungen hinweg zeigt sich, dass das durchschnittliche Preisniveau über dem wettbewerblichen Gleichgewichtspreis liegt. Es wird somit deutlich, dass sowohl menschliche als auch KI-basierte Akteure in der Lage sind, sich ohne explizite Absprachen auf Preise über dem Wettbewerbsniveau zu koordinieren. Die höchsten Gewinne und Marktpreise werden dabei von KI-basierte Agenten erzielt, wenn diese im Wettbewerb mit anderen Agenten stehen. Stehen KI-basierte Agenten jedoch im Wettbewerb mit menschlichen Entscheidern (ohne weitere Entscheidungsunterstützung), so sinken die Marktpreise und Unternehmensgewinne signifikant. Insbesondere fallen Preise und Gewinne unter das durchschnittliche Niveau in Märkten, in denen nur menschliche Entscheider in Konkurrenz stehen.

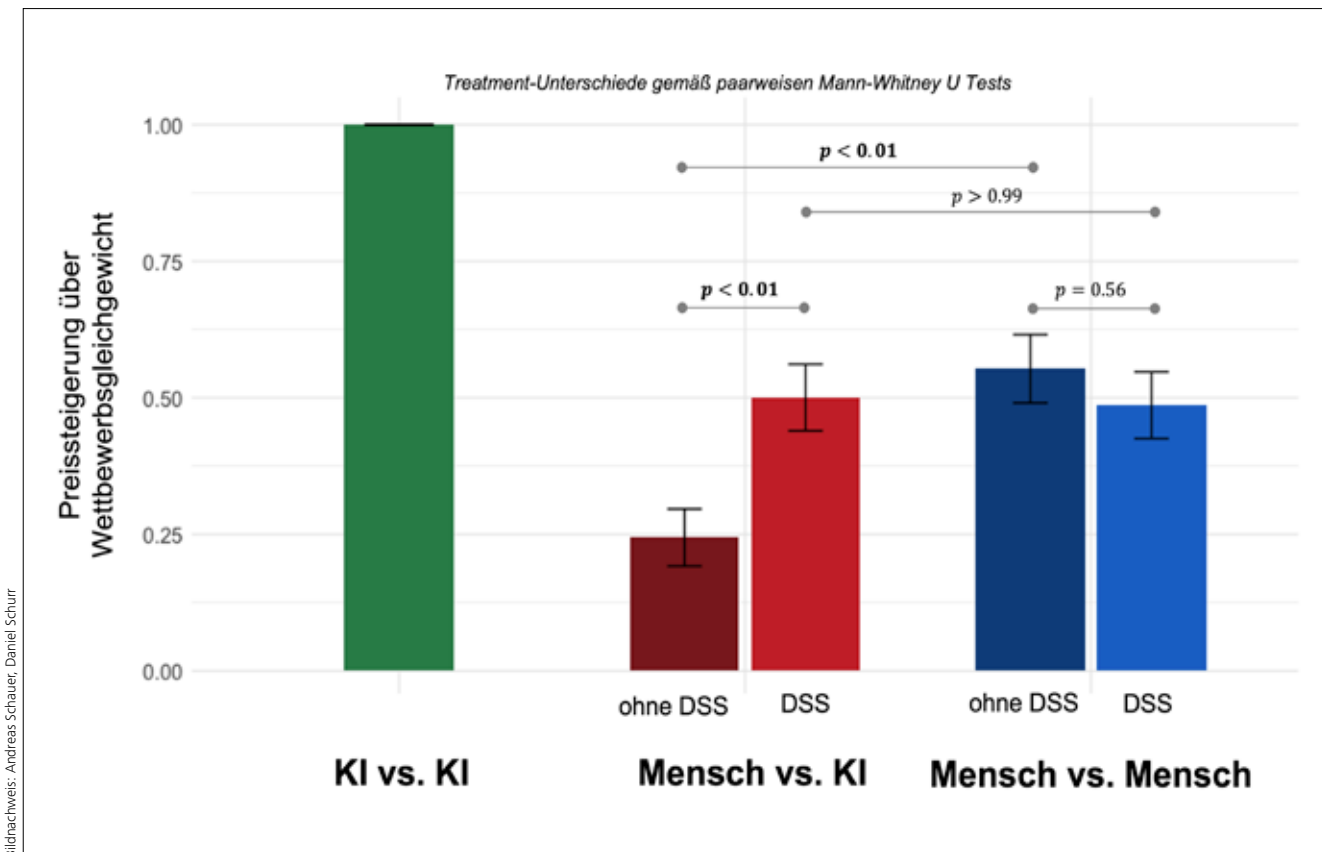
Es zeigt sich somit, dass in hybriden Marktumgebungen, in denen menschliche

und künstliche Intelligenz im Wettbewerb stehen, eine implizite Koordination auf hohe Preisniveaus schwieriger erscheint als in Marktumgebungen, in denen nur Menschen oder nur Algorithmen konkurrieren. Dieses Ergebnis ist aus theoretischer Sicht bedeutsam, da verwandte Studien, die den Wettbewerb zwischen Menschen und Algorithmen untersuchen, bisher einen monotonen Zusammenhang zwischen der Anzahl der algorithmischen Entscheider und dem Preisniveau bzw. der Unternehmensgewinne festgestellt haben. Die Ergebnisse des Forschungsprojekts zeigen nun, dass sich bisherige Erkenntnisse aus Studien zu regelbasierten, statischen Algorithmen nicht notwendigerweise auf dynamische KI-basierte Ansätze übertragen lassen. Während Menschen und KI-basierte Agenten im Wettbewerb untereinander vergleichsweise kooperativ agieren und hohe Preisniveaus erzielen, ergeben sich deutliche niedrigere durchschnittliche Preise und Unternehmensgewinne, wenn menschliche Entscheider und KI-basierte Agenten in direkter Konkurrenz stehen.

Auf Grundlage einer weiteren Versuchsanordnung, in der menschliche Teilnehmende nicht über die algorithmische Identität ihres Wettbewerbers informiert wurden, kann zudem gezeigt werden, dass dieses Marktergebnis durch eine geringere Fähigkeit zur Kooperation zwischen Menschen und KI und nicht allein durch eine geringere Bereitschaft zur Kooperation getrieben ist. Die Preisdaten geben zudem Hinweise darauf, dass menschliche



1 Digitale Handelsplattform für Menschen und KI-basierte Agenten



Bildnachweis: Andreas Schauer, Daniel Schurr

2 Durchschnittliches Preisniveau in den verschiedenen Versuchsanordnungen

che Entscheider und KI-basierte Agenten grundsätzlich unterschiedliche Strategien verfolgen, um Kooperation in digitalen Märkten zu erreichen.

Schließlich geben die Experimentergebnisse Aufschluss darüber, wie KI-basierte Entscheidungsunterstützung den Wettbewerb in digitalen Märkten beeinflussen kann. Stehen allein menschliche Entscheider im Wettbewerb miteinander, so verändert KI-basierte Entscheidungsunterstützung die Wettbewerbsdynamik und das Marktergebnis in den durchgeführten Experimenten nicht signifikant. In hybriden Marktumgebungen, in denen menschliche Entscheider mit KI-basierten Agenten im Wettbewerb stehen, führt die Verfügbarkeit von KI-basierter Entscheidungsunterstützung jedoch zu einem signifikanten Anstieg der Marktpreise und der Unternehmensgewinne. Im Ergebnis erreichen Preise und Gewinne somit das gleiche Niveau wie im Wettbewerb zwischen menschlichen Entscheidern. KI-basierte Entscheidungsunterstützungssysteme können somit die geringere Fähigkeit zur Kooperation zwischen Menschen und KI kompensieren, was in der Folge zu einer Abschwächung der Wettbewerbsintensität in digitalen Märkten führen kann.

Implikationen für Wirtschaft und Gesellschaft: Einsatz und Governance von KI in digitalen Märkten

Das Forschungsprojekt zeigt auf, welche wettbewerblichen Auswirkungen durch den zunehmenden Einsatz von KI-basierten Preissetzungsverfahren in digitalen Märkten zu erwarten sind. Die Ergebnisse zeigen zum einen, dass eine weitverbreitete Nutzung von KI-basierten Preissetzungs-Algorithmen die Wettbewerbsintensität in digitalen Märkten grundsätzlich reduzieren und somit zu höheren Preisen für Verbraucher:innen führen kann. Zum anderen machen die Ergebnisse jedoch deutlich, dass menschliche und künstliche Intelligenz häufig grundlegend unterschiedliche Wettbewerbs- und Kooperationsstrategien verfolgen und es daher gerade in hybriden Marktumgebungen mit menschlichen und KI-basierten Akteuren zu vergleichsweise wettbewerblichen Marktergebnissen kommen kann.

Die Forschungsergebnisse verdeutlichen zudem die bedeutsame Rolle von KI-basierten Entscheidungsunterstützungssystemen, die den Wettbewerb und die Preissetzung menschlicher Entscheider signifikant zum Nachteil von Verbraucher:innen beeinflus-

sen können. Für Unternehmen kann der Einsatz solcher Systeme hingegen zu signifikanten Gewinnsteigerungen führen. Zugleich sind damit aber auch Compliance-Risiken für Entscheidungsträger:innen in Unternehmen verbunden, da die Preisentscheidungen KI-basierter Systeme von Wettbewerbsbehörden als wettbewerbsfeindliches Verhalten im Verantwortungsbereich der einsetzenden Unternehmen gewertet werden könnten. Um diesen Risiken zu begegnen, sollten Unternehmen das Verhalten eigener KI-basierter Preissetzungs-Algorithmen in marktnahen Simulationsumgebungen testen und während des Einsatzes kontinuierlich überwachen.

Wettbewerbs- und Regulierungsbehörden sollten verstärkt eigene digitale Kompetenzen aufbauen und insbesondere vermehrt Markt- und Preisdaten sammeln, um den Einsatz von KI-basierten Preissetzungs-Algorithmen und deren Auswirkungen auf das Marktergebnis besser nachvollziehen zu können. Dies gilt umso mehr, da die experimentellen Ergebnisse zeigen, dass nicht notwendigerweise ein monotoner Zusammenhang zwischen der Anzahl der algorithmischen Entscheider und der Wettbewerbsintensität besteht. In Märkten in denen der Verdacht wett-

bewerbsfeindlichen Verhaltens durch »algorithmischer Kollusion« besteht, können Behörden auf Audits, Befragungen oder eigene Preissetzungsalgorithmen zurückgreifen, um genauer zu analysieren, in welchem Maße die eingesetzten KI-basierten Preissetzungsalgorithmen oder Entscheidungsunterstützungssysteme

den Preisbildungsprozess und die Wettbewerbsdynamik beeinflussen. Präventiv kann dies durch regulatorische Maßnahmen zur Erhöhung der Transparenz bezüglich der Nutzung von KI-basierten Algorithmen in Unternehmen und den zugrundeliegenden Datenquellen gestärkt werden.

WIR FÜHLEN

DICH WOHL.

LEADING
EMPLOYER
2023

#KörperGeistSeeleArbeit



Stellenangebote

Ober- / Fach- / Assistenzarzt (m/w/d)

Psychologe (m/w/d)

Psychologischer Psychotherapeut (m/w/d)

In den Heiligenfeld Kliniken arbeiten wir mit Körper, Geist und ganz viel Seele. Was für unsere Patientinnen und Patienten heilend ist, kann für uns nur förderlich sein. Bewirb dich in unserer psychosomatischen [Familienklinik in Waldmünchen](#) und lerne unser einzigartiges, ganzheitliches Therapiekonzept kennen.



Heiligenfeld
Kliniken

Infos und Bewerbung unter:
karriere.heiligenfeld.de

Kontakt: Tanja Meier
09972 308-800





Foto © privat

Andreas Schauer, M. Sc.,
ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Maschinelles Lernen, insbesondere Uncertainty Quantification an der Universität Regensburg. Zuvor war er von Juni 2021 bis August 2022 in der Forschungsgruppe Data Policies als wissenschaftlicher Mitarbeiter tätig. Er besitzt einen Master of Science in Wirtschaftsinformatik und einen Master of Science in Business Administration sowie einen Bachelor of Science in Wirtschaftsinformatik von der Universität Passau.



Foto © privat

Prof. Dr. Daniel Schnurr
ist seit August 2022 Inhaber des Lehrstuhls für Maschinelles Lernen, insbesondere Uncertainty Quantification an der Universität Regensburg. Zuvor leitete er die Forschungsgruppe Data Policies an der Universität Passau. Daniel Schnurr promovierte 2016 im Bereich Wirtschaftsinformatik am Karlsruher Institut für Technologie, wo er zuvor Informationswirtschaft (B.Sc. & M.Sc.) studierte. Daniel Schnurr ist Research Fellow des Centre on Regulation in Europe (CERRE).

Blick in die Wissenschaft 44/45 ■ 33

Notfallpläne für den Ernstfall testen

Prof. Dr. Maria Leitner

Die fortschreitende Digitalisierung und die Leistungssteigerung der digitalen Infrastrukturen der letzten Jahre hat die Informationssicherheit als entscheidendes Thema der Resilienz immer mehr in den Mittelpunkt aller digitalen Systeme, Anwendungen und in Organisationen gerückt. Eine Vielzahl an öffentlichen Beispielen von Cyberangriffen (z. B. Erpressungsversuche bei großen Verkehrsunternehmen, Fahrzeugherstellern, Stadtverwaltungen oder auch Universitäten) zeigt, dass diese im Alltag von allen Organisationen gleich ob in Wirtschaft, Wissenschaft oder auch in der Verwaltung angekommen sind.

Die immer stärkere Nutzung von digitalen Anwendungen in der Gesellschaft (z. B. mit personenbezogenen Daten) benötigt die Umsetzung höherer Schutzniveaus. Diese Anwendungen eröffnen oftmals potenziellen Angreifern eine größere Angriffsfläche, da sie lohnende Ziele für diese darstellen. Beispielsweise gab es noch vor zwanzig Jahren sehr wenige Geräte im Haushalt (Stichwort *Smart Home*), die mit dem Internet verbunden waren. Mittlerweile gibt es eine Vielzahl an Anwendungen und Geräten (z. B. das Handy, der Kühlschrank oder die Hausklingel mit Video), die konstant mit dem Internet verbunden sind. Die Digitalisierung hat ebenso die Industrie maßgeblich verändert. Ehemals isolierte Systeme haben nun oftmals Verbindungen in lokale Netze oder das Internet. Eine derartige Verbindung könnte zum Beispiel auch von Angreifenden ausgenutzt werden. Auch wenn die kontinuierliche Weiterentwicklung und Automatisierung von Technologien der Gesellschaft sehr viel Nutzen gebracht hat, so ist es jedoch auch wichtig, zusätzlich zu den technologischen Entwicklungen, die Sicherheit und Privat-

sphäre kontinuierlich bei diesen Transformationsprozessen mitzubetrachten.

Denn gleichzeitig haben sich auch die Angreifer weiterentwickelt und nutzen die Automatisierung immer öfter zu ihren Zwecken. Heutzutage ist es einfach, mit nur wenigen Klicks und vorgefertigten Anwendungen, gezielt eine Vielzahl von Computersystemen gleichzeitig anzugreifen. Durch dieses sich kontinuierlich weiterentwickelnde Angreiferfeld, ist es essenziell auch Techniken und Technologien weiterzuentwickeln, um rechtzeitig und systematisch Angriffe zu erkennen und das schadhafte Verhalten schnellstmöglich einzudämmen.

Eine zielführende Möglichkeit sich auf Cyberangriffe vorzubereiten sind Cybersicherheitsübungen. Übungen werden bereits seit Jahrzehnten in verschiedensten Bereichen zur nachhaltigen Vorbereitung genutzt. Beispielsweise werden bei der Feuerwehr oder Rettung regelmäßige Brand- oder Rettungsübungen durchgeführt. Übungen können auch im Bereich der Cybersicherheit zur Anwendung gebracht und zur Vorbereitung genutzt werden. In herausfordernden Übungssituationen (zum Beispiel bei einer simulierten Erpressung oder einem Datenklau) werden die Notfallpläne (engl. *Incident response plans*) einem Stress- und Realitätstest unterzogen. Dadurch können sich Organisationen spielerisch auf den Ernstfall vorbereiten und schneller reagieren.

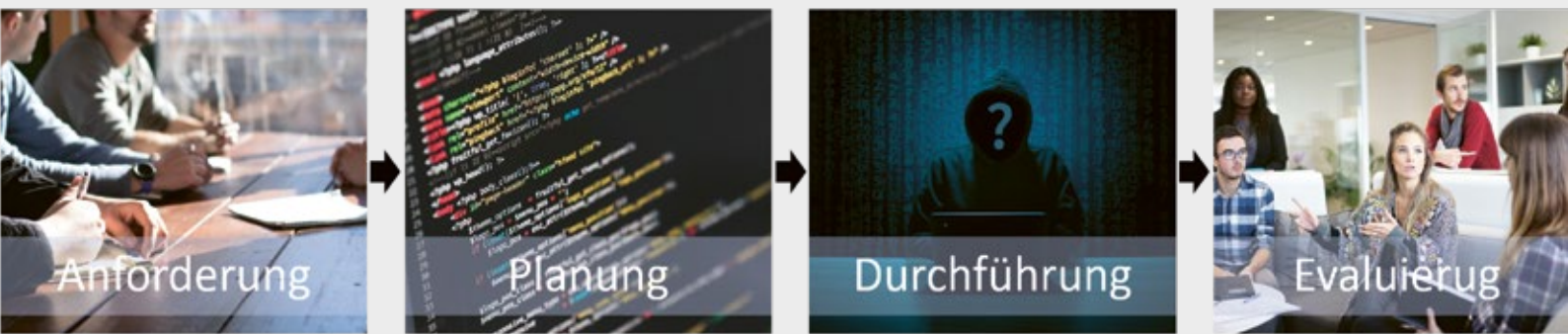
In der Forschung beschäftigen wir uns daher nicht nur damit wie konkrete zukunftsweisende Übungsumgebungen (z. B. welche Technologien) simuliert und umgesetzt werden können, sondern auch welche Angriffe und Techniken im Cybersicherheitsbereich wichtig sind [1]. Die Vorbereitung einer Übung kann in mehrere

Phasen eingeteilt werden [2]. Beispielhafte Ansätze werden bereits von der ENISA (siehe Abbildung 1), der Agentur der Europäischen Union für Cybersicherheit [3], spezifiziert.

Cybersicherheitsübungen – die Methodik

Ausgangsbasis ist eine Idee, konkrete Anforderungen und Ziele der Übung in einer ersten Phase der Anforderungsanalyse zu definieren. Es gibt heutzutage sehr viele verschiedene Cybersicherheitsübungen, die von diskussionsorientierten Ansätzen bis hin zu sehr technisch und hybrid-realisierten Ansätzen reichen. Sie alle verfolgen unterschiedliche Ziele und Resultate [4]. Die Wahl der Übung legt daher auch die weiteren Inhalte und Möglichkeiten insbesondere für Teilnehmende fest, sich im Rahmen einer solchen Übung einzubringen. Beispielsweise kann in einer diskussionsorientierten Übung sehr wenig auf technische Details eingegangen werden, da oftmals nur Papier und Stift verwendet werden können. In einer technischen Übung wird allerdings oftmals die Suche in den technischen Infrastrukturen ermöglicht und auch ein entsprechendes technisches Knowhow bei den Teilnehmenden vorausgesetzt.

In der zweiten Phase, der Übungsplanung, wird ein Drehbuch erstellt, das einzelne Geschichten und Szenarien definiert. Diese Szenarien helfen den Übungsverlauf erfolgreich zu planen. Die Herausforderung liegt darin, ein Szenario fesselnd, interaktiv, informativ, anonym, nachvollziehbar und sicherheitsfokussiert zu gestalten [2]. Ein Szenario kann zum Beispiel der Ausfall wichtiger technischer Infrastrukturen oder



1 Häufige Phasen einer Cybersicherheitsübung (basierend auf [2,3])

die Erpressung mit *Ransomware* sein. Ransomware ist eine Software, die besitzende Personen von ihren Geräten oder Teilen aussperrt, um dann eine Lösegeldforderung (engl. *Ransom*) zu stellen. Dabei wird häufig die Festplatte ganz oder teilweise verschlüsselt. Ausgehend von diesem Szenario werden dann beide Seiten der Übung geplant. Dabei wird berücksichtigt wie z. B. herausfordernde Situationen durch Einspielungen herbeigeführt werden können und wie sie im Sinne des Übungsziels gestaltet sein sollten. Einspielungen dienen dann dazu diesen Übungsverlauf zu realisieren und können dabei unter anderem auf Mittel wie Emails aber auch technische Informationen zu Angriffen (engl. *Cyber Threat Intelligence*) zurückgreifen. Auf der anderen Seite werden die zu erwarteten Rückmeldungen durch die Teilnehmenden analysiert, um auf Eventualitäten vorbereitet zu sein und um die Ziele der Übung zu unterstützen.

In technischen Übungen werden in der Übungsplanung auch Anforderungen für die Simulationsumgebungen definiert, die dann in weiterer Folge implementiert werden. Oftmals werden Cybersicherheitsübungen in virtuellen Umgebungen, sogenannten *Cyber Ranges* (z. B. [1, 5]) aufgebaut. Diese ermöglichen eine vielfältige und variable Softwareumgebung, wie zum Beispiel eine exemplarische IT-Landschaft eines kleinen- und mittleren Unternehmens. Es werden auch Angriffe auf die exemplarische Infrastruktur in der *Cyber Range* vorbereitet. Ein Probedurchlauf vor der eigentlichen Durchführung ist sehr empfehlenswert.

In der weiteren Phase der Durchführung steigt die Spannung, da das vorbereitete Szenario dann gestartet wird und die Teilnehmenden mit neuen Situationen konfrontiert werden. Die Einspielungen führen dazu, dass die Übungssituation zur Realität wird und spielerisch der Ernstfall erprobt werden kann. Es werden Cyberangriffe simuliert und dadurch die Notfallpläne der Organisationen und die Reaktionen der Teilnehmenden getestet. Zudem lernen die Teilnehmenden mit neuen und herausfordernden Situationen umzugehen.

In der Phase der Evaluierung, werden die wichtigsten Eckpunkte der Übung reflektiert und das Feedback der Teilnehmenden aufgenommen. Das ist insbesondere wichtig, um den Teilnehmenden die wichtigsten Erkenntnisse mit in die Praxis zu geben.

Neue Herausforderungen der künstlichen Intelligenz in Cybersicherheitsübungen meistern

Cybersicherheitsübungen sind eine ideale Testumgebung für neue Technologien. Neue Werkzeuge und Technologien der künstlichen Intelligenz (KI) können im Rahmen von Cybersicherheitsübungen eingesetzt und getestet werden. Auf Seite der Organisationen und Teilnehmenden kann in Übungen das Potenzial für KI-basierte Cyberangriffe durch Übungen besser verstanden und Notfallpläne adaptiert werden. Neu ist der Einsatz von KI in diesem Bereich keineswegs, denn bereits seit vielen Jahren werden beispielsweise Verfah-

ren zur Mustererkennung dazu genutzt, um Cyberangriffe zu erkennen. Heute gilt es, unter diesen angepassten technologischen Vorzeichen, neue Konzepte zur Erkennung und Minimierung der Auswirkungen zu entwickeln und diese durch Übungen in einem geschützten Rahmen zu erproben. Mit der Forschung können neue Grundlagen und Verfahren entwickelt werden, die eine schnellere Entdeckung und Bearbeitung von Angriffen sowie den dazu notwendigen Informationsaustausch unterstützen.

Über den Lehrstuhl

Der Lehrstuhl von Prof. Dr. Maria Leitner befasst sich mit zentralen Fragestellungen im Bereich Cybersicherheit, Automatisierung und KI. Es geht darum neue Herausforderungen in der Cybersicherheit mit Konzepten, Algorithmen und Anwendungen grundlagennahe zu erforschen und auch praxisorientierte Resultate zu erzielen. In langjährigen Kooperationen mit Wissenschaft, Behörden, Wirtschaft und Industrie werden neue Ideen gemeinsam erarbeitet und Ergebnisse evaluiert.

Literatur

- [1] M. Leitner u. a., »Enabling exercises, education and research with a comprehensive cyber range.«, *J Wirel Mob Netw. Ubiquitous Comput Dependable Appl*, Bd. 12, Nr. 4, S. 37–61, 2021.
- [2] M. Leitner, »A Scenario-Driven Cyber Security Awareness Exercise Utilizing Dynamic Polling: Methodology and Lessons Learned«, in *Proceedings of the 9th International Conference on Information Systems Security and Privacy, ICISSP 2023, Lisbon, Portugal, February 22–24, 2023*, P. Mori, G. Lenzini, und S. Furnell, Hrsg., SciTePress, 2023, S. 634–642
- [3] A. Ogee, R. Gavrilu, P. Trimintzios, V. Stavropoulos, und A. Zacharis, »The 2015 report on national and international cyber security exercises: Survey, analysis and recommendations«. ENISA, 2015. [Online]. Verfügbar unter: <https://www.enisa.europa.eu/publications/latest-report-on-national-and-international-cyber-security-exercises?v2=1>
- [4] S. Kucek und M. Leitner, »An Empirical Survey of Functions and Configurations of Open-Source Capture the Flag (CTF) Environments«, *J. Netw. Comput. Appl.*, Bd. 151, Feb. 2020
- [5] M. M. Yamin, B. Katt, und V. Gkioulos, »Cyber ranges and security testbeds: Scenarios, functions, tools and architecture«, *Comput. Secur.*, Bd. 88, Jan. 2020

DREI FRAGEN ...

Was ist Ihre größte fachliche Herausforderung?

Durch die alltäglichere Nutzung von Software in allen Lebensbereichen wird es immer mehr Daten geben, die Aktivitäten aller Art beschreiben. Sicherheit und Privatsphäre sind daher zentrale Fragestellungen für unser digitales Leben. Eine der vielen Fragestellungen ist dabei, wie wir Daten noch effizienter interpretieren können, um rascher Cyberangriffe zu entdecken.

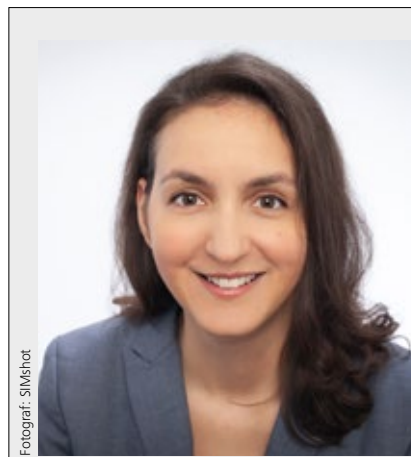
Wie verändert KI Ihr Fachgebiet?

KI ist schon lange in der Cybersicherheit verankert. KI wird beispielsweise bei der Filterung von unerwünschten Emails (engl. *Spam*) schon lange angewendet bzw. auch bei der Analyse von Netzwerkdaten, um Cyberangriffe zu erkennen. Der leichte Zugang zu KI verändert jedoch auch die Möglichkeiten der Angreifenden zum Bei-

spiel im Bereich der Desinformation. Es ist jetzt noch leichter geworden Falschinformationen automatisiert zu generieren. Dies gilt auch für die Nachbildung von Stimmen oder Videos. Es ist daher wichtig künftig auch Maßnahmen und Techniken zu entwickeln, die mögliche KI-generierte Inhalte erkennen können und auch die Menschen auf diese Möglichkeiten hinzuweisen, um effektive Schutzmaßnahmen zu treffen.

Welche Entwicklung in der Informatik wünschen Sie sich für die nächsten 5 Jahre?

Ich würde mir wünschen, dass das Wissen und die Denkweisen der Informatik und insbesondere der Informationssicherheit noch stärker an verschiedene Zielgruppen allen Alters vermittelt werden und damit langfristig zur digitalen Resilienz unserer Gesellschaft beitragen.



Fotograf: SIMshot

Prof. Dr. Maria Leitner ist Professorin für KI in der IT-Sicherheit und forscht an den Schnittstellen von KI, Cybersicherheit und Automatisierung. Davor war Sie Gastprofessorin für Informatik an der Universität Wien und Scientist am AIT Austrian Institute of Technology. Am AIT hat sie bis 2021 die AIT Cyber Range geleitet. Unter anderem hat Sie Cybersicherheitsübungen für insgesamt mehr als 350 Personen konzipiert, umgesetzt und angeleitet, wie beispielsweise eine nationale Cybersicherheitsübung. Prof. Dr. Maria Leitner hat zwei Bücher herausgegeben und über 50 von Fachleuten begutachtete wissenschaftliche Publikationen verfasst.



Maschinelles Lernen mit Anwendungen in den Naturwissenschaften

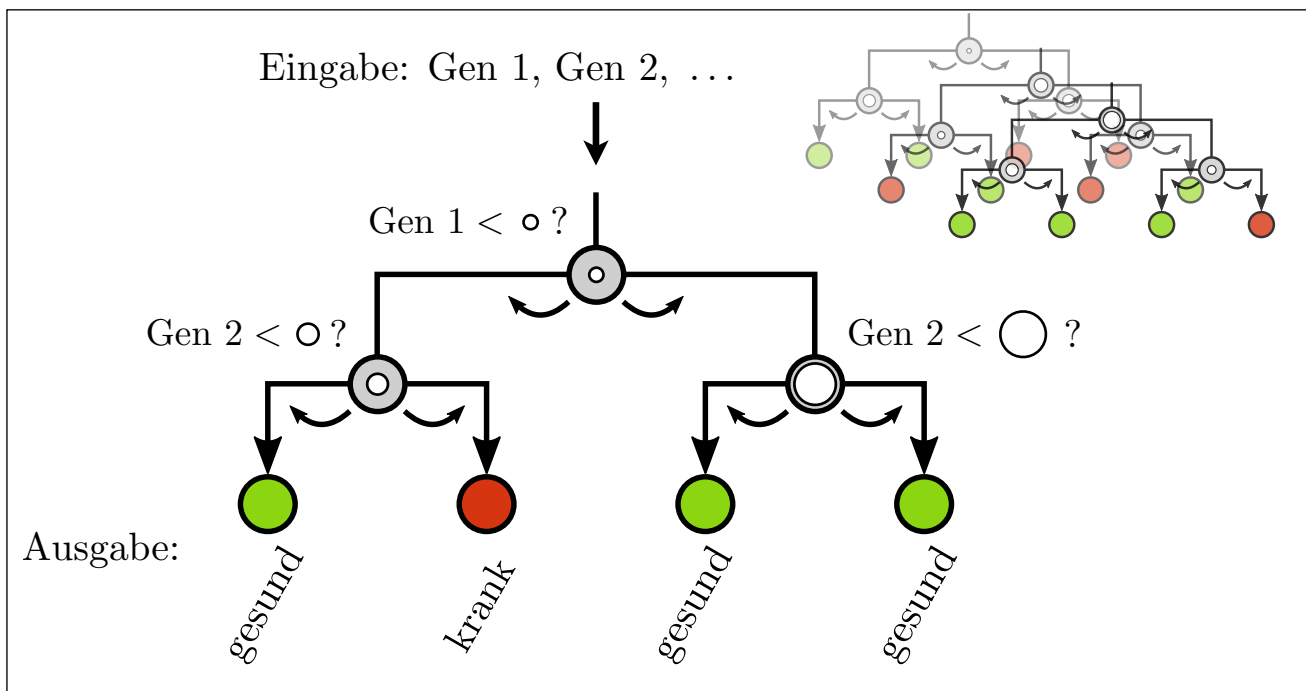
Prof. Dr. Merle Behr, Dr. Markus Schmitt

Eine der zentralen Triebfedern für die Entwicklung moderner Methoden des maschinellen Lernens ist die Verfügbarkeit zuvor unvorstellbar großer Datenmengen durch umfassende Digitalisierung und die weltweite Vernetzung über das Internet. Auf dieser Grundlage konnten zum Beispiel auf den Gebieten der Bilderkennung, der Verarbeitung natürlicher Sprachen und der automatischen Strategiefindung in komplexen Umgebungen bedeutende Durchbrüche erzielt werden, wodurch das maschinelle Lernen auch unseren Alltag

zunehmend mitbestimmt. In dieser Hinsicht zeigen insbesondere die neuesten Sprachmodelle wie GPT oder Bard zukünftige Möglichkeiten auf, die noch vor kurzem undenkbar erschienen, und illustrieren damit populär die Bedeutung dieser technologischen Umwälzung.

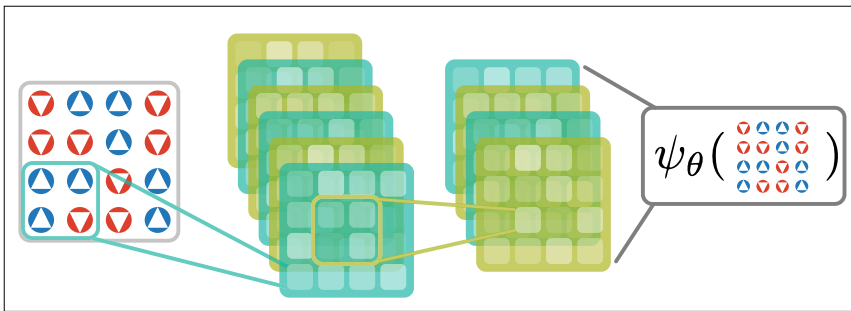
Auch in den Naturwissenschaften spielen die Methoden des maschinellen Lernens eine stetig wachsende Rolle. Viele Experimente oder Computersimulationen erzeugen große, hochdimensionale Datenmengen, aus denen die entscheidende

Information für die wissenschaftliche Erkenntnis extrahiert werden muss. Diese Anwendungen stellen spezielle Anforderungen an die verwendeten Algorithmen, da sie oft besonders präzise sein müssen, statistische Abweichungen zu quantifizieren sind und die Funktionsweise durchschaubar sein muss, um wissenschaftliche Schlüsse zu ziehen. Die Herausforderung besteht also darin, effiziente Algorithmen zu entwickeln, die idealerweise gleichzeitig den Analysemethoden der mathematischen Statistik zugänglich sind. Darüber



Bildnachweis: Markus Schmitt

1 Schematische Darstellung des Random Forest Algorithmus angewendet auf genetische Daten, um eine Krankheit vorherzusagen. Für ein Individuum, bei dem eine Krankheit vorhergesagt werden soll, werden unterschiedliche genetische Informationen (z. B. über Gen 1, Gen 2, usw.) als Eingabe an den Algorithmus übergeben. Der Algorithmus hat vorab aus Trainingsdaten unterschiedliche Entscheidungsbäume erlernt. Jeder Entscheidungsbaum liefert eine Vorhersage, ob das Individuum krank oder gesund ist. Die Vorhersagen der unterschiedlichen Entscheidungsbäume werden dann im Random Forest Algorithmus zu einer einzigen Vorhersage zusammengefasst.



2 Schematische Darstellung eines künstlichen neuronalen Netz, das eine quantenmechanische Wellenfunktion in komprimierter Form repräsentiert. In diesem Beispiel ordnet die Wellenfunktion jeder möglichen Kombination von »up«- (blau) und »down«- (rot) Zuständen der 16 quantenmechanischen Spins (links im Bild) eine komplexe Zahl (rechts im Bild) zu. Anstelle einer großen Nachschlagetabelle, in der alle Werte gespeichert sind, verarbeitet das neuronale Netz die gegebene Konfiguration und berechnet die zugehörige komplexe Zahl. Es ist dabei durch eine kleinere Menge von Gewichten (θ) parametrisiert und stellt in diesem Sinne eine komprimierte Form der Wellenfunktion dar.

hinaus hat die allgemeine Forschung an Methoden des maschinellen Lernens neue algorithmische Bausteine hervorgebracht, die als Teil wissenschaftlicher Computersimulationen auch die Erzeugung von Daten beschleunigen können.

Der Lehrstuhl »Maschinelles Lernen« an der Fakultät für Informatik und Data Science beschäftigt sich neben der grundlegenden Entwicklung und Analyse von Methoden des maschinellen Lernens insbesondere mit deren Anwendungen auf den Gebieten der Biologie, Medizin und Physik. Im Folgenden werden zwei Beispiele konkreter erläutert, in denen zwei der bekanntesten und erfolgreichsten Algorithmen des maschinellen Lernens verwendet werden: der Random Forest Algorithmus mit Anwendung in der Genetik und Künstliche Neuronale Netzwerke mit Anwendung in der Quantenphysik.

Genetische Interaktionen in Bäumen des Random Forest

Innerhalb zahlreicher Fachgebiete der Medizin und Biologie stellt sich eine essenzielle Frage: Wie hängen genetische Merkmale und phänotypische Eigenschaften, wie beispielsweise das Auftreten bestimmter Krankheiten, miteinander zusammen? Diese Verbindung ist in den meisten Fällen äußerst komplex. Zum einen beeinflussen oft vielfältige genetische Varianten das Erscheinungsbild einer Krankheit. Zum anderen interagieren diese genetischen Varianten häufig auf komplexe Weise miteinander, wodurch ihr Einfluss, etwa auf die Ausprägung einer Krankheit, nicht isoliert betrachtet werden kann. Methoden des Maschinellen Lernens stel-

len eine Möglichkeit dar, solche komplexen, nicht-linearen Zusammenhänge aus Datensätzen zu erlernen. Eine besonders effektive Gruppe von Algorithmen zur Darstellung von Interaktionseffekten sind sogenannte tree-ensemble Methoden, wobei der Random Forest Algorithmus als einer der prominentesten und erfolgreichsten Vertreter gilt. Im Gegensatz zu künstlichen neuronalen Netzen, die auf einer Netzwerkstruktur basieren (siehe nächster Abschnitt), nutzen Random Forests Baumstrukturen. Dabei werden verschiedene Entscheidungsbäume kombiniert, um Zusammenhänge – beispielsweise zwischen genetischen Variationen und Krankheitsbildern – darzustellen. Abbildung 1 zeigt eine schematische Darstellung eines solchen Random Forest Algorithmus. Die Arbeitsgruppe um Professorin Merle Behr widmet sich unter anderem der Herausforderung, wie sich Interaktionseffekte aus den Baumstrukturen des Random Forests auf eine leicht verständliche Art extrahieren lassen. Dieser Ansatz ermöglicht es Medizinern und Biologen, die komplexen Zusammenhänge, die der Random Forest erlernt hat, nachzuvollziehen. Letztendlich kann dies dazu beitragen, gezielte Therapiemaßnahmen für Krankheiten zu entwickeln. Zum Beispiel kooperiert die Arbeitsgruppe eng mit Medizinern und Biologen innerhalb des DFG Transregio TRR 374, wo Forschung an Nierenerkrankungen betrieben wird. Ein zentraler Schwerpunkt der Arbeitsgruppe Behr liegt dabei nicht nur auf der Entwicklung von Methoden und Algorithmen – beispielsweise zur Extraktion von Interaktionseffekten mittels Random Forests – sondern auch auf der mathematischen und statistischen Analyse dieser

Ansätze. Dies ermöglicht es einerseits, die Bedingungen zu identifizieren, unter denen die verwendeten Methoden valide Ergebnisse liefern. Andererseits lässt sich die Unsicherheit der Ergebnisse mathematisch präzise quantifizieren. Beides ist gerade bei medizinischen Anwendungen von großer Bedeutung, da hier eine besonders hohe Zuverlässigkeit und Interpretierbarkeit der Ergebnisse notwendig ist.

Simulation komplexer Quantensysteme mit künstlichen neuronalen Netzwerken

Die Simulation komplexer Quantensysteme stellt eine der großen Herausforderungen in der rechnergestützten Physik dar. Die mathematische Modellierung basiert auf einem Zustandsraum, der exponentiell mit der Zahl der im System enthaltenen Teilchen wächst, wodurch naive Herangehensweisen selbst die Kapazitätsgrenzen der größten Supercomputer sehr schnell erreichen. Für effizientere Algorithmen muss die zusätzliche Struktur physikalisch interessanter Zustände genutzt werden, um eine komprimierte Beschreibung zu erhalten. Während man die Besonderheiten vieler relevanter Situationen verstanden und in numerischen Methoden nutzbar gemacht hat, sind Systeme jenseits des Gleichgewichts oder in zwei räumlichen Dimensionen weiterhin äußerst schwierig zu behandeln. Um dieses Regime zu erschließen, das aktuell in den Fokus experimentell realisierter Quantensimulatoren rückt, verfolgt unsere Arbeitsgruppe Computational Quantum Science unter der Leitung von Dr. Markus Schmitt, einen neuen Ansatz, der auf der Kompression des quantenmechanischen Zustands durch künstliche neuronale Netzwerke beruht (siehe Abbildung 2). Während diese Methodenentwicklung einerseits von den allgemeinen Fortschritten im Bereich des Deep Learning profitiert, sind für die numerischen Simulationen gleichzeitig spezielle Herausforderungen zu bewältigen – zum Beispiel, weil zur Optimierung der neuronalen Netze anders als in herkömmlichen Anwendungen die Trainingsdaten aus diesen selbst generiert und mit den physikalischen Bewegungsgleichungen kombiniert werden müssen. Erste Anwendungen demonstrieren, dass auf diesem Weg neue physikalische Situationen numerisch zugänglich werden, wie etwa universelles Verhalten in der Phasenübergangsdynamik von zweidimensionalen Quantenmagneten.



Bildnachweis: Markus Schmitt

3 Gruppenbild der Mitglieder des Lehrstuhls für Maschinelles Lernen an der Fakultät für Informatik und Data Science. Hinten (von links nach rechts): Dr. Markus Schmitt (Gruppenleiter), Dr. Jonas Rigo, Wladislav Krinitsin, Nicolas Ihlo, Prof. Merle Behr (Lehrstuhlinhaberin); vorne (von links nach rechts): Laura Pietschmann (Sekretariat), Mohammad Abedi, Dr. Kata Vuk, Francesca de Franco

DREI FRAGEN ...

Welche Entwicklung im Bereich Data Science hat Sie im letzten Jahr am meisten überrascht?

Wie wahrscheinlich die meisten Wissenschaftler im Bereich Data Science fällt uns hier als erstes ChatGPT ein. Was uns auch überrascht hat, ist, dass Europa bis jetzt keine größeren Anstrengungen unternimmt, um ein eigenes Large Language Model zu etablieren.

Welche Entwicklung in dem Bereich Data Science wünschen Sie sich für die nächsten 5 Jahre?

Eine noch bessere Digitalisierung und Verfügbarkeit von Daten, die für wissenschaftliche Publikationen verwendet werden; außerdem eine Etablierung von Data Science als eines der grundlegenden Studienfächer in Deutschland.

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

In der Data Science findet häufig theoretische und angewandte Forschung getrennt voneinander statt. In unserer Arbeitsgruppe möchten wir beides möglichst gut miteinander verbinden, weil wir davon überzeugt sind, dass beide Bereiche voneinander profitieren.



Foto © Markus Schmitt

Prof. Dr. Merle Behr leitet seit 2022 den Lehrstuhl für Maschinelles Lernen an der Universität Regensburg. Nach ihrer Promotion in Mathematik an der Georg-August Universität Göttingen im Jahr 2018 verbrachte sie zweieinhalb Jahre als Postdoc und Neyman Visiting Assistant Professor am Department of Statistics der UC Berkeley. Vor ihrem Wechsel nach Regensburg war sie in der Pharmaforschung der Bayer AG tätig. Ihre Forschung konzentriert sich auf die statistische Analyse und Weiterentwicklung von Methoden des Maschinellen Lernens, insbesondere tree-ensemble Methoden, change-point Methoden, causal inference und blind source separation.

statistische Analyse und Weiterentwicklung von Methoden des Maschinellen Lernens, insbesondere tree-ensemble Methoden, change-point Methoden, causal inference und blind source separation.



Foto © Markus Schmitt

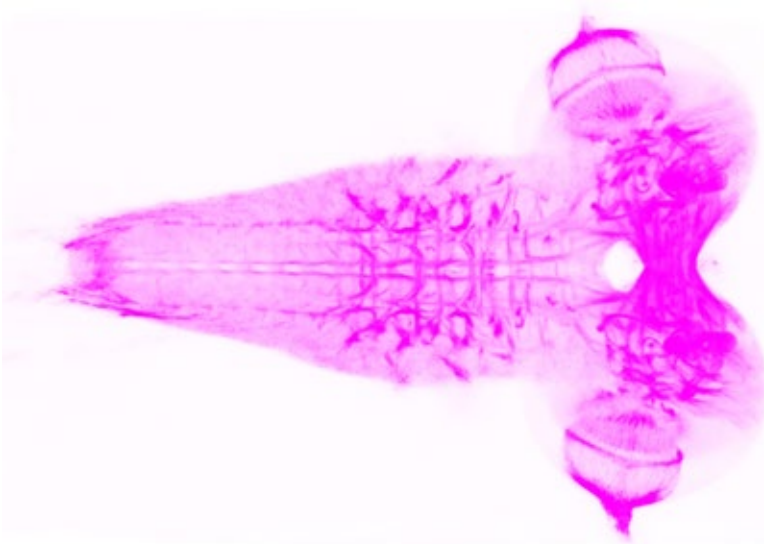
Dr. Markus Schmitt leitet die Gruppe "Computational Quantum Science" im Lehrstuhl Maschinelles Lernen an der Universität Regensburg. Gleichzeitig ist er Helmholtz Young Investigator Group Leader am Forschungszentrum Jülich. Nach seiner Promotion in Physik an der Georg-August-Universität Göttingen im Jahr 2018 absolvierte er verschiedene Postdoc-Stationen, z.B. als Leopoldina Postdoctoral Fellow an der UC Berkeley. Seine Forschung dreht sich um

komplexe Quantensysteme jenseits des Gleichgewichts, mit einem Fokus auf der Entwicklung von neuen numerischen Methoden, die Ideen des maschinellen Lernens mit einbeziehen.

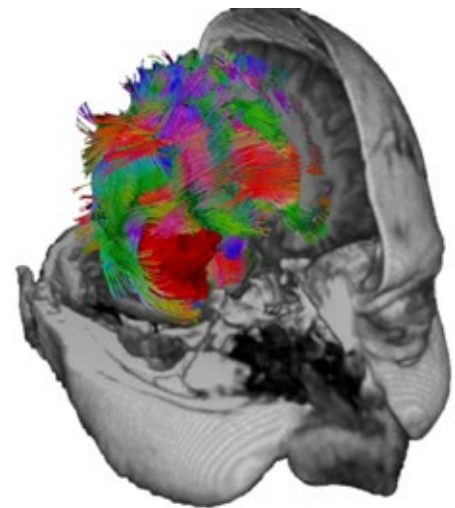
Automatisierte, KI-basierte Analyse von Bilddaten

Der Lehrstuhl für Bildverarbeitung

Prof. Dr.-Ing. Dorit Merhof



1 Die Fruchtfliege (*Drosophila melanogaster*) ist ein wichtiger Modell-Organismus in den Lebenswissenschaften, um grundlegende Forschungsfragen (u. a. zu Verhalten, Lernen, Krankheitsmechanismen) zu erforschen. Das Bild zeigt einen Lichtmikroskopie-Atlas des Larvengehirns, der durch Verfahren der Bildverarbeitung aus einer Vielzahl an Lichtmikroskopie-Aufnahmen erstellt wurde. Projekt: <https://gepris.dfg.de/gepris/projekt/441181781>



2 Die Diffusions-MRT Bildgebung ermöglicht eine Messung des Diffusionsverhaltens von Wasser im Gehirn und somit Aussagen über den Verlauf neuronaler Bahnen im Gehirn, was beispielsweise für die Planung von Gehirntumor-OPs gewinnbringend eingesetzt werden kann. Die Abbildung zeigt rekonstruierte neuronale Bahnen und deren Verlauf um einen Gehirntumor. Projekt: <https://gepris.dfg.de/gepris/projekt/417063796>

Die fundamentale Bedeutung visueller Information kommt bereits im Sprichwort »Ein Bild sagt mehr als tausend Worte« zum Ausdruck. Tatsächlich erfolgt die Informationsaufnahme beim Menschen zu über 80% über das visuelle System, so dass Bild- und Videodaten in verschiedenen Bereichen des Alltags, des Berufslebens und in vielen Forschungsdisziplinen von hoher Relevanz sind.

Die automatisierte, Algorithmen-basierte Analyse von Bild- und Videodaten

biete dabei eine Reihe von Vorteilen: Hierdurch wird eine konsistente, Betrachter-unabhängige Analyse möglich, die Analyse ist reproduzierbar, kann in geringer Zeit in großem Umfang durchgeführt werden und erlaubt es menschlichen Experten, sich stattdessen auf Tätigkeiten zu fokussieren, die bislang nicht durch Algorithmen übernommen werden können.

In den letzten Jahren haben KI-Verfahren basierend auf neuronalen Netzen in der Bildverarbeitung zunehmend an Be-

deutung gewonnen. Sofern ausreichend Bilddaten für das Training neuronaler Netze vorhanden sind, erreichen moderne KI-basierte Verfahren in vielen Fällen deutlich bessere Ergebnisse als bisherige klassische Bildverarbeitungsverfahren. Insgesamt tragen KI-basierte Verfahren in der Bildverarbeitung somit dazu bei, die Leistungsfähigkeit und die Einsatzmöglichkeiten von Bilddaten in einer Vielzahl von Anwendungen und Forschungsbereichen zu erweitern. Sie ermöglichen eine schnellere, genauere und

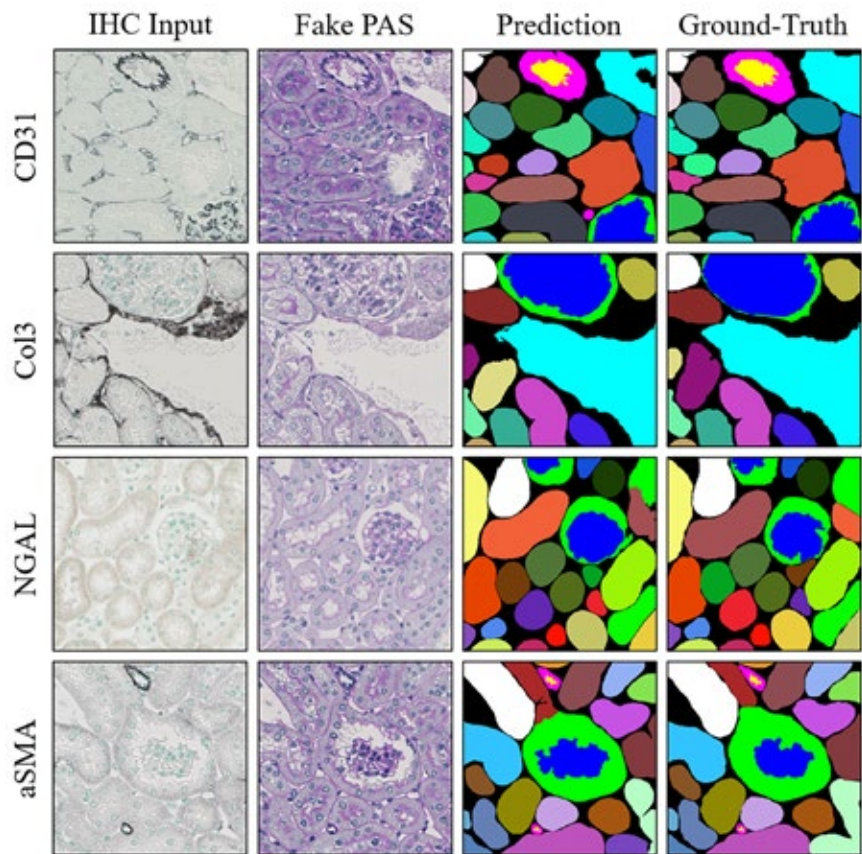
effizientere Analyse und Interpretation von Bildinformationen, was zu bedeutenden Fortschritten in Wissenschaft, Medizin, Industrie und vielen anderen Bereichen führt.

Eine besondere Herausforderung für Bildverarbeitung im Allgemeinen (und KI-basierter Bildverarbeitung im Speziellen) ist die Analyse biologischer und medizinischer Bilddaten. Dies liegt darin begründet, dass im biomedizinischen Kontext stetig neue, komplexe Bildgebungstechniken entwickelt werden (z. B. Super-Resolution Mikroskopie-Verfahren, hochdimensionale Multiplexing-Verfahren in der Histopathologie, oder neue MRT-Sequenzen für anatomische Bildgebung). Derartige innovative, komplexe Bildgebungsverfahren ziehen auch einen dringenden Bedarf an dedizierten Bildverarbeitungsalgorithmen nach sich, um die generierten Bilddaten optimal zu analysieren und daraus letztlich biologische oder medizinische Erkenntnisse oder Diagnosen gewinnen zu können.

In unserer aktuellen Forschung am Lehrstuhl für Bildverarbeitung beschäftigen wir uns daher mit unterschiedlichen Typen von neuronalen Netzen bzw. KI-Modellen, um verschiedene Problemstellungen – insbesondere im Kontext biomedizinischer Bilddaten – zu adressieren:

Kontinuierliches Lernen: Wir entwickeln neuartige Techniken, die es KI-Modellen ermöglichen, fortlaufend zu lernen und sich anzupassen, falls die zu analysierenden Bilddaten leicht variieren. In der klinischen Praxis treten solche Variationen häufig auf, bspw. bei Bilddaten unterschiedlicher (a) Kliniken, (b) Bildgebungsprotokolle, (c) Färbetechniken oder (d) Gerätehersteller. Kontinuierliches Lernen ist entscheidend um sicherzustellen, dass KI-Systeme Wissen und Fähigkeiten beibehalten und stetig anreichern, ohne vorherige Informationen zu vergessen, so dass die Systeme besser generalisieren und auch bei leicht unterschiedlichen Eingabe-Bildern leistungsfähig bleiben.

Selbstüberwachtes Lernen: Die Annotation von Bilddaten für das Training neuronaler Netze bringt einen großen manuellen Arbeitsaufwand mit sich. Wir entwickeln daher selbstüberwachte Algorithmen, die auch Bilddaten effektiv nutzen können, welche keine manuellen Annotationen aufweisen. Dieser Ansatz verringert nicht nur die Notwendigkeit für umfangreiche manuelle Annotationen, sondern verbessert auch die Skalierbarkeit und Anwendbarkeit von KI-Modellen.



3 Die Digitalisierung pathologischer Gewebeproben zur quantitativen, automatisierten, Benutzer-unabhängigen und reproduzierbaren Analyse mittels Bildverarbeitung gewinnt zunehmend an Bedeutung. Da Analyse-Algorithmen typischerweise auf spezifischen Färbungen trainiert wurden, können Färbungs-Translationen dazu beitragen, dass existierende Algorithmen in einer größeren Breite eingesetzt werden können. Im Bild sind verschiedene Immunohistochemische Färbungen (IHC Input) zu sehen, die in die Standard-Färbung PAS transliert wurden, so dass eine für die PAS-Färbung existierende automatische Segmentierung (prediction) weiterhin angewendet werden kann. Projekt: <https://gepris.dfg.de/gepris/projekt/459599325>

Generative Modelle: Generative Modelle wie Variational Autoencoder (VAE), Generative Adversarial Networks (GANs) und neuerdings Diffusions-Modelle haben einen festen Platz in der biomedizinischen Bildgebung und -Analyse eingenommen. Durch Generative Modelle gelingt es, Repräsentationsräume für biomedizinische Bilddaten zu finden, zusätzliche Beispieldaten zu generieren oder lückenhafte Bilddaten zu ergänzen. Diese Möglichkeiten sind von großem Wert für die Datenanreicherung, Simulation und Verbesserung der Robustheit von KI-basierten Verfahren zur Diagnose und Behandlungsplanung basierend auf medizinischen Bilddaten. In unserer Forschung arbeiten wir seit mehreren Jahren an der Weiterentwicklung Generativer Modelle für unterschiedliche Anwendungsszenarien.

Foundation-Modelle: Große Sprachmodelle – auch Foundation-Modelle genannt – können bereits anspruchsvolle

Text-bezogene Aufgaben automatisiert durchführen, ein Beispiel ist hier GPT-3. Inzwischen gibt es auch Foundation-Modelle, die gleichzeitig Einbettungen für Bilddaten und Texte erzeugen und diese assoziieren können. Im biomedizinischen Kontext bieten sich hier vielfältige Nutzungsmöglichkeiten (z. B. automatisierte, simultane Analyse medizinischer Bilddaten und der zugehörigen Arztbriefe). Wegen ihres großen Anwendungsspektrums bieten Foundation-Modells das Potential, nochmals deutlich fortgeschrittenere KI-Systeme zu entwickeln.

Integration multi-modaler Daten: Sowohl in der klinischen Forschung als auch für Fragestellungen zur Diagnose, Therapieplanung und Prognose werden heute vielfältige, komplexe Daten erhoben – angefangen von genetischen Analysen, über verschiedene -Omics-Daten bis hin zu histologischen und radiologischen Bilddaten. Die Integration solcher multi-modaler

Daten stellt eine große Herausforderung dar, die wir im Rahmen unserer Forschung mittels KI-basierter Methoden adressieren.

Optimierung des Energieverbrauchs: Energieeffiziente KI ist entscheidend für den nachhaltigen und ressourcenschonenden Einsatz von KI und somit für die Verbreitung von KI-basierten Verfahren. Durch Optimierungsverfahren auf algorithmischer Ebene sowie Betrachtung von untereinander abhängigen arithmetischen

Operationen werden wir für unterschiedliche KI-Modelle Optimierungen entwickeln, die den Energieverbrauch erheblich reduzieren, ohne die Leistung des KI-Modells zu beeinträchtigen. Dies hat nicht nur Vorteile für die Umwelt, sondern eröffnet auch Möglichkeiten für KI-Anwendungen in ressourcenbeschränkten Umgebungen.

Durch unsere Forschung in diesen Bereichen hoffen wir, zur Weiterentwicklung von KI-Technologien – insbesondere für biomedizi-

nische Anwendungen – beizutragen. Unser Ziel ist es, KI-Technologien zu entwickeln, die nicht nur leistungsfähiger und anpassungsfähiger sind, sondern auch nachhaltig eingesetzt werden können. Diese Verfahren bieten das Potential, für eine breite Palette von Anwendungen – insbesondere im Gesundheitswesen – wichtige Aufgaben zu erfüllen, menschliche Nutzer zu entlasten und einen Beitrag zu wichtigen medizinischen, gesellschaftlichen oder industriellen Aufgaben zu leisten.

DREI FRAGEN ...

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

In unserer Forschung arbeiten wir an robusten, flexiblen und nachhaltig einsetzbaren KI-Verfahren zur Analyse von Bilddaten, insbesondere im biomedizinischen Bereich. Unser Ziel ist es, durch eine schnellere, genauere und effizientere Analyse und Interpretation komplexer Bildinformationen zu Fortschritten in der Medizin und den Lebenswissenschaften beizutragen, beispielsweise in den Bereichen Krebsforschung und -Therapie, Immunologie, Hämatologie/Blutkrebs, oder Demenzforschung.

Wie verändert KI Ihr Fachgebiet?

Das Fachgebiet der Bildverarbeitung hat durch KI-basierte Verfahren einen grundlegenden Wandel erlebt, da KI-basierte Verfahren vielfach klassischen Bildverarbeitungsverfahren überlegen sind und im Bereich der KI-basierten Verfahren nach wie vor ein großer Forschungsbedarf zu verzeichnen ist.

Was ist Ihre größte fachliche Herausforderung?

Die KI-Entwicklungen haben in den letzten Jahren eine extreme Dynamik entwickelt. Mit den Neuerungen und der Vielfalt an Entwicklungen in der Forschung Schritt zu halten ist somit eine gewisse Herausforderung, aber gleichzeitig sehr interessant und spannend.



Foto © privat

Prof. Dr. Dorit Merhof

studierte Informatik (Abschluss: Diplom, 2003, mit Auszeichnung) an der Friedrich-Alexander Universität Erlangen-Nürnberg. Danach promovierte sie am Lehrstuhl für Graphische Datenverarbeitung ebendort (Abschluss: Dr.-Ing., 2007, mit Auszeichnung). Anschließend folgte ein zweijähriger Auslandsaufenthalt in Oxford, UK, wo Dorit Merhof eine Stelle als "Research Scientist I" bei Siemens Molecular Imaging innehatte. Ab Oktober 2009 war Dorit Merhof an der Universität Konstanz als Juniorprofessorin (W1) für Visual Computing tätig. Im Juli 2013 folgte sie einem Ruf auf eine W3-Professur an der RWTH Aachen und leitete dort den Lehrstuhl für Bildverarbeitung in der Fakultät für Elektrotechnik und Informationstechnik. Seit Oktober 2022 ist sie W3-Professorin in der Fakultät für Informatik und Data Science an der Universität Regensburg und leitet dort den Lehrstuhl für Bildverarbeitung.

Die Genome des Menschen

Forschungsschwerpunkte der Arbeitsgruppe für Algorithmische Bioinformatik

Prof. Dr. Birte Kehr

Im Jahr 2000 verkündeten der amerikanische Präsident Bill Clinton und der britische Premierminister Tony Blair die Entschlüsselung des menschlichen Genoms. Damit war ein Meilenstein der Wissenschaft erreicht, der seitdem aus der modernen biomedizinischen Forschung auf molekularer Ebene nicht mehr wegzudenken ist. Dennoch war die zunächst veröffentlichte DNA-Sequenz weit davon entfernt, das Genom eines jeden Menschen abzubilden. Einerseits war sie nicht vollständig, denn es fehlten die schwer zu entschlüsselnden Telomer- und Centromerregionen. Andererseits gibt es Millionen von Unterschieden zwischen den Genomen zweier beliebiger Menschen mit Ausnahme eineiiger Zwillinge, so dass eine lineare DNA-Sequenz immer nur das Genom eines einzelnen Menschen darstellen kann. Heute ist klar, dass sich die Wissenschaft nicht allein auf eine lineare DNA-Sequenz stützen kann, sondern dass zusätzlich die genomische Variation abgebildet werden muss. Hierzu trägt die Forschung der Arbeitsgruppe für Algorithmische Bioinformatik bei.

Variation in der Genomsequenz ist verantwortlich für die sichtbare Diversität unter uns Menschen, beispielsweise für unsere Augenfarbe oder Körpergröße, aber ebenso auch für Unterschiede, die zunächst nicht sichtbar sind. Denn unser Genom beeinflusst, wie hoch unser Risiko ist, Krankheiten zu bekommen, und wie gut ein Patient auf eine medizinische Behandlung anspricht. Je besser daher unsere Fähigkeiten werden, Variationen in der Genomsequenz zu erfassen, umso

besser können wir die Hintergründe von Krankheiten aufklären und umso mehr von diesem Wissen kann in die biomedizinische Forschung und letztendlich in die Medizin einfließen.

Bis es jedoch routinemäßig möglich wird, die Variation in Patientengenomen annähernd vollständig zu erfassen, gibt es noch große Herausforderungen zu bewältigen. Beispielsweise sind die technologischen Möglichkeiten, um DNA-Sequenz aus einem DNA-Molekül auszulesen, begrenzt. Aktuell können nur relativ kurze Sequenzabschnitte am Stück »sequenziert« werden, die anschließend mithilfe von Methoden der Informatik wieder zusammengesetzt werden. Um die Variation in einem Genom zu erfassen, werden Millionen solcher kurzer Sequenzabschnitte, die in der Fachsprache einfach »Reads« genannt werden, erzeugt. Die schiere Größe der Genomdatensätze erfordert einen sparsamen Umgang mit Rechenzeit und Speicherverbrauch. Gleichzeitig müssen Fehleraten der Sequenzierungstechnologien in der Analyse der Daten sorgfältig behandelt werden. Unter anderem damit beschäftigt sich die aktuelle Forschung im Feld der Genomik und Bioinformatik.

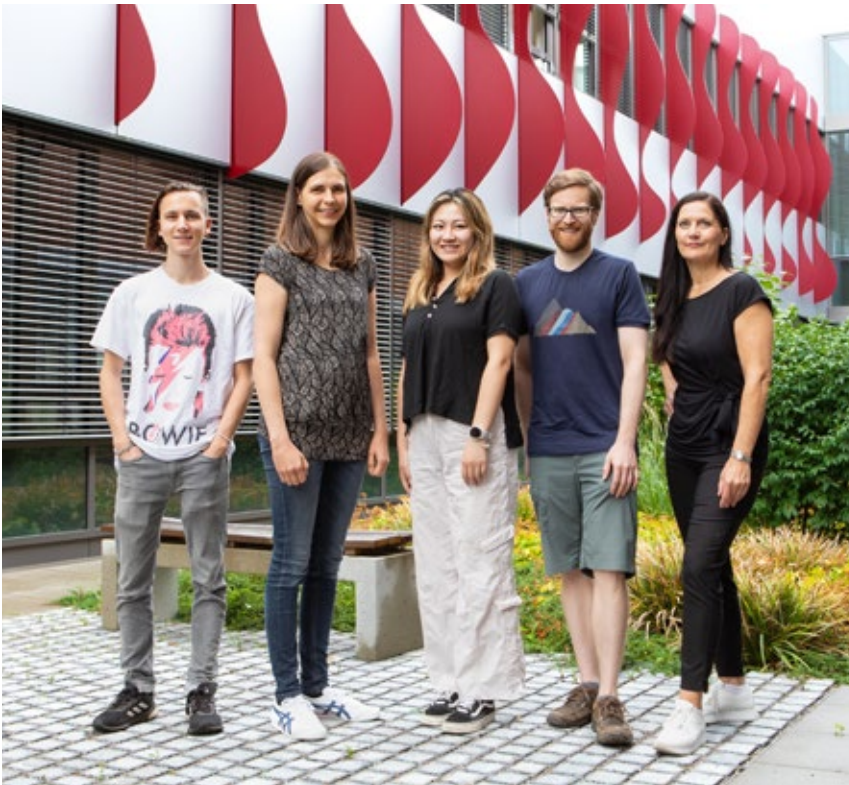
Bioinformatiker stellen sich Genome als lange Texte bestehend aus den Zeichen A, C, G und T vor. Hinter den Zeichen verbirgt sich die biochemische Zusammensetzung von DNA, die Nukleotide. Das menschliche Genom besteht aus gut 3 Milliarden aneinandergelagerten Nukleotiden, belegt also als Textdatei etwa 3 Gigabyte an Speicher auf einem Computer und würde ausgedruckt ein ganzes Regal voller Bücher füllen. Es ist in fast jeder Zelle unseres Körpers zu finden und ist dort in Chromosomen or-

ganisiert. Jedes Chromosom ist ein langes DNA-Molekül, aufgewickelt und verpackt in Proteine.

Im menschlichen Genom liegt jedes Chromosom – und damit alle 3 Milliarden Nukleotide – in zwei Kopien vor, da jeweils ein Chromosomensatz von der genetischen Mutter und ein Chromosomensatz vom genetischen Vater vererbt wird. Diese Kopien sind aktuellen Schätzungen zufolge zu etwa 99.6% identisch (NHGRI 2023), sofern unsere Eltern nicht nahe miteinander verwandt sind. Auf die Länge des menschlichen Genoms bedeuten 0.4% Abweichungen jedoch immer noch mehrere Millionen Unterschiede, die als Varianten bezeichnet werden. Natürlich findet man nicht nur zwischen den an uns von unseren Eltern vererbten Kopien des Genoms Millionen von Varianten, sondern auch beim Vergleich der Genome zweier beliebiger Menschen. Mindestens ein gewisser Anteil dieser Varianten kann physiologische Unterschiede zwischen uns Menschen erklären.

Die größte Zahl der Varianten sind Einzelnukleotidvarianten (engl. Single nucleotide variants, kurz SNVs), bei denen sich die Sequenz in genau einem Zeichen (Nukleotid) unterscheidet. Kleine Insertionen und Deletionen von 1–50 Nukleotiden werden als »Indels« bezeichnet. Beim Vergleich zweier menschlicher Genome findet man Millionen von SNVs und Hunderttausende Indels. Varianten, die einen Sequenzabschnitt von mehr als 50 Nukleotiden betreffen, bezeichnet man als »Strukturvarianten«. Zu ihnen zählen große Deletionen, Insertionen, Duplikationen, Inversionen, Translokationen und komplexe Varianten, an denen mehrere Sequenzabschnitte be-

Bildnachweis: UKR, Universitätsklinikum Regensburg



1 Die Arbeitsgruppe für Algorithmische Bioinformatik am Leibniz-Institut für Immuntherapie.

teilt sind. Aufgrund ihrer Größe betreffen Strukturvarianten weit mehr Nukleotide des menschlichen Genoms als SNVs und Indels zusammen, obwohl schätzungsweise nur 20–30 Tausend Strukturvarianten pro Genom vorkommen (Audano et al. 2019).

Die Arbeitsgruppe für Algorithmische Bioinformatik (Abb. 1) entwickelt neue Ansätze, um genomische Variation auf der Grundlage von Genomsequenzdaten zu beschreiben. Insbesondere interessiert uns, an welchen Positionen im Genom Strukturvarianten liegen und wie sie aussehen (Detektion) und, ob sie auf keiner, einer oder beiden Kopien der Chromosomen vorliegen (Genotypisierung). Dabei formulieren wir die konkreten Fragestellungen nach Möglichkeit als bekannte algorithmische Optimierungsprobleme, um sie dann mit etablierten Verfahren der Informatik lösen zu können. Fehlerwahrscheinlichkeiten

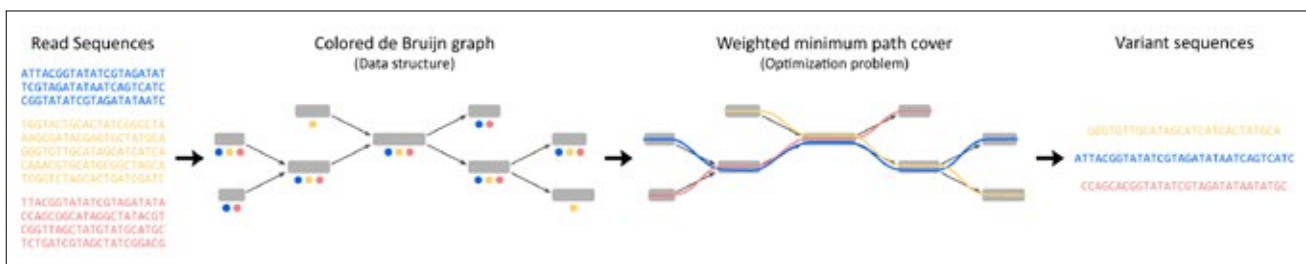
und andere Unsicherheiten in den Daten beschreiben wir mit statistischen Modellen und lassen sie als Gewichte in die Optimierungsprobleme einfließen.

Ein Beispiel für unsere Arbeiten ist die Skalierung von bestehenden Ansätzen für die Detektion und Genotypisierung von Strukturvarianten auf die stetig wachsende Zahl sequenzierter Genome (Krannich et al. 2022, Niehus et al. 2021). Wir wollen nutzen, dass die Daten von vielen Genomen sich gegenseitig ergänzen und die Evidenz für viele Varianten mit jedem analysierten Genomdatensatz wächst. Dadurch, dass die allermeisten genomischen Varianten vererbt werden, kommt ein großer Anteil von ihnen in mehr als einer Person vor. Jedes Kind erbt die Variation von seinen Eltern, die diese von den Großeltern und diese wieder von ihren Eltern geerbt haben, und so weiter. Als Folge der Ver-

erbungsregeln gibt es über alle Kontinente hinweg Varianten, die häufiger oder weniger häufig vorkommen. Für eine vererbte Variante gilt also, dass es umso wahrscheinlicher wird, sie in mehr als einem Genom zu finden, je mehr Genome wir gleichzeitig nach Varianten durchsuchen. Je mehr Daten (Reads) eine Variante wiederum beschreiben, umso sicherer können wir uns sein, dass es sich tatsächlich um eine Variante und keine Sequenzierungsfehler handelt. Varianten können durch eine Skalierung auf viele Genome also mit größerer Sicherheit und damit umfassender erfasst werden. Die Skalierung setzen wir um, indem wir die Read-Sequenzen in anderen Datenstrukturen repräsentieren als die Read-Sequenzen einzelner Genome, denn es muss ja zusätzlich abgebildet werden, welche Read-Sequenz aus welchem Genom stammt. Auf diesen Datenstrukturen formulieren wir wiederum andere Optimierungsprobleme, so dass schlussendlich neue Detektions- und Genotypisierungsansätze entstehen (Abb. 2).

Je nach Art der genomischen Varianten bieten sich andere Möglichkeiten, sie zu detektieren und genotypisieren, und entsprechend unterschiedlich sehen auch die Lösungsansätze aus. Beispielsweise erfordert die Beschreibung großer Insertionsvarianten, bei denen zusätzliche, unbekannt Sequenz beobachtet wird, dass diese zusätzliche Sequenz aus den kurzen Read-Sequenzen zusammengesetzt wird. Eine Deletionsvariante, bei der ein Sequenzabschnitt fehlt, ist dagegen durch eine Anfangs- und Endposition in einem Referenzgenom vollständig beschrieben.

Auch die Sequenzierungstechnologie, mit der die zugrundeliegenden Daten erzeugt werden, spielt eine große Rolle für die Ausgestaltung eines neuen Ansatzes. Technologien, die kurze Reads von 150 Zeichen auslesen, haben niedrige Fehlerraten und die Fehler, die sie machen, sind größtenteils Substitutionen. Diese Daten sind in vielen Projekten aus Kostengründen die erste Wahl und sind entsprechend in gro-



2 Die Suche nach genomischen Insertionsvarianten formuliert als das gewichtete Minimum Path Cover Optimierungsproblem.

Bildnachweis: Birte Kehr

ßer Zahl verfügbar. Technologien, die längere Reads von 10.000 bis über 100.000 Zeichen an einem Stück auslesen können, haben noch deutlich höhere Fehlerraten, wobei die meisten Fehler kurze Insertionen und Deletionen sind. Ein Großteil der heute bekannten Strukturvarianten konnte erst mit solch langen Reads detektiert werden und auch die eingangs genannten Telomer- und Centromerregionen wurden mit diesen Daten erstmals im Jahr 2022 zugänglich (Nurk et al. 2022).

Fest steht, dass all diese Variation in einer linearen DNA-Sequenz nicht abgebildet werden kann. Im Forschungsfeld wird seit über 10 Jahren diskutiert, welche Darstellung des menschlichen Genoms einschließlich der in der Bevölkerung existierenden Variation am geeignetsten ist. Sich verzweigende Graphdatenstrukturen scheinen unabdingbar, konnten sich aber in der Breite noch nicht durchsetzen. Ein Grund dafür mag sein, dass jegliche Ansätze zur Analyse von Sequenzierungsdaten auf ei-

nem linearen Referenzgenom basieren und für die neue Datenstruktur angepasst oder sogar neu entwickelt werden müssen. Der potenzielle Nutzen solch eines menschlichen »Pangenoms« ist jedoch unumstritten (Liao et al. 2023).

Zuletzt sei angemerkt, dass auch ein Pangenom immer unvollständig bleibt, weil es nur die bekannte Variation enthalten kann. Mit der Geburt eines Kindes gelangt jedoch neue Variation in die Bevölkerung. Die Anzahl neuer Mutationen pro Kind schätzen wir mit durchschnittlich etwa 120 zwar sehr gering (Jonsson et al. 2017, Kristmundsdottir et al. 2023), sie sind jedoch besonders für die Diagnose seltener Krankheiten von großer Bedeutung. Und auch die Veränderungen der Genomkopien in all unseren Zellen wird ein Pangenom nicht abbilden können. Bei jeder Zellteilung können Kopierfehler im Genom passieren, die sich über das Leben eines Menschen anreichern und nicht an Kinder vererbt werden. Im schlimmsten

Fall führen sie zu unkontrollierter Zellteilung und Krebs. Aber auch gesunde Zellen verändern sich mit der Zeit, angefangen mit stark expandierenden Immunzellen bis hin zu langlebigen Nervenzellen. Die unterschiedlichen Genome des Menschen weiter zu erforschen und somit Grundlagen für die Medizin zu schaffen, bleibt die Aufgabe der Bioinformatik in der Zukunft.

Referenzen

National Human Genome Research Institute (NHGRI). **Fact Sheet: Human Genomic Variation**. Retrieved October 6, 2023, from <https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-variation>

Audano PA, Sulovari A, Graves-Lindsay TA, ..., Eichler EE. **Characterizing the major structural variant alleles of the human genome**. *Cell* 2019; 176(3):663-675.e19

Krannich T, White WTJ, Niehus S, ..., Kehr B. **Population-scale detection of non-reference sequence variants using colored de Bruijn graphs**. *Bioinformatics* 2022; 38(3):604-611

Niehus S, Jonsson H, Schönberger J, ..., Kehr B. **PopDel identifies medium-size deletions simultaneously in tens of thousands of genomes**. *Nature Communications* 2021; 12:730

Nurk S, Koren S, Rhie A, ..., Phillippy AM. **The complete sequence of a human genome**. *Science* 2022; 376(6588):44-53

Liao WW, Asri M, Ebler J, ..., Paten B. **A draft human pangenome reference**. *Nature* 2023; 617: 312-324

Jonsson H, Sulem P, Kehr B, ..., Stefansson K. **Parental influence on human germline de novo mutations in 1,548 trios from Iceland**. *Nature* 2017; 549(7673):519-522

Kristmundsdottir S, Jonsson H, Harderson MT, ..., Stefansson K. **Sequence variants affecting the genome-wide rate of germline microsatellite mutations**. *Nature Communications* 2023; 14:3855

DREI FRAGEN ...

Was möchten Sie in den nächsten Jahren entwickeln?

Bei einer Immunreaktion teilen sich bestimmte Immunzellen sehr häufig und bilden sogenannte »Klone«. Ich möchte neue Ansätze entwickeln, die es erlauben, die Rolle von genomischen Veränderungen in Immunzellklonen zu untersuchen.

Die Lösung von welchem fachlichen Thema haben Sie sich leichter vorgestellt?

Bisher alle. Jede neue Aufgabe in der Forschung birgt unerwartete Wendungen. Es scheint eher die Ausnahme, dass die Lösung eines Themas so ausfällt, wie man sie sich zu Beginn vorstellt. Zeitpläne multipliziere ich daher prinzipiell mal drei.

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

Ein persönlicher Wunsch: Die Variation im Kringle-Repeat des LPA-Gens positionsgenau auflösen, dessen direkte Rolle für das genetisch bedingte Risiko für Herz-Kreislauf-Erkrankungen ungeklärt ist.



Foto © Universität Regensburg

Prof. Dr. Birte Kehr

leitet seit 2020 die Forschungsgruppe für Algorithmische Bioinformatik am Leibniz-Institut für Immuntherapie und hat eine W2-Professur an der FIDS inne. Zuvor baute sie eine Nachwuchsgruppe am Berlin Institute of Health und der Charité-Universitätsmedizin Berlin auf, nachdem sie drei Jahre lang bei deCODE generics/Amgen Inc. mit Sitz in Reykjavik, Island, gearbeitet hatte. Sie promovierte am Fachbereich Mathematik und Informatik der Freien Universität Berlin und erhielt ihr Diplom in Bioinformatik von der Friedrich-Schiller-Universität Jena.

Algorithmen zum Entschlüsseln der Genregulation

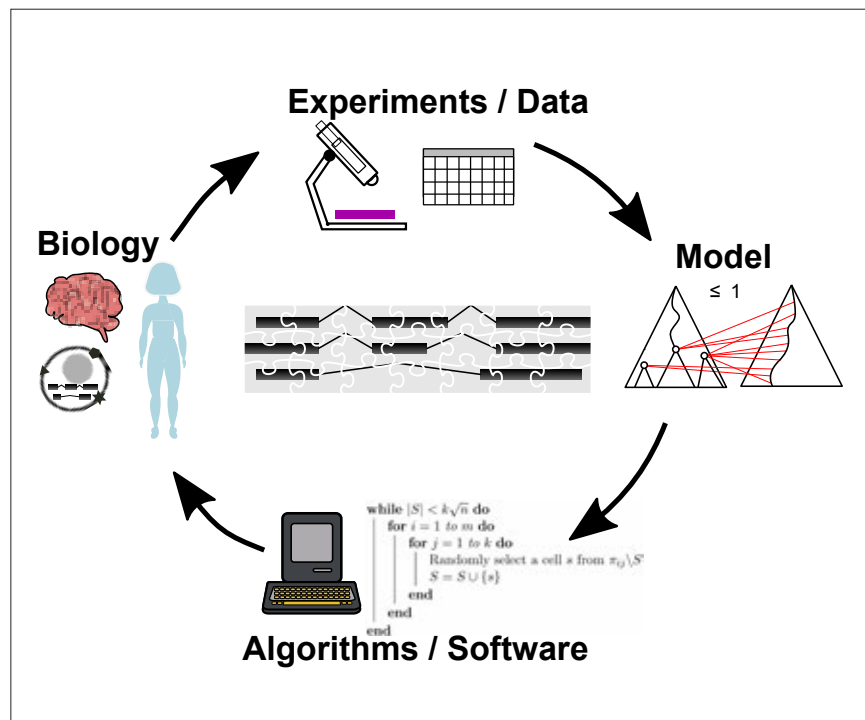
Dr. Francisca Rojas Ringeling, Prof. Dr. Stefan Canzar

Die Hochdurchsatz-Sequenzierung erzeugt riesige Mengen fragmentierter genetischer Daten. Unser Labor verwendet fortschrittliche Algorithmen, um diese Informationen wieder zusammensetzen und versteckte Muster zu erkennen.

Höhere Organismen speichern ihr genetisches Material als Desoxyribonukleinsäure (DNA) in den Zellkernen. In einem Prozess namens Transkription werden einzelne Abschnitte, die Gene, in Boten-Ribonukleinsäuren (mRNAs) umgewandelt. Anschließend entstehen durch den Translationsprozess Proteine als wichtigste Funktionseinheiten.

Die genaue Rekonstruktion der grundlegenden Proteinmuster in den genetischen Bauplänen des Lebens kann uns helfen, die Zellbiologie zu verstehen und spezifische Muster in Genen zu identifizieren, die Krankheiten hervorrufen. Die bevorzugte Methode zur Umwandlung von Proben zellulärer RNA und anderer zellulärer Merkmale in genetische Daten ist die (RNA-)Sequenzierung, die Hunderte Millionen kleiner Datenfragmente, sogenannte Reads, erzeugt. In unserem Labor entwickeln wir Algorithmen und Open-Source-Software, die genomische Reads zu molekularen Messungen und komplexeren Zellmustern zusammenfügt, um neue Einblicke in fundamentale Probleme der Biologie und menschliche Krankheiten zu gewinnen.

Für die Entwicklung präziser und effizienter Rechenmethoden kombinieren wir Techniken aus der sogenannten »large-scale« Optimierung, der mathematischen Programmierung, und des maschinellen Lernens. Beispielsweise haben wir (i) exakte und approximative Algorithmen für die Färbung von Graphen entwickelt, um die Auflösung experimenteller Proteinstrukturdaten zu erhöhen, (ii) neuronale Netze

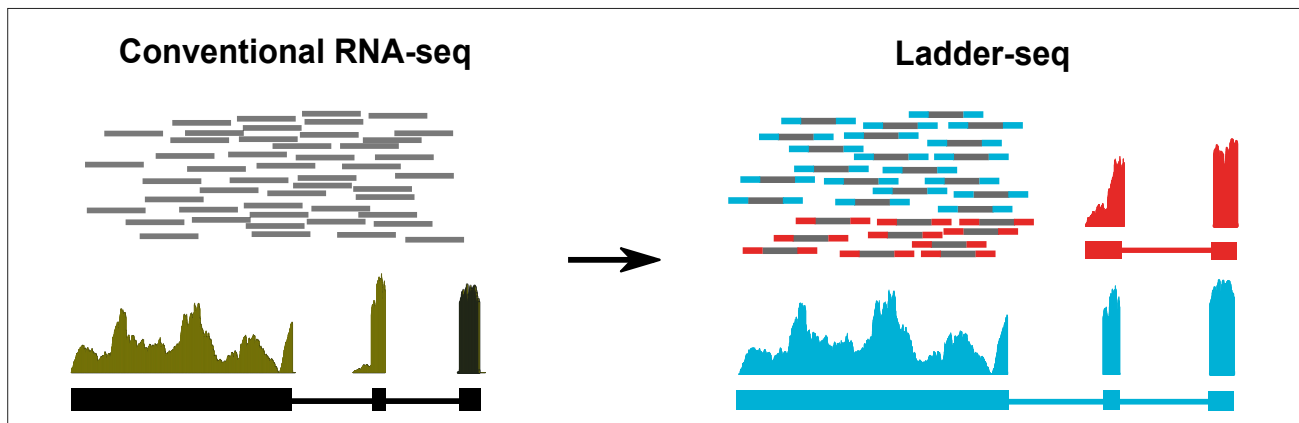


1 Wir arbeiten eng mit Biologen und Medizinern zusammen, um von einer konkreten biologischen Probleminstanz zu einer formalen Problemklasse zu abstrahieren, deren Lösungen für ein breites Spektrum biomedizinischer Fragestellungen nützlich sind. Basierend auf dem extrahierten Modell entwickeln und analysieren wir Algorithmen, die effizient sind in Theorie und Praxis, und die wir in flexible und benutzerfreundliche Softwaretools umsetzen.

verwendet, um hochdimensionale Einzelzell-Messungen in einen interpretierbaren niedrigdimensionalen Raum zu projizieren, und (iii) Dynamic-Time-Warping, das u. a. in der Spracherkennung zur Anwendung kommt, für den Vergleich von komplexen Differenzierungs-Trajektorien erweitert.

Wir übersetzen die entwickelten Algorithmen in hochwertige Softwarelösungen und wenden sie in enger Zusammenarbeit mit Biologen und Medizinern an, um neue Erkenntnisse, z. B. in der Genregulation, zu gewinnen (Fig. 1). In einer Kollaboration

mit dem Labor von Dr. Song (University of Pennsylvania) haben wir beispielsweise zur Entdeckung (i) des embryonalen Ursprungs adulter neuronaler Vorläuferzellen sowie (ii) der zeitlichen Regulation der Gehirnentwicklung von Mäusen durch chemische Modifikationen der einzelnen mRNA-Bausteine beigetragen. Dr. Feuchtinger von der Kinderklinik der LMU München haben wir dabei geholfen, die Expression des TIM-3-Gens mit einem erhöhten Rückfallrisiko bei Kindern mit Leukämie (ALL) in Verbindung zu bringen.



2 In Ladder-seq wird den Sequenzfragmenten Farbinformation hinzugefügt, die es unseren maßgeschneiderten Algorithmen erlaubt Transkripte zu identifizieren, die für konventionelles RNA-Seq unsichtbar sind.

Algorithmen für die Rekonstruktion des Transkriptoms

Das Zusammensetzen von vielen Millionen von RNA-Fragmenten aus der RNA-Sequenzierung (RNA-Seq) zu vollständigen molekularen Sequenzen, den sogenannten Transkripten, ähnelt dem Lösen eines Puzzles und ist sehr schwierig. Wir haben dafür die Methode *CIDANE* entwickelt, die kombinatorische Optimierung mit maschinellem Lernen vereint, um Transkripte präziser als existierende Methoden zu rekonstruieren. Anstatt zu versuchen, die Ergebnisse der genetischen Transkription rechnerisch zu rekonstruieren, haben wir in der Methode *McSplicer* ein vereinfachtes probabilistisches Modell des zugrunde liegenden Prozesses entwickelt. Das Modell basiert auf der Nutzung elementarer Bausteine von Transkripten, die in *McSplicer* mit einem iterativen Optimierungsalgorithmus effizient quantifiziert wird. Damit lassen sich vielfältige Folgen einzelner Mutationen effektiv beschreiben, wie wir in Experimenten mit RNA-Seq Daten von Patienten mit Autismus-Spektrum-Störungen veranschaulicht haben.

Durch die experimentelle Fragmentierung der RNA geht jedoch ein Informationsverlust einher, der durch rechnergestützte Methoden allein nicht wiederhergestellt werden kann. Unser Team hat deshalb Ladder-seq entwickelt, das Änderungen am RNA-Seq Protokoll mit maßgeschneiderten Algorithmen kombiniert, um Transkripte zu identifizieren, die für konventionelle RNA-Seq Methoden unsichtbar sind (Fig. 2). Bildlich gesprochen fügt ein zusätzlicher Schritt im Experiment den genetischen Puzzleteilen Farbinformationen hinzu, die von unseren Algorithmen genutzt werden, um die Puzzlestücke präziser

als bisher zusammensetzen. Mit Ladder-seq konnten wir so die Funktion regulatorischer Einheiten neuronaler Stammzellen im Gehirn von Mäusen entschlüsseln.

Effiziente Algorithmen für Einzelzell-Omics

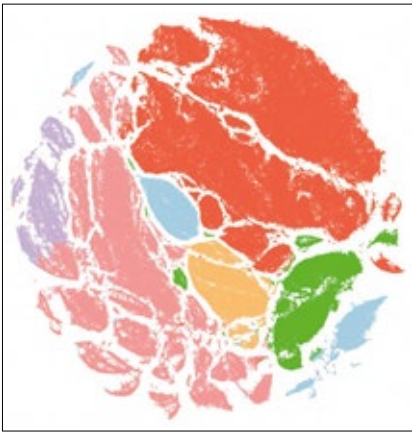
Experimentelle Methoden für die Sequenzierung von RNA in einzelnen Zellen haben die Forschung in der Biologie und der Medizin revolutioniert. Um das volle Potenzial der Einzelzell-Genomik auszuschöpfen, entwickeln wir Algorithmen, die die Lücke zwischen der Skalierbarkeit aktueller Analysemethoden und der schier Menge der erzeugten Daten schließen können. Gleichzeitig verknüpfen die von uns entwickelten Berechnungsmethoden mehrere durch neue Technologien gemessene Modalitäten, darunter Genom, Transkriptom, Epigenom und Proteom, um Zelltypen, z. B. im Kontext des Immunsystems, zu verfeinern.

Während heute die Genexpression von vielen Millionen Zellen gleichzeitig experimentell gemessen werden kann, stößt die rechnergestützte Interpretation der erzeugten Daten, um z. B. neuartige Zelltypen zu identifizieren, bei der Verwendung von existierenden Algorithmen schnell an ihre Grenzen. In der Einzelzell-Genomik kommen deshalb häufig vereinfachte Methoden zur Anwendung, die komplexe biologische Zusammenhänge nicht erfassen können. Im Gegensatz zu diesem »Mehr-Daten-weniger-Methoden«-Trend verfolgen wir in unserer Forschung einen »Relevante-Daten-mehr-Methoden«-Ansatz. So berechnet unsere Methode *Sphetcher* eine dünnbesetzte Repräsentation der Daten, die auf der einen Seite

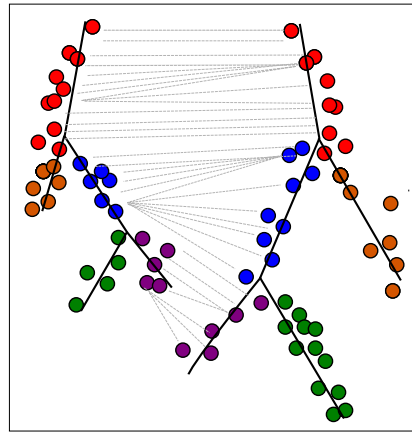
den experimentell erfassten molekularen Raum akkurat abdeckt, aber gleichzeitig die Analyse mit komplexen Algorithmen ermöglicht. Ein solcher sogenannter Einzelzell-»Sketch« kann seltene Zelltypen zum Vorschein bringen, wie wir mit der Identifikation einer seltenen Population von inflammatorischen Makrophagen in Nabelschnurblutzellen demonstriert haben. Ein Sketch erleichtert darüber hinaus die Visualisierung und den Austausch großer Datensätze, und beschleunigt aufwendige Analysen wie die Rekonstruktion von Differenzierungs-Trajektorien. Ein effizienter Algorithmus ermöglicht es *Sphetcher*, einen Sketch aus mehr als zwei Millionen Zellen von Mausembryonen in nur 16 min zu berechnen.

Eine weitere grundlegende Aufgabe in der Analyse der RNA-Expression in Einzelzellen ist die Identifizierung transkriptionell unterschiedlicher Zellgruppen (Fig. 3). Eine algorithmische Innovation ermöglicht es unserem Softwaretool *Spectre*, einen Datensatz mit zwei Millionen Zellen in weniger als einer halben Stunde zu analysieren. Der in *Spectre* implementierte Ensemble-Lernansatz ist außerdem in der Lage, multimodale Omics-Messungen, wie die Expression von mRNA und Oberflächenproteinen, zu kombinieren, um feinste transkriptomische Unterschiede zwischen Subpopulationen von z. B. Gedächtnis-T-Zellen aufzulösen.

Die Visualisierung von solchen multimodalen Omics-Daten erfolgt meist mittels einer t-SNE oder UMAP Einbettung (Fig. 3) der einzelnen Modalitäten, die anschließend manuell zu einem Gesamtbild zusammengesetzt werden müssen. Wir haben die beiden Algorithmen so erweitert, dass die berechneten Einbettungen mehrere Modalitäten gleichzeitig berücksichtigen



3 t-SNE Einbettung verschiedener Zelltypen des Nervensystems der Maus.



4 Trajan match einzelne Zellen zwischen biologischen Prozessen, um Veränderungen der Genexpression entlang komplexer Differenzierungs-Trajektorien zu vergleichen.

und so die visuelle Interpretation von multimodalen Omics-Daten deutlich vereinfachen. Unsere beiden Algorithmen *j-SNE* und *j-UMAP* sind im Softwarepaket *JVis* frei verfügbar und wurden bereits mehr als 50,000-mal heruntergeladen.

Die Einzelzellanalyse erlaubt darüber hinaus die Rekonstruktion von Trajektorien, die dynamische Veränderungen der RNA-


Expression während eines biologischen Prozesses beschreiben. Der Vergleich von Differenzierungs-Trajektorien von Immunzellen, die verantwortlich sind für unterschiedlich erfolgreiche Immunantworten gegen eine Infektion, kann z. B. Aufschluss geben über den zugrunde liegenden molekularen Regulationsmechanismus, und so die Behandlung von Infektionserkran-

kungen verbessern. Wir haben dazu die Methode *Trajan* entwickelt, die es erstmals ermöglicht, komplexe Differenzierungs-Trajektorien systematisch zu vergleichen (Fig. 4). *Trajan* löst dazu eine Variante des »Constraint-Matching«-Problems, das wir theoretisch mit Dynamic-Time-Warping (DTW) verknüpft haben. DTW wird für den Vergleich von zeitlichen Sequenzen, z. B. in der Spracherkennung, verwendet. In einem Vergleich einzelner Zelltrajektorien, die die Differenzierung menschlicher Muskeln und die myogene Reprogrammierung menschlicher Fibroblasten beschreiben, identifizierte *Trajan* die Korrespondenz zwischen Kernprozessen der beiden Trajektorien, anhand derer wir seit kurzem bekannte Hindernisse der myogenen Reprogrammierung reproduzieren konnten.

Wir werden auch in Zukunft Algorithmen und Open-Source-Software entwickeln, die dazu beitragen können, die vielen Teile des größeren Puzzles der Genregulationsnetzwerke in den Zellen zusammensetzen, und dadurch unser Verständnis der zellulären Identität und Funktion, im gesunden Körper und bei Krankheiten, verbessern.



Dr. Francisca Rojas Ringeling schloss 2009 ihr Studium der Medizin an der Universität von Chile in Santiago Chile ab. Nach Abschluss ihrer Facharztausbildung in Medizinischer Genetik am Klinischen Krankenhaus der Universität von Chile im Jahr 2013 zog sie in die Vereinigten Staaten und schrieb sich an der Johns Hopkins University School of Medicine für ihre Promotion in Humangenetik ein. Ihre Doktorarbeit im Labor von Dr. Hongjun Song und Dr. Guo-li Ming konzentrierte sich auf die posttranskriptionelle Regulierung der embryonalen Neurogenese mit einem Schwerpunkt auf Genomik und Bioinformatik. Im Jahr 2019 wechselte sie als Postdoktorandin in das Labor von Dr. Stefan Canzar am Genzentrum der Ludwig-Maximilians-Universität, bis sie 2023 als Research Assistant Professor an die Penn State University kam.



Prof. Dr. Stefan Canzar studierte Informatik in München und war von 2005 bis 2008 am Max-Planck-Institut für Informatik, Abteilung für Algorithmen und Komplexität, tätig. Er erhielt seinen binationalen Dokortitel in Informatik von der Universität des Saarlandes (Deutschland) und der Université Henri Poincaré (Frankreich) im Jahr 2008. Nach seiner Postdoc-Phase am Centrum Wiskunde & Informatica (Amsterdam) und am Johns Hopkins Institute of Genetic Medicine (Baltimore) war er ab 2014 als Research Assistant Professor am Toyota Technological Institute in Chicago tätig. Ab 2016 leitete er eine unabhängige Forschungsgruppe am Genzentrum der LMU München, bis er 2023 als Associate Professor an die Penn State University wechselte. Seit Oktober 2023 ist er Inhaber des Lehrstuhls für Algorithmische Bioinformatik an der Universität Regensburg.



Mit Hilfe von Daten Immunprozesse entschlüsseln

Der Lehrstuhl Computational Immunology

Prof. Dr. Florian Erhard

Unser Immunsystem ist zweifellos eines der erstaunlichsten Wunderwerke der Natur. Es ist ein hochkomplexes Netzwerk aus Zellen, Proteinen, Organen und Geweben, das ständig daran arbeitet, uns vor einer schier endlosen Vielfalt von Krankheitserregern und Gefahren zu schützen. Doch wie schafft es das Immunsystem, diese komplexe Aufgabe zu bewältigen, ohne uns selbst zu schaden? Zu verstehen, wie diese Balance zwischen der Abwehr von Krankheiten und der Vermeidung einer übermäßigen Reaktion gegen den eigenen Körper funktioniert, ist eine der faszinierendsten Herausforderungen in der heutigen Biomedizin.

So genannte »omics« Technologien haben in den letzten Jahren einen atemberaubenden Fortschritt erlebt. Diese Technologien können zwar einen riesigen Berg an Daten über alle möglichen biologischen Vorgänge generieren, die Art und Weise wie aus den Daten Informationen gewonnen werden müssen stellt uns aber vor große Herausforderungen: Stellen Sie sich ein Puzzle mit Millionen von Teilen vor. Sie kennen das Motiv nicht, das das fertige Puzzle zeigen wird. Zudem fehlen viele Teile, viele Teile sind leider identisch und oft sind Teile schwer beschädigt. Ein solches Puzzle gilt es nun zu lösen, denn das fertige Puzzlemotiv würde uns eine neue Erkenntnis über einen Aspekt des Immunsystems liefern.

Unser Forschungsansatz ist multidisziplinär und tief in den Schnittstellen zwischen Biologie, Informatik und Statistik verwurzelt. Wir entwickeln innovative computergestützte Methoden, um spezielle Arten von solchen Puzzles lösbar zu machen,

d.h. Informationen aus »omics«-Daten zu ziehen und daraus Erkenntnisse zu gewinnen. Doch dies ist nur ein Teil unserer Mission. Unser zentrales Interesse gilt dem Immunsystem und seiner Funktionsweise. Unsere Forschung reicht von der Identifizierung von zellulären Erkennungsmerkmalen, mit denen das Immunsystem kranke von gesunden Zellen unterscheiden kann, bis zur Untersuchung der molekularen und dynamischen Wechselwirkungen zwischen Proteinen und Genen, die bei der primären Immunabwehr eine Rolle spielen.

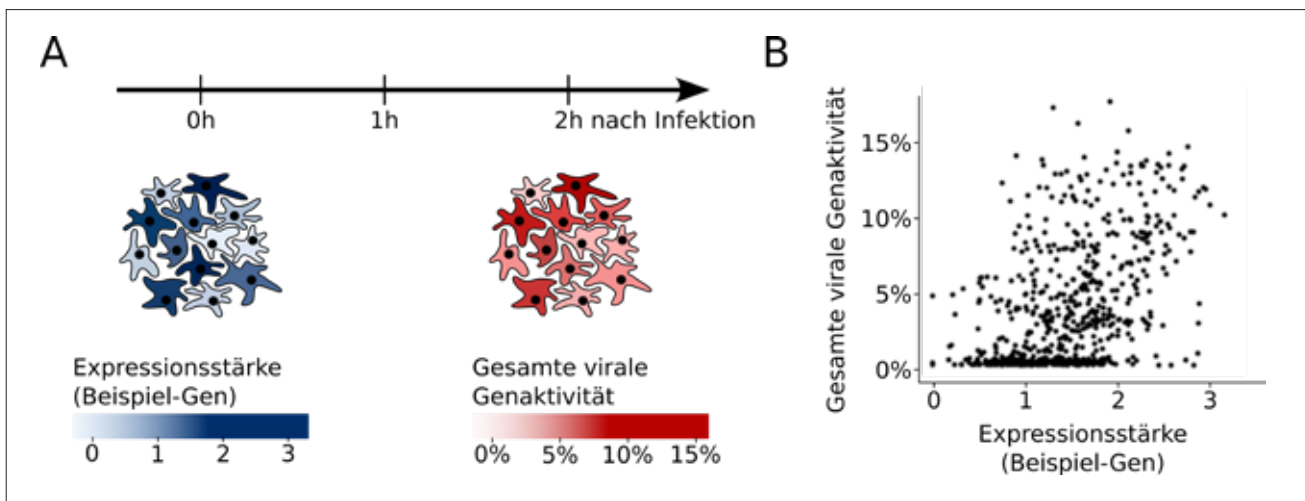
Wie funktioniert das zellintrinsic Immunsystem

Beinahe jede unserer Zellen im Körper hat erstaunliche Fähigkeiten, sich und andere Zellen gegen Krankheitserreger wie Viren oder Bakterien zu verteidigen: Nach Infektion erkennen sie spezielle Muster in diesen Erregern und setzen daraufhin komplexe Abwehrmechanismen in Gang. Dazu gehören Entzündungsreaktionen, das Ausschütten des Botenstoffs »Interferon«, welcher Nachbarzellen alarmiert, und die darauf folgende Produktion von hundert von verschiedenen Effektorproteinen, die auf viele unterschiedliche Weisen Viren und Bakterien abwehren. Zu diesen Effektorproteinen gehören zum Beispiel Enzyme, die virales Genmaterial zerschneiden können, Adhäsionsmoleküle, die die Freisetzung neuer Viruspartikel verhindern, oder Proteine, die zum Zelltod führen.

All diese Prozesse geschehen nicht isoliert – sie interagieren miteinander und mit anderen zellulären Faktoren, was den Grad

der Abwehr bestimmt. Zudem ist das Timing wichtig: Wann welche Prozesse wie stark aktiviert werden, ist entscheidend, ob Krankheitserreger erfolgreich bekämpft werden können oder die Oberhand gewinnen. Für normale Körperfunktionen müssen Zellen außerdem nach erfolgreicher Bekämpfung der Infektion aus diesem alarmierten Zustand in den Normalzustand zurückkehren können. Dieses ganze System ist in seiner Komplexität genau so ausbalanciert, um auf der einen Seite eine rasche und starke Immunantwort hervorzurufen, auf der anderen Seite aber eine andauernde Überreaktion zu verhindern. Viren und Bakterien haben im Lauf ihrer Evolution gelernt, diese komplexen Vorgänge zu unterlaufen, und können dieses System aus der Balance bringen. Man weiß zum Beispiel, dass schwere Verläufe von COVID-19 mit einer verstärkten Entzündungsreaktion einhergehen, aber gleichzeitig die Interferon-Antwort durch das Virus unterdrückt wird.

Um solche zeitlichen Verläufe in einzelnen Zellen erstmals messbar zu machen, haben wir zusammen mit experimentellen Gruppen ein neues Verfahren entwickelt: Dieses Verfahren basiert auf herkömmlicher Einzelzellsequenzierung, mit deren Hilfe die Aktivität tausender Gene in tausenden bis zehntausenden von Zellen festgestellt werden kann. Anders als herkömmliche Einzelzellsequenzierung, die nur Momentaufnahmen liefert und die Zellen dabei zerstört, liefert unser *scSLAM-seq* [2] Verfahren zusätzlich Information über zeitliche Dynamiken. Diese Zusatzinformation äußert sich zum Beispiel darin, dass wir damit den genauen Zustand aller



1 Das Konzept »Heterogeneity sequencing« **A** Bei einem Experiment werden viele Zellen (hier dargestellt: 13 Zellen) zum Zeitpunkt 0 mit Viren infiziert. In einzelnen Zellen weist die Expressionsstärke jedes Gens eine natürliche Heterogenität auf, d. h. das Gen ist unterschiedlich stark aktiv in jeder Zelle (in unterschiedlichen Blautönen dargestellt). Zwei Stunden (2h) später hat das Virus mehr oder weniger erfolgreich in den einzelnen Zellen dafür gesorgt, dass seine eigenen Gene aktiviert wurden (in unterschiedlichen Rottönen dargestellt). Gene, für die die Expression zum Zeitpunkt 0 und der Erfolg der viralen Genexpression korreliert sind, sind Kandidaten für pro- oder antivirale Faktoren: Ihre Expression vor Infektion begünstigt (positive Korrelation) oder vermindert die Infektion (negative Korrelation). Um solche Korrelationen bestimmen zu können, benötigt man Messungen von Expressionsstärken vor **und** nach Infektion. **B** Graphische Darstellung der Expressionsstärke zum Zeitpunkt 0 und der viralen Genaktivität für hunderte von Zellen (Punkte) eines potenziellen pro-viralen Faktors.

Zellen vor **und** nach der Infektion messen können anstatt nur vorher **oder** nachher [2]. Mit dieser Art von Daten sind gänzlich neue Analysekonzepte möglich (Abb. 1).

Der Knackpunkt an scSLAM-seq ist die Datenanalyse [3]. Die Art und Weise, wie die Informationen über die zeitliche Dynamik in den Daten hinterlegt sind, macht es zum Beispiel nötig, nicht nur feste Werte aus den Daten abzuleiten (z. B. »welcher Anteil der gesamten Aktivierung stammt aus der letzten Stunde«), sondern die Mess-Unsicherheit abzubilden (»die Daten belegen, dass dieser Anteil im Bereich 10-30 % liegt«).

Erst diese von uns entwickelten Analysemethoden für scSLAM-seq machen es möglich, die zeitliche Dynamik dieser komplexen Vorgänge der zellintrinsic Immunabwehr vollständig zu erfassen. Aufbauend auf solchen Messungen werden wir Modelle entwickeln, mit dem wir die genauen zeitlichen Abläufe vorhersagen können: Wann werden in welchen Zellen welche Signalwege aktiv, was bewirken im Einzelnen die aktivierten Faktoren und wie beeinflussen sich unterschiedliche Zellen in ihrer Nachbarschaft.

Aufklären der Funktionen und Mechanismen von kryptischen Peptiden

Die meisten Zellen haben neben der zellintrinsic Immunabwehr außerdem die Fähigkeit, einen anderen Teil des Immunsystems in Gang zu setzen: Sie tragen sogenannte humane Leukozytenantigene Klasse I (HLA-I) auf ihrer Zelloberfläche. Diese Strukturen zeichnen sich verantwortlich, Peptide, das sind Fragmente von Proteinen, die in der Zelle hergestellt werden, nach außen zu präsentieren. Spezialisierte Immunzellen namens T-Zellen können unterscheiden, ob es sich bei einem präsentierten Peptid um ein körpereigenes oder fremdes handelt. »Fremd« heißt zum Beispiel, dass das Peptid aus einem viralen Protein stammt, oder aufgrund einer in einer Tumorzelle entstandenen Mutation als nicht mehr körpereigen erkannt wird. Eine T-Zelle, die ein fremdes Peptid erkennt, kann die präsentierende Zelle daraufhin abtöten und so Infektionskrankheiten oder das Entstehen von Tumoren eindämmen oder verhindern.

Mit Hilfe einer Kombination mehrerer »omics«-Technologien und neuen Datenanalysemethoden [4,5] haben wir vor kurzer Zeit entdeckt, dass bis zu 10 % aller präsentierten Peptide nicht etwa Fragmente von bekannten Proteinen, sondern das Produkt von sogenannter *kryptischer*

Translation ist: Bekannte Proteine üben im Allgemeinen wichtige molekulare Funktionen aus und überdauern für mehrere Stunden oder sogar Tage in unseren Zellen bevor sie einem natürlichen Abbauprozess unterliegen. Kryptische Translation hingegen stellt (zum Teil extrem kleine) Produkte her, die unmittelbar nach ihrer Herstellung sofort abgebaut werden. Lange stellte sich die Frage, warum Zellen die Energie dafür aufwenden, kryptische Peptide herzustellen.

Eine Antwort darauf ist, dass sie den T-Zellen als erweiterte Erkennungsmerkmale zur Verfügung stehen. Aus dieser Entdeckung ergeben sich mehrere Fragestellungen, die wir in Zukunft erforschen werden:

- Zusammen mit experimentellen Kollaborationspartnern evaluieren wir den Nutzen von kryptischen Peptiden für immuntherapeutische Ansätze.
- Mit Hilfe weiterentwickelter »omics«-Messungen und unseren Analyse-Methoden werden wir dem Zusammenhang zwischen kryptischer Translation und der Präsentation kryptischer Peptide quantitativ untersuchen
- Wir werden erforschen, warum so viel kryptische Translation stattfindet, obwohl unsere Zellen eine ganze Reihe von Kontrollmechanismen besitzen, die eine erfolgreiche Herstellung von funktionellen Proteinen sicherstellen.

Wir leben zweifellos in aufregenden Zeiten, in denen die Synergie zwischen fortschrittlicher Rechenleistung, innovativen KI-Methoden und modernen Technologien zur Datengewinnung die Immunologie revolutioniert. Diese multidisziplinären Fortschritte ermöglichen es uns, die komplexen Prozesse des Immunsystems tiefer zu durchdringen als je zuvor. Während wir in der Grundlagenforschung immer weiter voranschreiten, ist es ermutigend zu sehen, wie erste Entwicklungen bereits den Weg zu den Patienten finden und einen positiven Einfluss auf die Medizin haben.

Literatur

- [1] Erhard, F. *et al.* scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571**, 419–423 (2019).
- [2] Erhard, F. *et al.* Time-resolved single-cell RNA-seq using metabolic RNA labelling. *Nat Rev Methods Primers* **2**, 1–18 (2022).
- [3] Jürges, C., Dölken, L. & Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34**, i218–i226 (2018).
- [4] Erhard, F. *et al.* Improved Ribo-seq enables identification of cryptic translation events. *Nature Methods* **15**, 363–366 (2018).
- [5] Erhard, F., Dölken, L., Schilling, B. & Schlosser, A. Identification of the Cryptic HLA-I Immunopeptidome. *Cancer Immunol Res* **8**, 1018–1026 (2020).

DREI FRAGEN ...

Welches Fachbuch lesen Sie gerade?

Finite Mixture Models von Geoffrey McLachlan und David Peel. Diese Klasse von Modellen ist unglaublich wichtig für verschiedenste biologische Daten, und dieses Buch bietet einen hervorragenden Überblick über die theoretischen Grundlagen, Algorithmen für Parameterschätzung, Probleme der Identifizierbarkeit von Parametern und illustriert die Anwendung dieser Modelle in verschiedenen Bereichen.

Welche Entwicklung im Bereich Data Science hat Sie im letzten Jahr am meisten überrascht?

Maschinelle Lernverfahren werden seit jeher in der Bioinformatik für verschiedenste Probleme eingesetzt. Auf Konferenzen hat man aus meiner Sicht immer sehr schnell erkannt, ob die oder der Vortragende aus der Ecke »Machine learning« kam und aufregende neue Methoden auf biologische Daten angewendet hat, oder ob der Vortrag aus der Biologie-Ecke kam, d. h. ein spannendes biologisches Problem angegangen wurde und dazu maschinellen Lernen verwendet wurde. Auf der kürz-

lich stattgefundenen ISMB, der wichtigsten Bioinformatik-Konferenz, war schön zu sehen, dass diese beiden Lager nicht (mehr) existieren. Es wurden ausschließlich relevante und hochspannende biologische Fragestellungen präsentiert, die mit modernen und auf die Problemstellung passenden Methoden des maschinellen Lernens gelöst wurden.

Was ist Ihre größte fachliche Herausforderung?

Jede Datenanalyse-Methode von »omics«-Daten trifft gewisse Annahmen. Viele dieser Annahmen sind vereinfachend und damit bei genauer Betrachtung schlichtweg falsch. Bei manchen Annahmen hat das keine oder sehr geringe Auswirkungen auf Ergebnisse, die Verletzung anderer Annahmen kann aber Ergebnisse drastisch verfälschen. Welche Annahmen zu welcher von diesen beiden Kategorien gehören, erfordert oft ein sehr genaues Verständnis der Biologie und der Messmethode sowie der explorativen Analyse vieler Datensätze. Dieses Einschätzen von vereinfachenden Annahmen halte ich für meine größte fachliche Herausforderung.



Foto © Margit Scheid

Prof. Dr. Florian Erhard hat in München Bioinformatik studiert und 2014 dort am Lehrstuhl von Prof. Dr. Ralf Zimmer promoviert. Nach einer kurzen Postdoc-Phase wurde er im Jahr 2016 zum Juniorprofessor für Systemvirologie an der Julius-Maximilians-Universität Würzburg berufen. Dort forschte er an der Genomik von Herpesviren, an kryptischen Translationsmechanismen und der Präsentation von Produkten kryptischer Translation via HLA-I, sowie an der Dynamik von genregulatorischen Prozessen unter akuten Perturbationen von Zellpopulationen oder auch einzelnen Zellen. In seiner Arbeitsgruppe wurden zu diesem Zweck mehrere maßgeschneiderte bioinformatische Methoden neu entwickelt, die auch breite Anwendung außerhalb seiner eigenen Arbeit finden. Florian Erhard ist darüber hinaus Mitglied in mehreren Forschungsverbänden wie dem durch die EU geförderten VIROINF, dem bayerischen Corona-Konsortium FOR-COVID, den DFG Forschungsgruppen DEEP-DV und „Fortschrittliche Konzepte in der zellulären Immunkontrolle von Zytomegalieviren“, und den Sonderforschungsbereichen DECIDE und „Cardioimmune interfaces“. Seit März 2023 ist er nun Inhaber des Lehrstuhls Computational Immunology der Universität Regensburg.

Maschinelles Lernen enthüllt den verborgenen Prozess der Tumorentstehung

Linda Hu, Andreas Lösch, Prof. Dr. Rainer Spang

Künstliche Intelligenz und maschinelles Lernen halten in immer größerem Maße Einzug in diverse Forschungsfelder, darunter auch die Medizin. Ihre Bedeutung erstreckt sich nicht nur auf die Verbesserung von Diagnoseverfahren und Therapieentscheidungen, sondern sie erlauben es auch, bisher völlig undenkbar Einblicke in die Natur von Krankheiten zu gewinnen, wie das folgende Beispiel veranschaulicht.

Tumore entwickeln sich, lange bevor sie klinisch nachweisbar sind. Auf genetischer und molekularer Ebene ereignen sich in dieser Phase bereits für den Krankheitsverlauf entscheidende Weichenstellungen. Die Krebszellen sammeln zahlreiche Mutationen an, adaptieren durch diese interne Signalwege und verändern ihren Stoffwechsel, es kommt zur Neubildung von Blutgefäßen im Gewebe, und häufig tritt eine Entzündungsreaktion auf. All diese Vorgänge ereignen sich an Orten, an denen wir sie klinisch nicht beobachten können. Sie lagen bisher im Dunklen.

Innovative Methoden des maschinellen Lernens bringen diese verborgenen Prozesse nun ans Licht. Wissenschaftler des Regensburg-Erlanger Forschungsverbundes TRR 305, darunter Rainer Spang, Linda Hu und Andreas Lösch vom Lehrstuhl für Statistische Bioinformatik, entwickeln neue Lernverfahren, die in der Lage sind, aus späten klinischen Momentaufnahmen den Verlauf der Krankheit von Beginn an zu rekonstruieren.

Daten zu einem Tumor werden erst nach seiner klinischen Entdeckung durch eine Gewebentnahme erfasst. Zu diesem Zeitpunkt haben bereits zahlreiche genetische Ereignisse stattgefunden, und wir

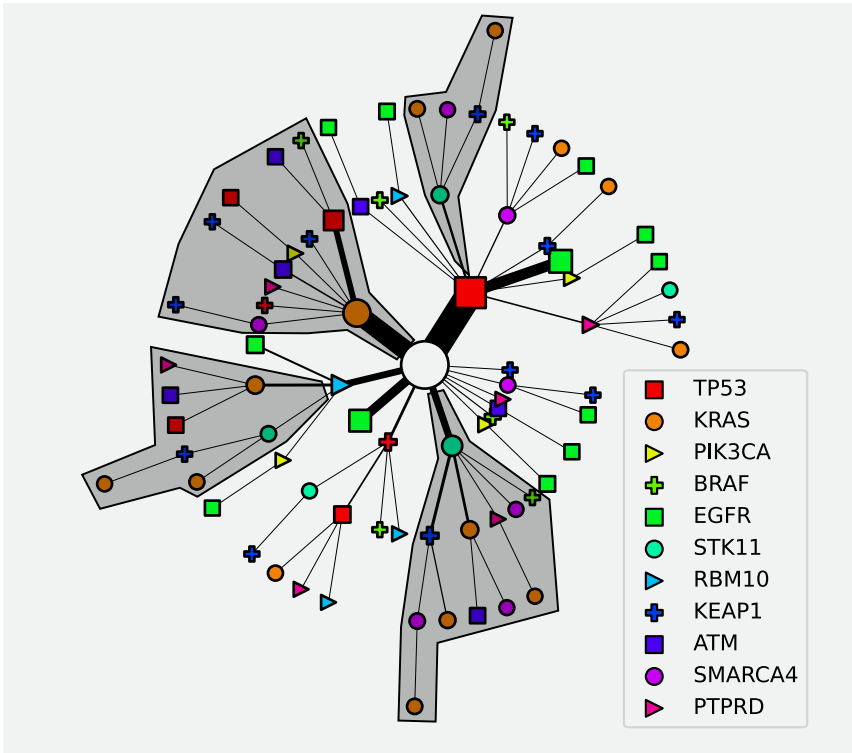
können Tumor für Tumor erfassen, welche Veränderungen eingetreten sind. Die Herausforderung besteht nun darin, aus diesen statischen Momentaufnahmen, in der Regel ist es nur eine pro Patient*in, den dynamischen Prozess ihrer Entstehung nachzuvollziehen.

Um die Methode zu verstehen, ist es hilfreich, sich eine analoge Situation in unserem Alltag vorzustellen, bei der wir ein ähnliches Problem intuitiv bewältigen. Stellen Sie sich also vor, Sie machen einen Stadtspaziergang und kommen an einer Kreuzung an. Dort beobachten Sie, dass sich bereits vier Ereignisse zugetragen haben: Ein Krankenwagen ist eingetroffen, es hat geschneit, es hat einen Unfall gegeben, und die Straße wurde gestreut. Es sollte nicht schwer sein, diese vier Ereignisse in eine sinnvolle zeitliche Reihenfolge zu bringen, da Sie wissen, dass beispielsweise Schnee auf den Straßen die Unfallgefahr erhöhen kann, oder dass ein Krankenwagen oft gerufen wird, wenn sich zuvor ein Unfall ereignet hat. Ähnlich gehen wir bei einem Tumor vor, der drei spezifische Mutationen aufweist, Blutgefäße entwickelt hat und metastasiert ist. Hierbei gibt es jedoch eine zusätzliche Schwierigkeit: Wir sind uns a priori nicht im Klaren darüber, welchen Einfluss jedes dieser fünf Ereignisse auf die Entstehungsrate der anderen vier hatte. Diese bedingten Raten müssen wir aus den Daten destillieren. Wenn ein Ereignis nach seinem Auftreten die Rate eines anderen Ereignisses erhöht, ist es wahrscheinlich, dass beide Ereignisse häufig im selben Tumor auftreten. Im Gegensatz dazu, wenn zwei Ereignisse ihre Raten gegenseitig verringern, treten sie nur selten im selben Tumor auf.

Der Umkehrschluss ist weitaus komplexer und weniger intuitiv. Daher greift man auf Methoden des Maschinellen Lernens zurück, um aus Momentaufnahmen zunächst bedingte Entstehungsraten zu berechnen und anschließend ein dynamisches Modell der Tumorentwicklung zu erstellen. An unserem Lehrstuhl haben wir ein generatives Modell entwickelt, mit dem wir Millionen realistischer Progressionswege virtueller Tumore generieren können, ohne dass tatsächlich jemand erkranken muss. Während Momentaufnahmen dieser KI-generierten Fortschrittswege genauso aussehen wie Mutationsbilder realer Tumore, zeichnen sich die virtuellen Tumore dadurch aus, dass ihre gesamte Entwicklungschronik transparent ist.

In diesen virtuellen Tumorchroniken (Abbildung 1) können wir direkt nachvollziehen, in welcher Reihenfolge Mutationen auftreten, die den Krebs aggressiver machen. In Abbildung 1 wurden solche typische Abfolgen für Patienten errechnet, von denen nur eine Momentaufnahme zur Verfügung stand. Eine typische Abfolge für Lungenkarzinome von Rauchern ist beispielsweise STK11-KEAP1-SMARCA4-KRAS (Abbildung 1 auf 6 Uhr). SMARCA4 kann Metastasenbildung bedeuten, während KRAS für beschleunigte Krebszellteilung steht. Unsere aktuellen Modelle beschränken sich noch ausschließlich auf die Rekonstruktion der Mutationsgeschichte der Tumore, nicht-genetische Ereignisse berücksichtigen wir noch nicht, werden dies aber bald in Angriff nehmen.

Ähnlich wie bei generativen Sprachmodellen, wie ChatGPT, planen wir in der Zukunft unser Modell mit spezifischen



Bildnachweis: Y. Linda Hu, Andreas Lösch

1 Typische Entwicklungsmuster für Mutationen in Lungenkarzinomen, errechnet aus vielen Momentaufnahmen. Jedes farbige Symbol steht für eine bestimmte Mutation. Jeder Pfad, vom gesunden Zustand in der Mitte nach außen, kennzeichnet mit der Pfaddicke wie viele Patienten diesen Weg vermutlich durchlaufen hatten. Für Raucher typische Abfolgen sind mit grau hinterlegt.

Referenzen

Schill, R., Solbrig S., Wettig, T., Spang R., Modelling cancer progression using Mutual Hazard Networks, *Bioinformatics*, Volume 36, Issue 1, January 2020, Pages 241–249, <https://doi.org/10.1093/bioinformatics/btz513>

Klever, M., Georg, P., Grasedyck, L., Schill, R., Spang, R., Wettig, T.: Low-rank tensor methods for markov chains with applications to tumor progression models. *Journal of Mathematical Biology* 86(1) (Dec 2022). <https://doi.org/10.1007/s00285-022-01846-9>

Rupp, K., Schill, R., Süskind, J., Georg, P., Klever, M., Lösch, A., Grasedyck, L., Wettig, T., Spang, R.: Differentiated uniformization: A new method for inferring markov chains on combinatorial state spaces including stochastic epidemic models (2021). <https://doi.org/10.48550/ARXIV.2112.10971>, <https://arxiv.org/abs/2112.10971>

»Prompts« zu starten. So initiieren wir den Entwicklungsprozess nicht bei einem gesunden Individuum, sondern mit einer Momentaufnahme des Tumors eines Patienten. Das generative Modell setzt die Geschichte des Tumors des Patienten virtuell fort. Erneut können wir Millionen möglicher Fortsetzungen generieren und

dann analysieren, welche Gemeinsamkeiten sie aufweisen. Wir sind der Überzeugung, dass dieser virtuelle Einblick in den zukünftigen Verlauf einer Krankheit Ärzten maßgeblich dabei hilft, eine optimale personalisierte Therapiestrategie für ihre Patienten zu entwickeln.

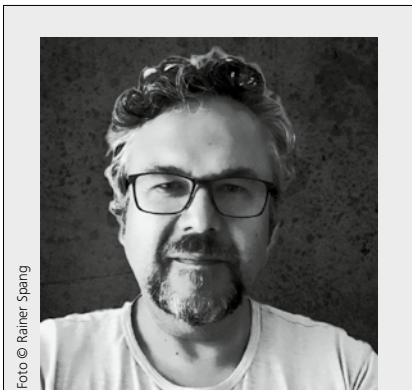


Foto © Rainer Spang

Prof. Dr. Rainer Spang

leitet den Lehrstuhl für statistische Bioinformatik an der Fakultät für Informatik und Data Science. Seine Forschung verbindet statistische Methodenentwicklung mit angewandter Data-Science in der Krebsforschung. Seit über 8 Jahren spielt dabei die Rekonstruktion der Tumorprogression eine wichtige Rolle in der Forschung seines Lehrstuhls.

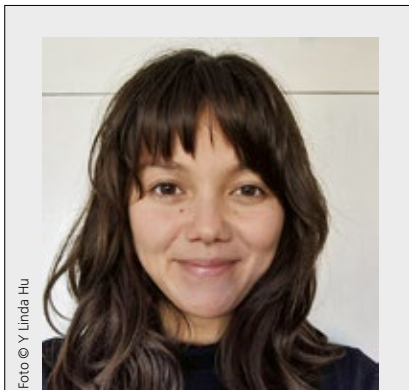


Foto © Y. Linda Hu

Y Linda Hu, M. Sc.,

ist wissenschaftliche Mitarbeiterin am Lehrstuhl für statistische Bioinformatik an der Universität Regensburg. Seit April 2023 promoviert sie bei Prof. Dr. Rainer Spang, nachdem sie zuvor, ebenfalls an der Universität Regensburg, ihr Mathematikstudium mit dem Master abschloss. Sie forscht als Teil des DFG-geförderten transregionalen Sonderforschungsbereichs zu Metastasenentwicklung an der statistischen Modellierung von Tumorprogression.

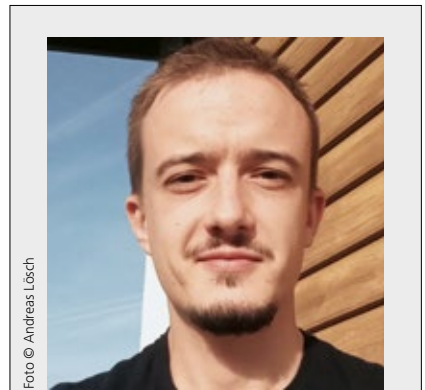


Foto © Andreas Lösch

Andreas Lösch, M. Sc.,

ist wissenschaftlicher Mitarbeiter am Lehrstuhl für statistische Bioinformatik an der Universität Regensburg. Vor der Promotion bei Prof. Dr. Spang schloss er das Masterstudium der Biologie an der Universität Regensburg ab. Er forscht als Teil des DFG-geförderten transregionalen Sonderforschungsbereichs zu Metastasenentwicklung an der statistischen Modellierung von Tumorprogression.

Allgegenwärtige Mensch-Maschine-Interaktion

Entwicklung, Forschung und Infrastruktur der Medieninformatik

Prof. Dr. Niels Henze, Vertr.-Prof. Dr. Johanna Bogon, Dr. Raphael Wimmer, Prof. Dr. Christian Wolff

Einleitung und Entwicklung des Fachs Medieninformatik

Im Umfeld des Internet-Booms rund um das Jahr 2000 war an der Universität Regensburg eine erste Professur für Medieninformatik eingerichtet und ab 2003 besetzt worden (Prof. Wolff). Diese war zunächst im fachlichen Kontext der Informationswissenschaft angesiedelt. In Vorbereitung auf den doppelten Abiturjahrgang in Bayern konnte ab 2006 der Ausbau des Fachs sowie die Einrichtung eines Medieninformatik-Studiengangs als Teilfach im kombinatorischen Bachelor der philosophischen Fakultäten geplant und umgesetzt werden (Hammwöhner et al., 2012). Zusammen mit dem 2010 gestarteten Studiengang und dem ab 2013 etablierten Masterstudiengang (Master of Science Medieninformatik) waren zwischenzeitlich bis zu 700 Studierende im Fach Medieninformatik immatrikuliert. Im Zuge des Ausbaus konnte eine zweite Professur für Medieninformatik eingerichtet und besetzt werden (Prof. Henze). Beide Professuren sind Teil des Instituts für Information und Medien, Sprache und Kultur, in dem Querschnittfächer wie allgemeine und vergleichende Sprachwissenschaft, Informationswissenschaft, Medieninformatik, Medienwissenschaft und vergleichende Kulturwissenschaft angesiedelt sind. Auch nach dem Wechsel der Professuren für Medieninformatik in die Fakultät für Informatik und Data Science im Jahr 2022 wird diese interdisziplinäre und interfakultäre Institutsstruktur weitergeführt.

Interaktion mit allgegenwärtigen Computern

Computer sind im Alltag der meisten Menschen heute unverzichtbar geworden. Sie haben sich zu essentiellen Werkzeugen entwickelt, die in nahezu allen Lebensbereichen eine herausragende Rolle spielen. Ob in beruflichen, Bildungs-, kommunikativen oder Freizeitkontexten - Computer und die damit verbundenen Technologien prägen unser tägliches Leben in vielfältiger Weise. Die kontinuierliche Steigerung der Rechenleistung von Computern sowie die Weiterentwicklung der darauf ausgeführten Software haben in der Informatik zu einem Paradigmenwechsel geführt. Heute steht nicht mehr die Frage im Vordergrund, welche Aufgaben Computer übernehmen können, sondern vielmehr, wie Computer Menschen bestmöglich bei der Bewältigung von Aufgaben unterstützen können. Ein zentraler Fokus liegt dabei darauf, Menschen dabei zu helfen, ihre Ziele mithilfe von Computern auf effektive, effiziente und zufriedenstellende Weise zu erreichen.

Der Lehrstuhl für Medieninformatik an der Universität Regensburg, die Arbeitsgruppe für Physisch-Digitale Affordanz (Dr. Wimmer) sowie der durch Dr. Johanna Bogon vertretene Lehrstuhl für Mensch-Maschine-Interaktion erforschen die Interaktion zwischen Menschen und Computern. Der Forschungsfokus liegt auf der Gestaltung und Verbesserung der Interaktion zwischen Menschen und Technologie, sowohl in der heutigen Zeit als auch für die Zukunft. Die Zielsetzung besteht darin, die

Interaktion mit Computern zu verstehen und zu verbessern, die digitale Transformation von Arbeitsabläufen zu unterstützen und die Verwendung von computergestützten Verfahren in anderen Disziplinen wie den Geistes- und Sozialwissenschaften zu erforschen.

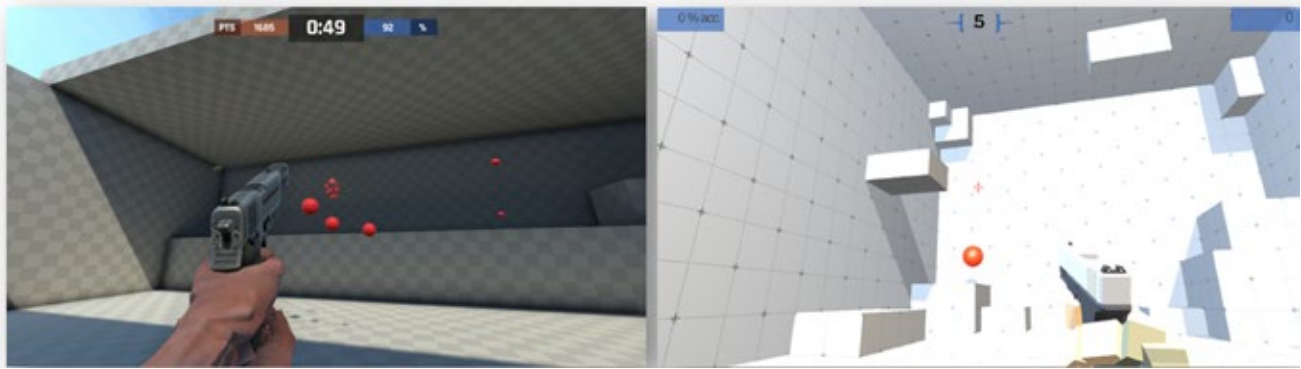
Zeitliche Aspekte der Mensch-Computer-Interaktion

Zeit spielt eine grundlegende Rolle bei der Interaktion mit Computern. Die Art und Weise, wie wir zeitliche Aspekte interaktiver Systeme wahrnehmen und in unsere Handlungen integrieren, beeinflusst, wie effizient und effektiv wir Aufgaben erledigen, sei es in virtuellen Umgebungen, bei der Nutzung von digitalen Werkzeugen oder beim Spielen von Videospielen. Diese zeitlichen Aspekte sind von großer praktischer Relevanz und haben Auswirkungen auf die Benutzerfreundlichkeit und die Gestaltung von Technologien. Unsere Forschung zielt darauf ab, die Zusammenhänge zwischen Zeit, Handlung und Wahrnehmung zu verstehen und Lösungen zu entwickeln, die die Mensch-Computer-Interaktion verbessern und die Nutzung digitaler Technologien noch nahtloser gestalten.

Um zu verstehen, wie sich zeitliche Aspekte interaktiver Systeme auf die Interaktion zwischen Menschen und Computern auswirken, untersuchen wir, wie sich Erkenntnisse aus der kognitiven Psychologie auf die Mensch-Computer-Interaktion übertragen lassen. Wir wissen aus der Ko-



Bildnachweis David Halhuber



1 Die linke Seite zeigt einen Screenshot aus dem Zieltrainer-Spiel Aim Lab. Die rechte Seite zeigt einen Screenshot eines für eine Studie entwickelten Spiels mit eingebauten Vorperioden-Ereignis-Paarungen. Beide Screenshots zeigen die Perspektive des Spielers, die virtuelle Waffe des Spielers und Spielziele.

gnitionsforschung, dass Menschen anhand vergangener Zeit unterschiedliche Ereignisse antizipieren. Wenn wir beispielsweise ein Dokument an den Drucker senden, so erwarten wir, dass wir bald hören, wie das Dokument gedruckt wird. Wenn dieses Geräusch nach einer bestimmten Zeit noch nicht aufgetreten ist, verändert sich unsere Erwartung. Wir nehmen nun an, dass ein Problem vorliegt und wir möglicherweise Papier nachfüllen müssen. Unsere Arbeiten zeigen, dass wir solche Verknüpfungen zwischen bestimmten Vorperioden (die Zeit, die vor Eintreten des Ereignisses vergeht) und bestimmten Ereignissen gezielt nutzen können, um Interaktionen mit Computern nahtloser zu gestalten. So konnten wir zeigen, dass Menschen in

einem Videospiel durch systematisch eingebaute Vorperioden-Ereignis-Paarungen implizit schneller auf Ereignisse reagieren können, ohne sich dieser zeitlichen Regelmäßigkeiten explizit bewusst zu werden (Halhuber et al., 2023).

Des Weiteren untersuchen wir, wie sich die Interaktion mit Computern auf die Zeitwahrnehmung auswirkt. Hierzu haben wir zum Beispiel die Ergebnisse von Zeitproduktionsaufgaben in der realen Welt mit den Ergebnissen von Zeitproduktionsaufgaben in virtueller Realität verglichen. Dadurch konnten wir zeigen, dass virtuelle Realität nicht unsere Zeitwahrnehmung verändert, aber einen Einfluss auf unsere Erwartungen bezüglich der Dauer von physikalischen Prozessen hat (Bogon et al.,

2023). Zudem erforschen wir die Effekte von Latenz – die Zeit, die zwischen einer Eingabe der Nutzenden und der Reaktion des Computers verstreicht. Wir konnten zeigen, dass unabhängig von Ausgabemodalität und Kontext Latenz immer negative Effekte auf die Nutzenden hat.

Verkörperung von Avataren in virtuellen Realitäten

Die Interaktion in virtuellen Welten, wie sie in Virtual Reality (VR) erlebt wird, stellt einen weiteren Schwerpunkt unserer Forschungsarbeit dar. Ein Aspekt, den wir erforschen, sind Effekte, die durch die Repräsentation des Benutzers verursacht werden. Die Re-



Bildnachweis Dr. Martin Kocur

2 Die reale Welt (links) und die virtuelle Szene bestehend aus nichtsportlichen, durchschnittlichen und sportlichen Avataren (von links nach rechts, weibliche oben, männliche unten) auf einem stationären Fahrrad.

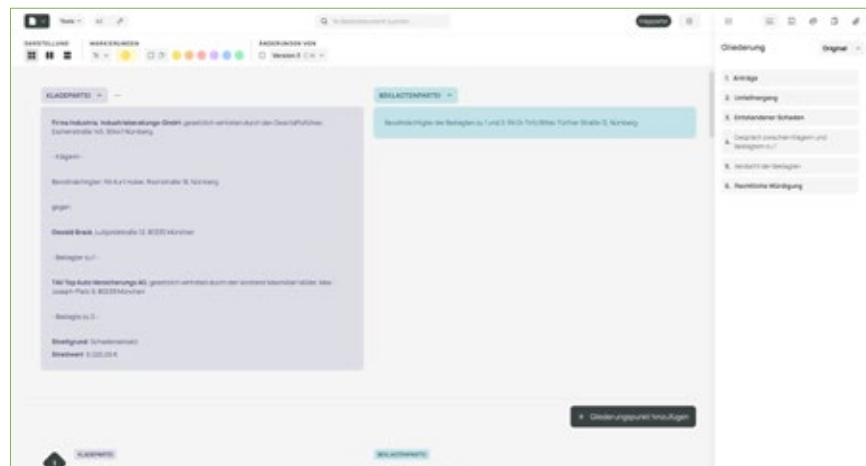
präsentation des Benutzers in virtuellen Umgebungen wird, nach dem Sanskrit-Wort für Abstieg, Avatar genannt. VR ermöglicht es Benutzern, jeden möglichen Avatar zu verkörpern. Wenn wir ein Fitnessspiel spielen, verwenden wir vielleicht einen sportlich aussehenden Avatar. Wenn wir uns dann in einem (virtuellen) Spiegel betrachten, sehen wir den Avatar, der sich genauso bewegt wie wir. Durch die synchrone Bewegung unseres eigenen Körpers und des Avatars und die Wahrnehmung des Avatars aus der Egoperspektive werden wir zu einem gewissen Teil zu diesem Avatar.

Unsere Arbeiten zeigen, dass Avatare und ihre realitätsnahe Gestaltung viele positive Effekte haben. Die sorgfältige Gestaltung von Avataren verstärkt nicht nur das Gefühl, der Avatar zu sein, sondern kann auch dazu führen, dass wir Aufgaben schneller absolvieren und weniger Fehler machen. Avatare verursachen aber auch darüber hinaus weitere Effekte. Das Aussehen von Avataren beeinflusst unsere Wahrnehmung und Verhalten - solche Verhaltensänderungen, die auf stereotypen Einschätzungen beruhen, werden als Proteus-Effekt bezeichnet, benannt nach dem griechischen Gott Proteus, der über die Fähigkeit zur Gestaltverwandlung verfügt.

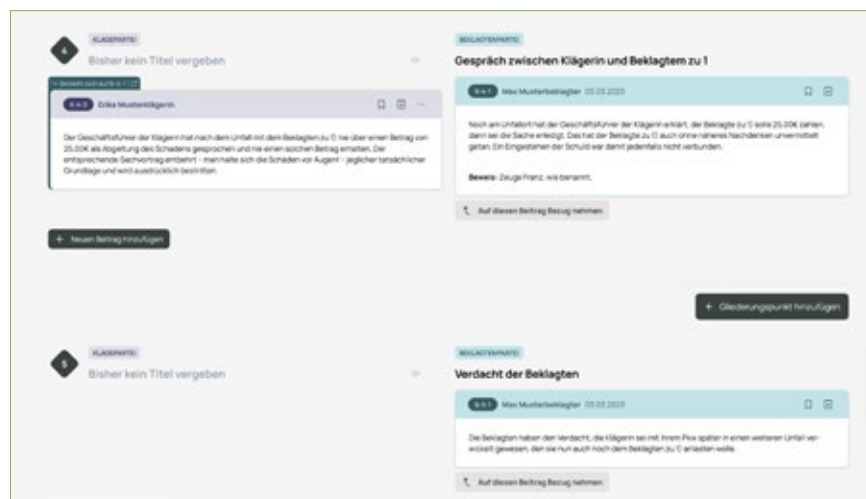
Unsere Arbeiten zeigen, dass der Proteus-Effekt auch Auswirkungen auf physiologische Messwerte hat. Wenn wir einen sportlich aussehenden Avatar verkörpern und in VR auf einem Ergometer Fahrrad fahren (Abb. 2), haben wir nicht nur das Gefühl, dass die sportliche Belastung weniger anstrengend ist, als wenn wir einen un-sportlich aussehenden Avatar verkörpern. Unsere Studien zeigen, dass ein sportlicher Avatar auch zu einem niedrigeren Puls führt (Kocur et al., 2021). Verwandte Effekte können sich auch bei anderen Aktivitäten und Avataren zeigen. Richtig gestaltet können Avatar und virtuelle Umgebung beispielsweise die Wärmeregulierung des menschlichen Körpers und das subjektive Temperaturempfinden beeinflussen (Kocur et al., 2023).

Usability Engineering und User Centered Design für interaktive Systeme in der Mensch-Maschine-Interaktion

Die Konzeption, Entwicklung und Evaluierung gebrauchstauglicher interaktiver Systeme ist seit jeher einer der wichtigsten Schwerpunkte der universitären Medien-



3 Screenshot Basisdokument, oben (quer) Funktionsleiste, rechts dritte Spalte für Kommentare, Hinweise, Neuordnung, in der Mitte zwei Spalten für den Parteivortrag (Quelle: <https://app.parteivortrag.de/>)



4 Screenshot Basisdokument »side-by-side«-Ansicht (Beiträge von Klage- und Beklagtenseite direkt gegenübergestellt), Quelle: <https://app.parteivortrag.de/>

informatik. Standen in den Anfangsjahren vor allem die Gestaltung interaktiver *betrieblicher* Informationssysteme im Mittelpunkt, hat sich in den vergangenen Jahrzehnten der Gestaltungsspielraum deutlich erweitert, weil interaktive Systeme zum selbstverständlichen Arbeitsmittel in allen Lebenslagen geworden sind. Durch Entwicklung multimedialer Techniken, heterogener Interaktionsgeräte vom Smartphone bis zum Desktop und nicht zuletzt die unterschiedlichen Interaktionsumgebungen wie grafische Benutzerschnittstellen haben sich vielfältige Anwendungsmöglichkeiten ergeben. Ein wesentlicher methodischer Fokus liegt dabei auf der nutzerzentrierten Gestaltung (*User Centered Design*) und einem mittlerweile ganzheitlichen Begriff des Nutzererlebens (*User Experience*), verbunden mit einer Betrachtung aller Schritte und Aspekte dieser Erfahrungen (*User Journey*).

In der Forschung spielen Fragen der technischen und gestalterischen Entwicklung von Benutzerschnittstellen sowie deren Evaluierung eine wichtige Rolle. Neben der Bewertung der Gebrauchstauglichkeit an sich werden ästhetische und hedonische Aspekte des Nutzererlebens (*Joy of Use*) untersucht, ebenso wie weitere Merkmale der Nutzer*innen (Alter, Geschlecht, Persönlichkeit, Erfahrung).

Ein Beispiel für derartige Gestaltungsprozesse in der Mensch-Maschine-Interaktion ist ein derzeit vom niedersächsischen und bayerischen Justizministerium gefördertes Projekt zur Entwicklung eines interaktiven elektronischen Basisdokuments, das von den Lehrstühlen für Medieninformatik (Prof. Wolff) und Zivilprozessrecht (Prof. Althammer) durchgeführt wird. Unter dem Titel »Strukturvorgaben für den Parteivortrag im Zivilprozess«



wird ein Prototyp eines elektronischen Basisdokuments für den Parteivortrag der Prozessbeteiligten (Klageseite, Beklagenseite, Richterschaft) in einem Reallabor an vier Landgerichten (Hannover, Landshut, Osnabrück, Regensburg) untersucht. Im Vorlauf zum Projekt wurden in mehreren Iterationsphasen mit Hilfe agiler Methoden (*User Stories*, *Legal Design Thinking*) zunächst gruppenspezifische Anforderungen der Anwalt- bzw. Richterschaft ermittelt und evaluiert. Im zweiten Schritt erfolgte das grundlegende Design und anschließend die funktionale Implementierung eines Prototyps, der als webbasierte Browser-Anwendung realisiert wurde und mit den Vorgaben und Rahmenbedingungen des elektronischen Rechtsverkehrs (ERV) und seiner technischen Umsetzung in den Justizverwaltungen Niedersachsens und Bayerns kompatibel ist. Eine besondere Herausforderung der Evaluierung liegt dabei in der Tatsache begründet, dass sie im Realbetrieb der Justiz, d. h. für »echte« Rechtsstreitigkeiten bei freiwilliger Teilnahme erfolgt. Dies stellt auch eine besondere Herausforderung bei der Evaluierung dar, da einerseits ein Maximum an ökologischer Validität gegeben ist (Erprobung im Realbetrieb), dies gleichzeitig aber die Steuerungsmöglichkeiten der Testläufe und die Erfassungsmöglichkeiten erheblich einschränkt. Der Softwareprototyp ist ein Vehikel, um Erkenntnisse über die künftigen Möglichkeiten der Werkzeugunterstützung beim Parteivortrag zu gewinnen und eine entsprechende Weiterentwicklung des Zivilprozessrechts vorzubereiten (Althammer et al., 2023).

Im Kern setzt die Benutzerschnittstelle die Idee einer Relationstabelle als Softwareanwendung um, in der die Argumente von Klage- und Beklagenseite jeweils in einer eigenen Spalte gegenübergestellt werden (Übersicht zur Benutzerschnittstelle in Abb. 3), wobei unterschiedliche Darstellungsmöglichkeiten genutzt werden können (z. B. side-by-side-Sicht, vgl. Abb. 4). Die Softwarelösung bietet zudem umfangreiche Möglichkeiten zur Darstellung, Umordnung und Annotation des Sachvortrags. Mit Hilfe von Suchfunktionen und Hervorhebungen lässt sich umfangreicher Vortrag besser erschließen, ein PDF-Export stellt die technische Kompatibilität mit den Vorgaben des elektronischen Rechtsverkehrs her und kann über das besondere elektronische Anwaltspostfach (beA) versandt werden.

Weitere Informationen zu dem Projekt finden sich unter <http://www.basisdokument.de> und <http://www.parteeivortrag.de>. Das Projekt hat ein intensives Echo in der Fachöffentlichkeit ausgelöst und ist insofern ein gutes Beispiel für Transfer als third mission der Hochschulen. Erfreulich ist, dass auch die zunächst sehr skeptische Bundesrechtsanwaltskammer im September 2023 zur Erprobung der Software aufgerufen hat (<https://t1p.de/BRAKempfehl-Basisdokument>).

Das Projekt ist ein Beispiel für die mittlerweile langjährige Beschäftigung der Regensburger Medieninformatik mit Themen der Rechtsinformatik bzw. der Legal Technology. Die Medieninformatik hat zu unterschiedlichen Themenfeldern (Text Mining, Information Retrieval, Rechtsvisualisierung, Sprache und Recht) gearbeitet und an der Konzeption der beiden 2020 bzw. 2021 eingeführten Studiengänge LL. M. Legal Tech (Weiterbildungsmaster) und LL.B. Digital Law mitgewirkt, für die sie auch in der Lehre Beiträge leistet.

Digitale Geisteswissenschaften – digital humanities

Schon fast seit Gründung der Universität Regensburg gibt es Aktivitäten im Bereich der digitalen Geisteswissenschaften. Nicht zuletzt durch ihre ursprüngliche Ansiedlung an der Fakultät für Sprach-, Literatur- und Kulturwissenschaften ist dies ein aktives und vielfältiges Tätigkeitsfeld der Medieninformatik (Burghardt & Wolff, 2014). Zusammen mit dem Lehrstuhl für Informationswissenschaft wurde 2017 ein Masterstudiengang für Digital Humanities eingerichtet und gemeinsam mit den Universitäten München (LMU) und Erlangen-Nürnberg (FAU), die vergleichbare Studiengänge etabliert haben, zum Campusnetzwerk *téchnē* erweitert. Aus Sicht der Informatik ist die große Vielfalt möglicher Fragestellungen in den Geistes- und Sozialwissenschaften reizvoll – nicht zuletzt hat die Gesellschaft für Informatik vor einigen Jahren die Frage nach der Bewahrung des digitalen Kulturerbes zu einer ihrer Grand Challenges erklärt (vgl. <https://gi.de/grand-challenges>). Auf diesem Feld finden etwa Studien zur Analyse sozialer Medien statt, die besondere Herausforderungen für die automatische Analyse und Klassifikation der sprachlichen Besonderheiten bieten, aber auch Probleme der Bildklassifikation und der Informationsextraktion aus

Bildern aufwerfen. Ein weiteres Arbeitsfeld ist die Entwicklung und Evaluierung softwaregestützter Arbeitswerkzeuge für die geisteswissenschaftliche Forschungsarbeit, z. B. im Bereich der Annotation unterschiedlicher Medien. Hier sind sowohl Fragen der gebrauchstauglichen Gestaltung als auch der Einsatz aktueller KI-Methoden (*deep learning*, *large language models*) zu bearbeiten.

Im Rahmen eines DFG-Projektes im Schwerpunktprogramm *Computational Literary Studies* (SPP 2207) zum Thema *Emotions in Drama* entstanden in den vergangenen Jahren umfangreiche Arbeiten zur automatischen Identifikation von Sentiment im Sinne der Erkennung von Emotionen in historischen Theaterstücken. Dazu konnten bereits vorhandene digitale Korpora deutscher Dramen vom Barock bis zur Klassik (ca. 1600–1850) genutzt werden. Zu den Herausforderungen der Forschung, die in Zusammenarbeit mit PD Dr. Katrin Dennerlein, Germanistin in Würzburg, durchgeführt wird, gehören einerseits die Erarbeitung eines spezifischen Emotionsmodells, das für die literaturwissenschaftliche Forschung geeignet ist und sich erheblich von den sonst im Bereich Sentiment-Analyse und Opinion Mining verwendeten Modellen unterscheidet, die typischerweise aus der psychologischen Forschung stammen. Ein zweites Problem war die Frage der Vorgehensweise bei der Emotionsannotation solcher Dramen als Voraussetzung für die Anwendung von Verfahren des maschinellen Lernens. Die Entwicklung und Standardisierung der Annotationspraxis sowie die Ausbildung der dafür eingesetzten Fachexpert*innen gehören ebenfalls zu den Arbeitsaufgaben im Projekt. Auf der Basis umfangreicher Dramenannotationen konnten Klassifikationsstudien mithilfe aktueller Transformer-Modelle durchgeführt werden, wobei unterschiedlich komplexe Klassifikationsaufgaben bewältigt wurden (Dennerlein et al., 2023). Eine besondere Herausforderung ist die für die vorliegenden Dramen teilweise noch nicht standardisierte Orthographie sowie die ältere Sprachstufe (Frühneuhochdeutsch). Aufbauend auf den vorklassifizierten Dramen lassen sich unterschiedliche Fragestellungen rechnergestützt beantworten: Dazu gehört die Analyse des Emotionsverlaufs in unterschiedlichen dramatischen Genres über das ganze Stück hinweg, also etwa der Vergleich zwischen Tragödien und Komödien. Eine andere Frage ist, inwiefern sich Regieanweisun-



Bildnachweis: Martin Brockelmann

5 3D-Modell des Future Interaction Lab, die Trennwände können entfernt werden und erlauben eine flexible Raumnutzung

gen hinsichtlich des Vorhandenseins von Emotionen vom eigentlichen Dramentext unterscheiden und wie sich dies historisch entwickelt hat. Durch das umfangreiche digital verfügbare Dramenmaterial lässt sich das in den digitalen Geisteswissenschaften etablierte Konzept des *Distant Reading*, das an die Stelle der typischen literaturwissenschaftlichen Arbeitsweise der intensiven Analyse kleiner Textmengen (*Close Reading*) die computergestützte automatische Analyse großer Textmengen setzt, auf geradezu paradigmatische Weise umsetzen.

im Bereich der Virtual Reality, der auch mit einer leistungsfähigen Motion Tracking-Anlage ausgestattet ist und als technische Grundlage beispielsweise für die Studien zu sozialer Interaktion und Proteus-Effekt in virtueller Realität genutzt wird (Abb. 6).

In einer langfristigen Kooperation mit der OTH Regensburg (Prof. Mottok) sowie dem Lehrstuhl für Erziehungswissenschaft (Prof. Gruber) konnte 2019 ein *Eye-Tracking Lab and Classroom* einge-

richtet werden, in dem 15 hochwertige Eye-Tracking-Systeme sowohl als Desktopgeräte als auch als Integration mit VR-Brillen verfügbar sind. Ein besonderer Themenschwerpunkt der Forschung dort ist im Bereich Software Engineering und Software Engineering Education bzw. Informatikdidaktik angesiedelt, d. h. die Eye Tracker werden genutzt, um Arbeitsprozesse in der Softwareentwicklung bzw. der Softwaretechnik-Ausbildung näher zu untersuchen (Gruber et al., 2020).

Forschungsinfrastruktur – die Labore der Medieninformatik

Die voranstehend geschilderte Vielfalt und Bandbreite der in der Forschung der Medieninformatik bearbeiteten Fragestellungen ist nur durch die Verfügbarkeit einer entsprechenden technischen Infrastruktur denkbar und möglich. Bereits mit Einrichtung des Studiengangs Medieninformatik konnte ein Future Interaction Lab (FIL) eingerichtet werden, das vor allem für Studien im Bereich Usability eingesetzt wird (Abb. 5).

Seit 2018 existiert daneben ein größerer Laborkomplex, der in der TechBase, dem Regensburger Gründerzentrum neben dem Universitätscampus, angesiedelt ist. Neben einer Werkstatt, in der Hardware-Prototypen gebaut werden können, gibt es dort einen größeren Raum für Studien



Bildnachweis: Team Medieninformatik

6 VR-Studio in der Tech Base, Kameras des Motion Tracking-Systems sind an den Traversen angebracht; im eingesetzten Bild eine Versuchsperson im Motion Tracking-Anzug und mit VR-Brille



Literatur

Althammer, C., Bauer, J., Böhm, V., Fehle, J., Mielke, B., & Wolff, C. (2023, 30. März 2023). Das Basisdokument geht ins Reallabor: zur Evaluation des Einsatzes bei Gericht In Rechtsinformatik als Methodenwissenschaft des Rechts / Legal Informatics as Science of Legal Methods. Proceedings des 26. Internationalen Rechtsinformatiksymposiums IRIS 2023, Salzburg. https://jusletter-it.web-law.ch/issues/2023/30-maerz-2023/das-basisdokument-ge_d958f6123c.html

Bogon, J., Högerl, J., Kocur, M., Wolff, C., Henze, N., & Riemer, M. (2023). Validating virtual reality for time perception research: Virtual reality changes expectations about the duration of physical processes, but not the sense of time. Behavior Research Methods.

Burghardt, M., & Wolff, C. (2014). Digital Humanities: Buzzword oder Strukturwandel in den Geisteswissenschaften? Stand und Perspektiven anhand Regensburger Beispiele. Blick in die Wissenschaft: Forschungsmagazin der Universität Regensburg, 25, 39-47.

Dennerlein, K., Schmidt, T., & Wolff, C. (2023). Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century. Digital Scholarship in the Humanities. <https://doi.org/10.1093/llc/fqad046>

Gruber, H., Wolff, C., Mottok, J., Bazo, A., Hauser, F., & Schreitetter, S. (2020, 9-12 Nov. 2020). Tutorial on Software Engineering Education in Co-Located Multi-User Eye-Tracking-Environments. In 2020 IEEE 32nd Conference on Software Engineering Education and Training (CSEET), München.

Halbhuber, D., Thomaschke, R., Henze, N., Wolff, C., Probst, K., & Bogon, J. (2023). Play with my Expectations: Players Implicitly Anticipate Game Events Based on In-Game Time-Event Correlations. In Proceedings of the 2023 International Conference on Mobile and Ubiquitous Multimedia.

Hammwöhner, R., Knäusel, H., & Wolff, C. (2012). Informationswissenschaft und Medieninformatik an der Universität Regensburg. Information - Wissenschaft & Praxis, 63(5), 285-291. <https://doi.org/10.1515/iwp-2012-0066>

Kocur, M., Habler, F., Schwind, V., Woźniak, P. W., Wolff, C., & Henze, N. (2021). Physiological and perceptual responses to athletic avatars while cycling in virtual reality. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.

Kocur, M., Jackermeier, L., Schwind, V., & Henze, N. (2023). The Effects of Avatar and Environment on Thermal Perception and Skin Temperature in Virtual Reality. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.



Foto © Andreas Schmid

Dr. Raphael Wimmer (geb. 1980) ist akademischer Rat am Lehrstuhl für Medieninformatik. Zu seinen Forschungsschwerpunkten zählen Hard- und Software für interaktive Systeme, Computer Vision und Werkzeuge zur Interaktion mit digitalen Daten



Foto © Andreas Schmid

Vertr.-Prof. Dr. Johanna Bogon (geb. 1982) vertritt derzeit den Lehrstuhl für Mensch-Maschine-Interaktion an der Universität Regensburg. Sie ist promovierte Kognitionspsychologin (2012, Nijmegen) und war als Postdoc am Forschungszentrum für Kognitive Interaktionstechnologie (CITEC) der Universität Bielefeld, sowie am Lehrstuhl für Allgemeine und Angewandte Psychologie und am Lehrstuhl für Medieninformatik der Universität Regensburg tätig. Ihre Forschungsschwerpunkte beinhalten die Interaktion von Zeit, Wahrnehmung und Handlung und prädiktive Prozesse im Kontext der Mensch-Maschine-Interaktion und Virtueller Realität.



Foto © Andreas Schmid

Prof. Dr. Niels Henze (geb. 1978) ist seit 2018 Professor für Medieninformatik. Er promovierte 2012 an der Universität Oldenburg und war von 2013 bis 2018 Juniorprofessor für Soziokognitive Systeme an der Universität Stuttgart. Seine Forschungs konzentriert sich auf die Mensch-Computer Interaktion mit Fokus auf Mixed Reality, die Verwendung von Modellen in interaktiven Systemen und mobile Mensch-Computer Interaktion.



Foto © Andreas Schmid

Prof. Dr. Christian Wolff (geb. 1966) ist seit 2003 Professor für Medieninformatik an der Universität Regensburg. Er ist promovierter Informationswissenschaftler (1994, Universität Regensburg) und habilitierter Informatiker (2000, Universität Leipzig). Zu seinen Forschungsschwerpunkten zählen: Mensch-Maschine-Interaktion, (Multimedia-)Software Engineering sowie automatische Sprachverarbeitung, insbesondere mit Anwendungen in den Geistes- und Sozialwissenschaften sowie der Rechtsinformatik.

Wissen aus dem Internet – Genug, genau, geprüft und geeignet?

Informationswissenschaft in Regensburg

Prof. Dr. Udo Kruschwitz, Prof. Dr. Bernd Ludwig, PD Dr. David Elsweiler

Am Lehrstuhl für Informationswissenschaft und an der Professur für Informationslinguistik befassen wir uns damit, wie KI Menschen helfen kann, ihr Wissen und ihre Fähigkeiten zu erweitern. Der Fokus ist dabei nicht die Automatisierung von menschlichen Tätigkeiten mit Methoden der KI und des maschinellen Lernens, sondern die Entwicklung eines tiefen Verständnisses davon, wie Menschen sich mit den neuartigen Werkzeugen des Informationszeitalters Wissen und Informationen aus digitalen Quellen aneignen oder von ihnen Hilfestellungen geben lassen, wie sie Tätigkeiten verrichten sollen. Ziel unserer Arbeit ist es einerseits vorherzusagen, wie Menschen in einer Situation vorgehen werden, um für sie die beste Information schnell bereitstellen zu können, und andererseits zu analysieren, welche Qualität verfügbare Information hat. Um unsere Ziele zu erreichen, führen wir unter anderem komplexe Experimente durch und entwickeln Software, die den Zugang zu verfügbarer Information mittels KI-Methoden vereinfacht.

Das Informationszeitalter und seine Folgen

Wir leben im Informationszeitalter, in dem das Internet und soziale Medien eine immer zentralere Rolle spielen. Die *Möglichkeiten*, die sich daraus ergeben, sind phänomenal. Was vor wenigen Jahren noch unmöglich erschien wie etwa die automatische Übersetzung von Texten in über 100 Sprachen, exakte Routenbeschreibungen von A nach B oder die Beantwortung von Fragen per Knopfdruck sind nur einige Beispiele.

Die Nutzung des Internets birgt allerdings auch vielfältige *Risiken* für das menschliche Entscheidungsverhalten, welche sich nicht selten zu gesamtgesellschaftlichen Problemen entwickeln. Diese Risiken sind alltäglich [7], bleiben bei den Betroffenen oft unerkannt und können fatal sein, wenn sie Entscheidungen bezüglich medizinischer Behandlungen oder politischer Wahlen betreffen [3]. Wenn Menschen durch das Internet surfen, stoßen sie beispielsweise häufig auf schädliche Inhalte oder treffen Entscheidungen auf Basis bestimmter Fehlannahmen. Solche Inhalte reichen von Fehlinformationen oder sogenannten Fake News bis hin zu Hassrede (z. B. beleidigenden, sexistischen oder rassistischen Kommentaren) und sind massenweise auf Social-Media-Plattformen zu finden.

Es ist allerdings nicht immer leicht, die Qualität von Information richtig zu bewerten. Menschen sind dabei oft von ihren Vorkenntnissen, ihren Einstellungen und ihrem Handlungskontext beeinflusst.

Fehlannahmen (kognitive Verzerrungen) sind dabei Denkfehler sowie Fehler beim Wahrnehmen und Urteilen. Solche Fehler sind nicht online-spezifisch, sondern allgemeine menschliche Neigungen, die sich im Internet fortsetzen und durch unausgewogene Suchmaschinen-Algorithmen und unausgewogene Inhalte im Internet verstärkt werden [1]. Beispiele sind der *Confirmation Bias*, d. h. die Neigung, Informationen bei der Informationssuche zu bevorzugen, die die eigenen Erwartungen bestätigen, und das *Anchoring*, der *unbewusste* Einfluss von Umgebungsreizen auf menschliche Entscheidungen.

All dies kann zu menschlichen Fehlentscheidungen führen mit teils gravierenden individuellen Folgen (z. B. psychische Erkrankungen, Geldnot) als auch gesamtgesellschaftlichen Schäden (z. B. Volksverhetzung, Wissenschaftsskepsis). Folglich ist es entscheidend, sich diesen problematischen Aspekten des menschlichen Informationsumgangs im Internet zu widmen und Lösungsansätze zu entwickeln. Für diesen Zweck erscheinen drei Arten von Herangehensweisen sinnvoll:

- Die zugrundeliegenden Probleme müssen *verstanden* werden, damit beispielsweise entsprechende KI-Modelle in jeder Situation die wahrscheinlichste Bewertung von Information durch Informationssuchende vorhersagen können.
- Problematische Inhalte können durch KI-basierte *intervenierende* Gegenmaßnahmen identifiziert und klassifiziert werden.
- Noch weiter gehen *Assistenzmaßnahmen*, z. B. durch Bildung und Software-Werkzeuge zur Unterstützung der menschlichen Entscheidungsfindung und Handlungsausführung in definierten Situationen.

Alle drei Herangehensweisen sind zentrale Säulen in der Forschung am Lehrstuhl für Informationswissenschaft an der Universität Regensburg.

Verstehen: Informationsumgang

Grundlagenforschung zu Fehlannahmen beziehungsweise kognitiven Verzerrungen bei Informationssuchenden kann



Bildnachweis: Yanpu Chen

1 Das Team des Lehrstuhls für Informationswissenschaft an der Universität Regensburg (nicht im Bild: Melanie Kilian).

Bedingungen, Einflussfaktoren und unterschiedliche Verhaltensauswirkungen, die im Zusammenhang mit menschlichen Fehlentscheidungen stehen, aufdecken. Dies schafft eine empirische Basis für die Entwicklung effektiver Interventions- und Präventionsmaßnahmen. Wir haben Fehlannahmen beziehungsweise kognitive Verzerrungen und menschliche Entscheidungsfindung in verschiedenen Kontexten ergründet: bei der Auswahl von Suchergebnissen während der Nutzung von Internet-Suchmaschinen, bei der Auswahl von Rezepten auf Online-Rezeptportalen und beim Stellen von Informationsanfragen an Mitarbeitende eines Flughafeninformationsstands.

Unsere Suchmaschinen-Studien haben gezeigt, dass Entscheidungen von mehreren Faktoren beeinflusst werden und dass deren Auswirkungen von Mensch zu Mensch sowie von Situation zu Situation variieren können. Im Rahmen einer Studie zur Auswahl von Suchmaschinenenergebnissen [4] haben wir Auswirkungen des Confirmation Bias auf die Beurteilung der Glaubwürdigkeit von Treffern erforscht. Dabei konnten wir zehn verschiedene Faktoren identifizieren, die Entscheidun-

gen bei der Auswahl von Suchergebnissen beeinflussen. Interessanterweise konnte derselbe Einflussfaktor sowohl positive als auch negative Auswirkungen auf die Glaubwürdigkeit eines Suchergebnisses haben. Ein Einflussfaktor stellte beispielsweise die Internetadresse dar, die mit den jeweiligen Suchergebnissen in Verbindung stand¹. Internetadressen wurden je nach Versuchsperson als unterschiedlich glaubwürdig wahrgenommen – einige Versuchspersonen hielten zum Beispiel *Die Welt* für eine seriöse Quelle, andere hingegen nicht.

Unsere Studien zur Auswahl von Rezepten auf Online-Rezeptportalen offenbarten unter anderem, dass Menschen schlecht beurteilen können, welche Gerichte am meisten Fett enthalten (richtige Beurteilung in nur 51,1 Prozent der Fälle), und dass die Beurteilungen wenig miteinander übereinstimmten.

All diese Ergebnisse legen nahe, dass Informationen im Internet als verifiziert bzw. nicht-verifiziert gekennzeichnet sein sollten und dass Internetnutzende zusätzliche Informationen brauchen, um bessere Entscheidungen treffen zu können.

Unsere Forschungen verdeutlichen darüber hinaus, dass Informationssuchende

nicht nur während ihrer Entscheidungen, also während der Phase der Auswahl von Alternativen, Unterstützung benötigen, sondern dass Fehlannahmen, die früher im Suchprozess auftauchen, z. B. beim Formulieren von Suchanfragen, Fehlentscheidungen vorprogrammieren können. Interviewdaten, die wir im Rahmen einer Studie zur Informationssuche bei Flughafengästen (z. B. Flugreisenden, Abholenden) [6] erhoben haben, zeigen, dass Fehlannahmen zu Informationsanfragen führen können, die nach außen hin zwar legitim wirken, aber letztendlich gefährden, dass das übergeordnete Ziel des Informationssuchenden erreicht wird. Ein Beispiel sind Situationen, in denen Informationsschende aufgrund von Fehlannahmen nach unzureichenden Informationen für die Erledigung ihrer Aufgaben suchen (z. B. wenn Personen nach dem Weg zu einem italienischen Restaurant fragen, weil sie eine Pizza Mallorquina essen möchten, die eigentlich ein spanisches Gericht ist).

¹ z. B. www.welt.de oder www.faz.net



2 Der virtuelle Agent in Aktion

(Quelle: <http://optapeb.de>)

Intervention: Erkennung von Fake News

Das Risiko, das Fehlinformationen für die menschliche Entscheidungsfindung darstellen, kann zum Teil eingedämmt werden, indem intervenierende Gegenmaßnahmen implementiert werden. Verfahren zur automatischen Erkennung von Fehlinformationen bilden eine mögliche Basis für Interventionsmaßnahmen (z. B. für das Löschen erkannter Fake News).

Ursprünglich hatten sich die Methoden des maschinellen Lernens hauptsächlich auf die Texte selbst konzentriert, also versucht, auf der Grundlage des sprachlichen Inhalts eines Nachrichtenartikels diesen als Fake News zu erkennen. Noch effektiver wird die Erkennung aber durch Einbeziehung zusätzlicher, kontextbezogener Informationen – z. B. Informationen darüber, wie sich die betrachteten Nachrichtenartikel in sozialen Medien verbreiten und wie Internetnutzende mit diesen Artikeln interagieren [11]. In unseren eigenen Arbeiten haben wir nachweisen können, dass unter Einbeziehung kontextbezogener Signale die Erkennung von Falschinformation als Graphklassifikationsproblem interpretiert robuste und effektive Ergebnisse liefert [2].

Assistenzmaßnahmen: Menschen- und Aufgabenorientierte Informationssysteme

Assistenzmaßnahmen sind für zahllose Anwendungen vorstellbar. Ein Beispiel ist die Simulation einer psychotherapeutischen Sitzung mit Hilfe eines virtuellen Agenten (siehe Abb. 2). Hier wirken verschiedene KI-Modelle zusammen: Eines für das psychotherapeutische Personal, das vorherzusagen kann, wie Fachleute versuchen, ihr Behandlungsziel je nach Reaktion der zu Behandelnden zu erreichen; und ein anderes für die zu Behandelnden, das ihren mentalen und psychischen Zustand aus verschiedenen Messwerten und beobachteten Reaktionen erkennen kann. Hier nutzen wir Deep Learning-Modelle.²

Andere, gesellschaftlich wichtige Beispiele finden sich bei der Verarbeitung natürlicher Sprache. In diesem Teilbereich der KI wurden in den letzten Jahren erstaunliche wissenschaftliche Fortschritte erzielt. Diese Entwicklungen so einzusetzen, dass Menschen beim Umgang mit sozialen Medien praktische Kompetenzen vermittelt werden, sehen wir als eine wichtige Komponente unserer Arbeit. Ein Beispiel ist der Einsatz multimodaler Algorithmen für

maschinelles Lernen, die in einen Social-Media-Begleiter integriert werden können, der die Internetnutzenden beim Umgang mit Fake News und anderen Bedrohungen in den sozialen Medien unterstützt und aufklärt [8].

Neben der Ergründung von Bedingungen für Fehlentscheidungen und der Bekämpfung von Internetinhalten, die solche Fehlentscheidungen begünstigen können, wie beispielsweise Fake News, gibt es die Möglichkeit, Fehlentscheidungen vorzubeugen, indem Informationssysteme – von Websuchmaschinen bis hin zu virtuellen Sprachassistenten – entwickelt werden, die die menschliche Entscheidungsfindung positiv unterstützen. Denkbar wäre hier – wie bereits oben beschrieben –, dass verifizierte Informationen in Suchergebnislisten kenntlich gemacht werden und dass Informationssysteme ihre Antworten durch zusätzliche Informationsboxen ergänzen, in denen korrekte Kontextinformationen zum angefragten Themenbereich zu finden sind [9]. Proaktive Informationssysteme [10]

² Das Video https://www.youtube.com/watch?v=3_Ex6UjK6co zeigt unser System im Einsatz.

sind eine weitere Präventionsmaßnahme, also Informationssysteme, die Informationen nicht nur auf der Basis von Suchanfragen liefern, sondern Menschen an der Hand nehmen und sie leitend durch zu erledigende Aufgaben führen (z. B. mittels sog. Task-Tracking), wie etwa ein virtueller Koch- oder Reiseassistent. Solche Systeme nehmen Menschen Entscheidungen ab, sodass Fehlentscheidungen in vielen Fällen erst gar nicht entstehen. In einer In-situ-Studie an einem Flughafen [5] haben wir herausgefunden, dass selbst Systeme mit rudimentären Task-Tracking-Fähigkeiten Menschen bei der Aufgabenbewältigung unterstützen und einen positiven Einfluss darauf haben können, wie Menschen die Aufgaben wahrnehmen.

Ausblick

Informationswissenschaft, KI und maschinelles Lernen bilden zusammen das spannende Forschungsfeld *Human-Centred AI*, das Menschen den Alltag leichter machen und ihnen helfen soll, sich in der Informationsflut zurechtzufinden, die aus der Digitalisierung resultiert. *Human-Centred AI* ist nicht nur spannend, sondern auch gesellschaftlich höchst relevant, weil da-

mit Informationen für andere Zwecke aufbereitet werden können als für die sie erstellt wurden. ChatGPT ist eines der aktuell prominenten Beispiele dafür. Die Forschung in der Zukunft darf nicht nur diese Systeme weiterentwickeln, sondern muss auch Lösungen dafür bereitstellen, ob die Antworten, die Informationssuchende von KI-Technologie auf ihre Fragen erhalten, wirklich genug, genau, geprüft und geeignet sind: Informationswissenschaft – am Puls der Zeit!

Literatur

[1] Leif Azzopardi. "Cognitive Biases in Search: A Review and Reflection of Cognitive Biases in Information Retrieval". In: *Proceedings of CHIIR '21*. 2021.

[2] Gregor Donabauer und Udo Kruschwitz. "Exploring Fake News Detection with Heterogeneous Social Media Context Graphs". In: *Proceedings of ECIR '23*. 2023.

[3] Amira Ghenai, Mark D. Smucker und Charles L.A. Clarke. "A ThinkAloud Study to Understand Factors Affecting Online Health Search". In: *Proceedings of CHIIR '20*. 2020.

[4] Markus Kattenbeck und David Elsweiler. "Understanding credibility judgements for web search snippets". In: *Aslib Journal of Information Management* (2019).

[5] Markus Kattenbeck, Melanie A. Kilian, Matthias Ferstl, Florian Alt und Bernd Ludwig. "Towards task-sensitive assistance in public spaces". In: *Aslib Journal of Information Management* (2019).

[6] Melanie A. Kilian und David Elsweiler. "Right Answers to Wrong Questions: The Dysfunctional Nature of Information Needs". In: *In review* (2023).

[7] Salman Bin Naeem, Rubina Bhatti und Aqsa Khan. "An exploration of how fake news is taking over social media and putting public health at risk". In: *Health Information & Libraries Journal* (2021).

[8] Dimitri Ognibene, Gregor Donabauer, Emily Theophilou, Sathya Bursic, Francesco Lomonaco, Rodrigo Wilkens, Davinia Hernández-Leo und Udo Kruschwitz. "Moving Beyond Benchmarks and Competitions: Towards Addressing Social Media Challenges in an Educational Context". In: *Datenbank-Spektrum* (2023).

[9] Anna-Marie Ortloff, Steven Zimmerman, David Elsweiler und Niels Henze. "The Effect of Nudges and Boosts on Browsing Privacy in a Naturalistic Environment". In: *Proceedings of CHIIR '21*. 2021.

[10] Chirag Shah, Ryen White, Paul Thomas, Bhaskar Mitra, Shawon Sarkar und Nicholas Belkin. "Taking Search to Task". In: *Proceedings of CHIIR '23*. 2023.

[11] Kai Shu, Suhang Wang und Huan Liu. "Beyond News Contents: The Role of Social Context for Fake News Detection". In: *Proceedings of WSDM '19*. 2019.

FRAGEN AN BERND LUDWIG

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

Ich möchte Modelle statistisch-relationaler KI für die Kooperation von Menschen mit KI-gesteuerten Geräten entwickeln. Diese Modelle sollen erklären, wie der Gegenstandsbereich, in dem gerade kooperiert wird, mit dem Sprechen darüber in der Mensch-Maschine-Interaktion interferiert. Das, was man sagt, provoziert nämlich, was man tut, und das wiederum beeinflusst, was man sagt und worüber man spricht. Die neuen Modelle sollen klassische »pipeline«-Ansätze dafür ablösen, in denen Interaktion und Aktion in einer Domäne als sequentielle Vorgänge gesehen werden, die man separat modellieren und nacheinander abarbeiten kann.

Was ist Ihre größte fachliche Herausforderung?

Die Symbiose von statistischem und relationalem Wissen. Rein regelbasierte Ansätze für die Analyse von Mensch-Maschine-Interaktion haben in der Vergangenheit immer wieder Probleme bereitet, weil sie sehr komplex sind und dennoch mit Ausnahmefällen kaum umgehen können. Rein statistische Ansätze lernen nur aus Beispielen und funktionieren in konkreten, aber sehr unwahrscheinlichen Situationen oft nicht optimal. Im Natural Language Engineering ist diese Problematik besonders gravierend, da sich aktuelle Deep Learning-Modelle immer weiter von (computer-)linguistischen Theorien entfernen. Es sieht fast so aus, als ob disjunkte Fragestellungen behandelt werden. Diese Kluft zu schließen, ist eine große Herausforderung.

Was möchten Sie in den nächsten Jahren entwickeln?

Wir möchten ein Beispiel für ein sprachbasiertes Assistenzsystem realisieren, das eine Theorie umsetzt, wie solche Systeme an verschiedene Domänen angepasst werden können, die ganz anderes Wissen erfordern. Die sprachliche Kommunikation funktioniert aber immer nach denselben Prinzipien. Ihre Operationalisierung soll das Assistenzsystem leisten.

Wie verändert KI Ihr Fachgebiet?

Natural Language Engineering war schon immer ein Gebiet der Künstlichen Intelligenz und des Maschinellen Lernens. Die KI hat nicht mein Fachgebiet verändert, sondern eher das Fachgebiet die KI. Wichtige Entwicklungen wie Transformer wurden an Beispielen aus dem Natural Language Engineering erarbeitet.

FRAGEN AN DAVID ELSWEILER

Welches Problem möchten Sie gern durch Ihre Forschung lösen?

Es ist zunehmend offensichtlich geworden, dass Algorithmen in eine Vielzahl gesellschaftlicher Probleme verwickelt sind. Sie tragen beispielsweise zur Verbreitung von Fehlinformationen und zur Verschärfung gesellschaftlicher Spaltungen bei. Paradoxerweise werden aber Algorithmen oft als Lösungen für genau diese Probleme angesehen. Ein bemerkenswerter Wandel vollzieht sich jedoch gerade, der sich von der »techno-lösungsorientierten« Perspektive entfernt und stattdessen den Schwerpunkt darauf legt, eine ausgewogenere Synergie zwischen Algorithmen und menschlicher Handlungsfähigkeit zu schaffen.

Indem wir lernen und verstehen, wie Individuen mit Systemen interagieren, welche Auswirkungen diese Interaktionen auf sie haben, welche inhärenten Vorurteile existieren und wie sie in der Lage sind, diese Vorurteile zu bewältigen und zu reduzieren, können wir dazu beitragen, diesen Herausforderungen zu begegnen.

Was ist Ihre größte fachliche Herausforderung?

Die Vielzahl von (manchmal latenten) Variablen, die im Spiel sind. Wenn Menschen handeln, geschieht dies niemals im luftleeren Raum, das heißt, jede Handlung wird von unzähligen persönlichen, sozialen, kulturellen und situativen Variablen beeinflusst. Dies macht das Untersuchungsobjekt von Natur aus komplex (aber spannend).

Was möchten Sie in den nächsten Jahren entwickeln?

Ich bin derzeit in mehreren laufenden Projekten engagiert, die sich mit der Untersuchung des Informationsverhaltens befassen und die gewonnenen Erkenntnisse zur Entwicklung interaktiver Informationssysteme nutzen. Diese Projekte umfassen die Untersuchung des Verhaltens in verschiedenen Umgebungen, von Küchenumgebungen bis hin zu Übergangsbereichen wie Flughäfen. Sie beinhalten auch Initiativen, die darauf abzielen, die Reflexion zu fördern, um Verhaltensänderungen zu

unterstützen, beispielsweise um mit dem Rauchen aufzuhören oder sich gesünder zu ernähren. Darüber hinaus konzentrieren sich einige dieser Projekte darauf, die Suchstrategien im Web zu modifizieren, um den Einfluss von algorithmischen Verzerrungen zu verringern

Wie verändert KI Ihr Fachgebiet?

Mein Forschungsbereich wurde kontinuierlich von Technologien, insbesondere Künstlicher Intelligenz, geprägt und hat diese ihrerseits geprägt. Aktuelle Fortschritte in der Künstlichen Intelligenz, wie beispielsweise LLMs, haben auf mehrere bemerkenswerte Weisen erhebliche Auswirkungen gehabt. Erstens haben sich die Kosten für die Entwicklung funktionsfähiger Assistenzsystemprototypen drastisch reduziert. Zweitens haben sie den Prozess der Generierung und Simulation von Datensätzen und Annotationen vereinfacht und damit unsere Arbeitsmethoden grundlegend verändert. Dies sind zweifellos aufregende Zeiten!



Foto © UR, Antonia Profs

Prof. Dr. Udo Kruschwitz

ist seit 2019 Professor für Informationswissenschaft an der Universität Regensburg. Vorher war er über 20 Jahre an der University of Essex tätig. Sein Hauptforschungsgebiet ist die Schnittstelle zwischen Information Retrieval (IR) und Natural Language Processing (NLP). Er ist derzeit Chair der Information Retrieval Specialist Group bei der British Computer Society.



Foto © UR, Antonia Profs

Prof. Dr. Bernd Ludwig

ist seit August 2011 Professor für Informationslinguistik an der Universität Regensburg. Zuvor arbeitete er am Lehrstuhl für Künstliche Intelligenz an der Friedrich-Alexander-Universität Erlangen-Nürnberg, an der Freien Universität Bozen und der Università degli Studi di Trento. Sein Forschungsschwerpunkt ist der Zusammenhang zwischen Sprache und Handeln in zweckrationalen Mensch-Maschine-Interaktionen. Er ist Gutachter bei und Organisator von verschiedenen internationalen Tagungen und Mitherausgeber der Zeitschrift Künstliche Intelligenz.



Foto © UR, Lst. Informationswissenschaft

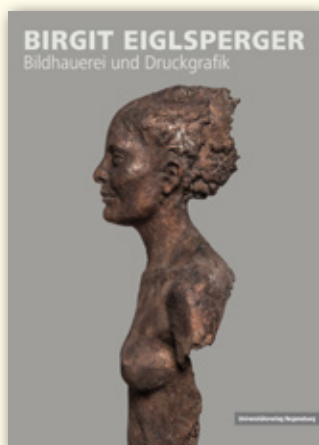
PD Dr. David Elweiler

ist Akademischer Oberrat am Lehrstuhl für Informationswissenschaft an der Universität Regensburg. Seine Forschung konzentriert sich darauf, das Informationsverhalten zu verstehen, mit dem Hauptziel, Informationssysteme zu entwerfen, die mit der natürlichen Denkweise und dem Verhalten von Menschen in Einklang stehen. In seinen neuesten Arbeiten liegt der Fokus auf Gesundheitsaspekten, darunter die Analyse des Verhaltens bei der Nutzung von Online-Rezepten, Empfehlungsalgorithmen und der Entwicklung von Informationssystemen, die Menschen bei der Umsetzung von Verhaltensänderungen unterstützen.

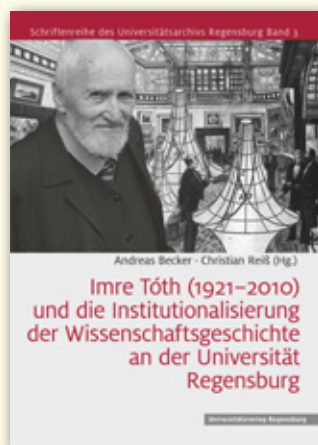
Universitätsverlag Regensburg



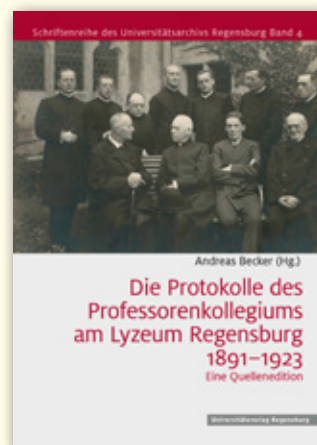
Birgit Eiglsperger
MENSCHENBILDER
Reihe: Kunst und Wissenschaft
1. Aufl. 2023
96 S., 21 x 27 cm
Klappenbrochur fadengeheftet
ISBN 978-3-86845-175-7
€ 16,90



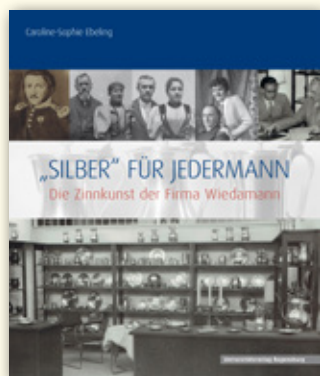
Birgit Eiglsperger
BIRGIT EIGLSPERGER
Bildhauerei und Druckgrafik
1. Aufl. 2018
128 S., 21 x 27 cm
Softcover fadengeheftet
ISBN 978-3-86845-156-6
€ 14,95



Andreas Becker /
Christian Reiß (Hrsg.)
IMRE TÓTH (1921–2010)
*und die Institutionalisierung der
Wissenschaftsgeschichte an der
Universität Regensburg*
Reihe: Schriftenreihe des Universitätsarchivs Regensburg, Bd. 3
1. Aufl. 2021
152 S., 10 Farbabb., 8 s/w-Abb.,
17 x 24 cm, Hardcover Fadengeheftet
ISBN 978-3-86845-166-5
€ 29,95



Andreas Becker (Hg.)
**DIE PROTOKOLLE DES
PROFESSORENKOLLEGIUMS
AM LYZEUM REGENSBURG
1891–1923**
Eine Quellenedition
Reihe: Schriftenreihe des Universitätsarchivs
Regensburg, Bd. 4
152 S. 6 s/w-Abb.,
17 x 24 cm, Hardcover fadengeheftet
ISBN 978-3-86845-170-2
€ 29,95



Caroline-Sophie Ebeling
»SILBER« FÜR JEDERMANN
Die Zinnkunst der Firma Wiedemann
1. Aufl. 2021
352 S., 17 Farbabb., 500 s/w-Abb.,
24 x 28 cm, Hardcover fadengeheftet
ISBN 978-3-86845-165-8
€ 50,00



Caroline-Sophie Ebeling
WEG ZUR FORM
*Die Zinngießert tradition der Familie
Wiedemann*
1. Aufl. 2016
144 S., 17 Farb- und 76 s/w-Abb.
17 x 24 cm, Hardcover fadengeheftet
ISBN 978-3-86845-141-2
24,95 €



Maximilian Lukesch
Michael Mayrock
**GRUNDLAGEN DES
CONTROLLING**
Übungsaufgaben und Lösungen
1. Aufl. 2023
96 S., 17 x 24 cm
Softcover fadengeheftet
ISBN 978-3-86845-172-6
ISBN 978-3-86845-173-3 (E-Book)
€ 10,00



Stephan Weber
**KONSTRUKTION EINER
METHODE FÜR DAS
SYSTEMATISCHE MAPPING IM
COMPLIANCE-
MANAGEMENT**
Reihe: Bankinformationen, Bd. 31
1. Aufl. 2022
517 S., 15,5 x 23,5cm
Softcover klebegebunden
ISBN 978-3-86845-169-6
ISBN 978-3-86845-174-0 (E-Book)
€ 39,95

Universitätsverlag Regensburg



Cordula Böll, Dieter Weiss (Hrsg.)

**50 JAHRE PHYSIK AN DER
UNIVERSITÄT REGENSBURG**

1. Aufl. 2020

176 S., 21 x 28 cm

Hardcover fadengeheftet

ISBN: 978-3-86845-162-7

€ 19,95

