Künstliche Intelligenz: Wie verlässlich ist sie?

Die Entscheidungen selbstlernender Systeme müssen nachvollziehbar sein

von Visvanathan Ramesh

Künstliche Intelligenz (KI), also intelligente Software, führt heutzutage Aufgaben aus, die man einst nur Menschen zutraute. Schon heute ist sie in vielen Bereichen unserer Gesellschaft angekommen – man denke an selbstfahrende Fahrzeuge, medizinische Diagnostik, Übersetzungsprogramme, persönliche Gesprächsassistenten, Suchfunktionen und Robotik. Doch wie weit können wir KI-Systemen vertrauen?

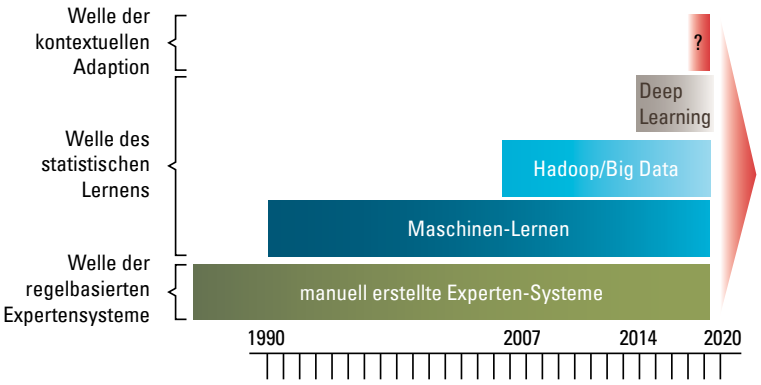
Die Ursprünge der Künstlichen Intelligenz liegen im Jahr 1955, als Prof. John McCarthy ein zweimonatiges Sommerseminar organisierte, mit dem er seiner Idee, denkende Maschinen zu erschaffen, einen Schritt näherkommen wollte. Seine Hypothese war, dass grundsätzlich alle Aspekte des Lernens und anderer Merkmale der Intelligenz so genau beschrieben werden können, dass eine Maschine gebaut werden kann, die diese Vorgänge simuliert. Er regte die Teilnehmer dazu an herauszufinden, wie man Maschinen dazu bringen kann, Sprache zu benutzen, zu abstrahieren und Konzepte zu entwickeln; kurz: Probleme von der Art zu lösen, die zurzeit dem Menschen vorbehalten sind. Außerdem sollten auch diese Systeme sich selbst weiter verbessern können. [1]

Heute hat sich der Begriff Künstliche Intelligenz (Artifical Intelligence) aufgefächert: Um Systeme mit unterschiedlichen Fähigkeiten zu kategorisieren, wird zwischen Artificial Narrow Intelligence (ANI), Artificial General Intelligence (AGI) und Artificial Super Intelligence (ASI) unterschieden. Unter ANI fallen Systeme, die eng gefasste Aufgaben ausführen können. Dazu gehören beispielsweise automatisierte Gesprächsassistenten im Kundenservice, die für einen klar definierten Aufgabenbereich genutzt werden. AGI kann als ein Versuch gesehen werden, Prof. McCarthys ursprüngliche Idee von künstlicher Intelligenz weiterzuverfolgen: nämlich grundsätzliche Strukturen zu entwerfen, mit denen sich menschliche Intelligenz nachbilden lässt.

Mit dem Begriff ASI werden Systeme beschrieben, die die menschliche Intelligenz sogar übertreffen. Das prominenteste Beispiel dafür ist das

1 Beim Go-Spiel ist intelligente Software dem Menschen inzwischen überlegen.

Die drei Wellen der Künstlichen Intelligenz (KI)



KI-System AlphaGo von DeepMind, das Menschen beim Go-Spiel überlegen ist. Obwohl ein solches System vielen vielleicht am unheimlichsten erscheint, weil sie fürchten, dass es, wie der Computer HAL in »2001: Odyssee im Weltraum«, die Macht übernehmen könnte, ist für die Sicherheit eines Systems etwas anderes entscheidend: dass man seine Entscheidungen nachvollziehen kann.

Fragen über den Zustand der Welt beantworten

Ein KI-System kann man als ein System verstehen, das – eingebettet in eine »Welt« – erfasste Daten übersetzt und Fragen zum Zustand der Welt beantwortet. Die Antwort setzt sich aus mehreren Entscheidungen zusammen, z. B. aus den Ergebnissen einzelner elementarer Berechnungen, die das KI-System nacheinander, parallel oder in Kombination anwendet. Dabei können die Antworten durch mehrfache Wiederholung des Prozesses verfeinert werden. Je nachdem, ob das System ein explizites Modell der Welt oder des eigenen Rechenprozesses besitzt, kann es seine Antwort erklären oder nicht. Und das ist entscheidend.

Heutzutage werden für KI-Systeme häufig Entscheidungsbäume oder tiefe neuronale Netze verwendet. Entscheidungsbäume haben den Vorteil, dass sie eine Sequenz von Tests beinhalten, die für Menschen sichtbar und verständlich dargestellt werden können. So kann man nachvollziehen, wie das System zu seiner Antwort gekommen ist. Neuronale Netze hingegen sind nicht inspizierbar, aber ihr Verhalten kann mit entsprechenden Tools visualisiert werden.

KI: Revolution oder Evolution?

Auch wenn im Zusammenhang mit Künstlicher Intelligenz immer wieder von einer Revolution gesprochen wird, handelt es sich um eine über 60 Jahre andauernde Evolution. Diese erfolgte in drei Wellen: [2] In einer ersten Welle entstanden Expertensysteme. Dafür übersetzte man menschliches Expertenwissen in fest definierte Regeln. Diese waren jedoch relativ unflexibel, da sie mit Mehrdeutigkeit und Unsicherheiten der realen Welt nicht gut zurechtkamen. Diese Unzulänglichkeit führte zur zweiten Welle der KI, in der statistische Methoden des maschinellen Lernens zum Einsatz kommen. Man erstellt statistische Modelle, deren Parameter vom KI-System mithilfe großer Mengen von Trainingsdaten und durch Optimierungsalgorithmen gelernt werden. Heute benutzt man tiefe neuronale Netzwerke, um eine Vielzahl eng gefasster Aufgaben (ANI) auszuführen. Der Erfolg der zweiten KI-Welle und damit einhergehend die zahlenmäßige Explosion der Anwendungen wurden in den letzten zehn Jahren insbesondere durch den Fortschritt in den Bereichen Rechenleistung, Trainingsalgorithmen und weltweit verfügbarer Computernetzwerke ermöglicht. Am wichtigsten war jedoch eine stetig zunehmende Menge an gesammelten und gespeicherten Daten, die die Basis des Lernens bilden.

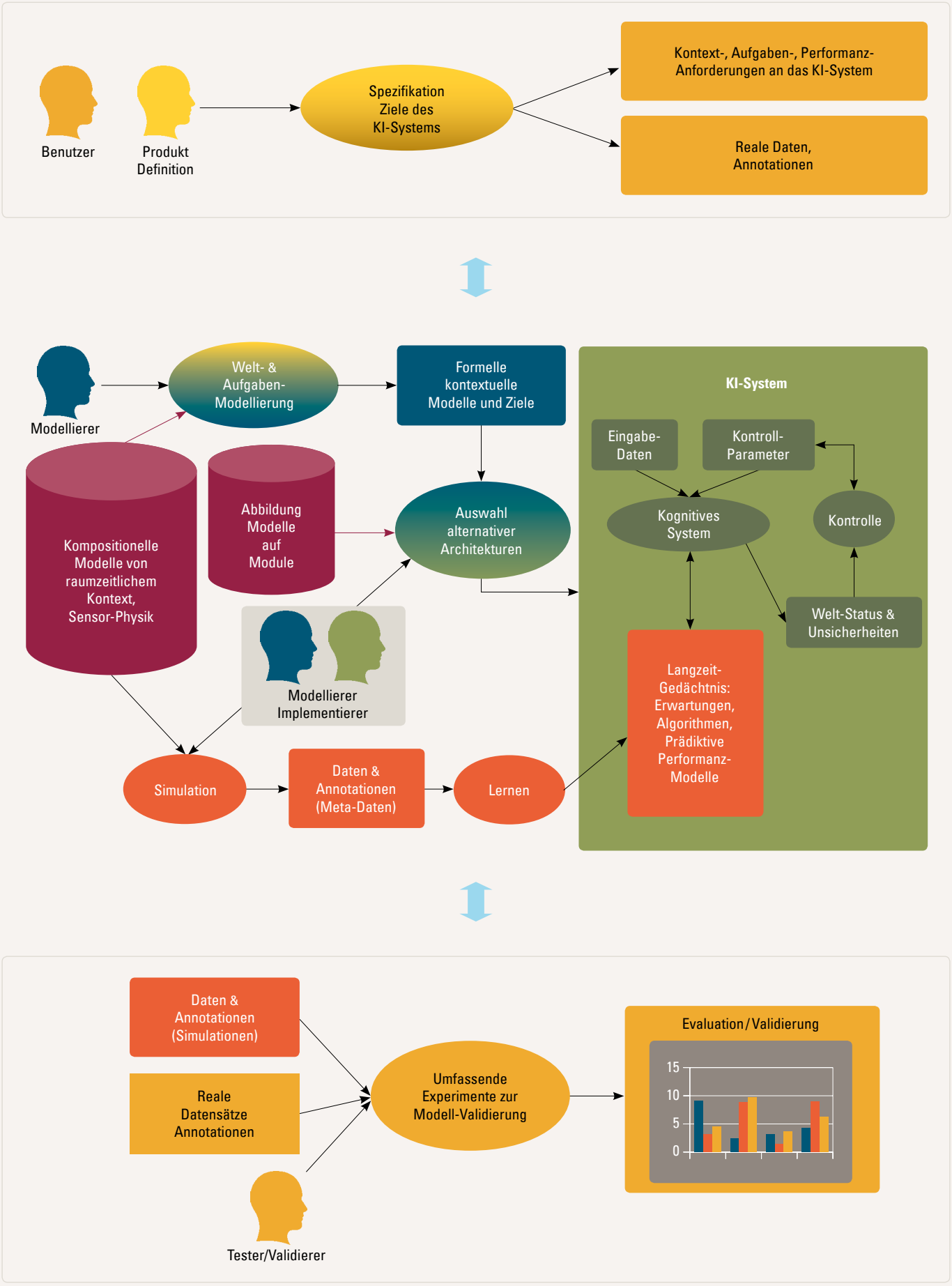
Auch die zweite Welle weist Unzulänglichkeiten auf, die aktuell von der Forschung behandelt werden:

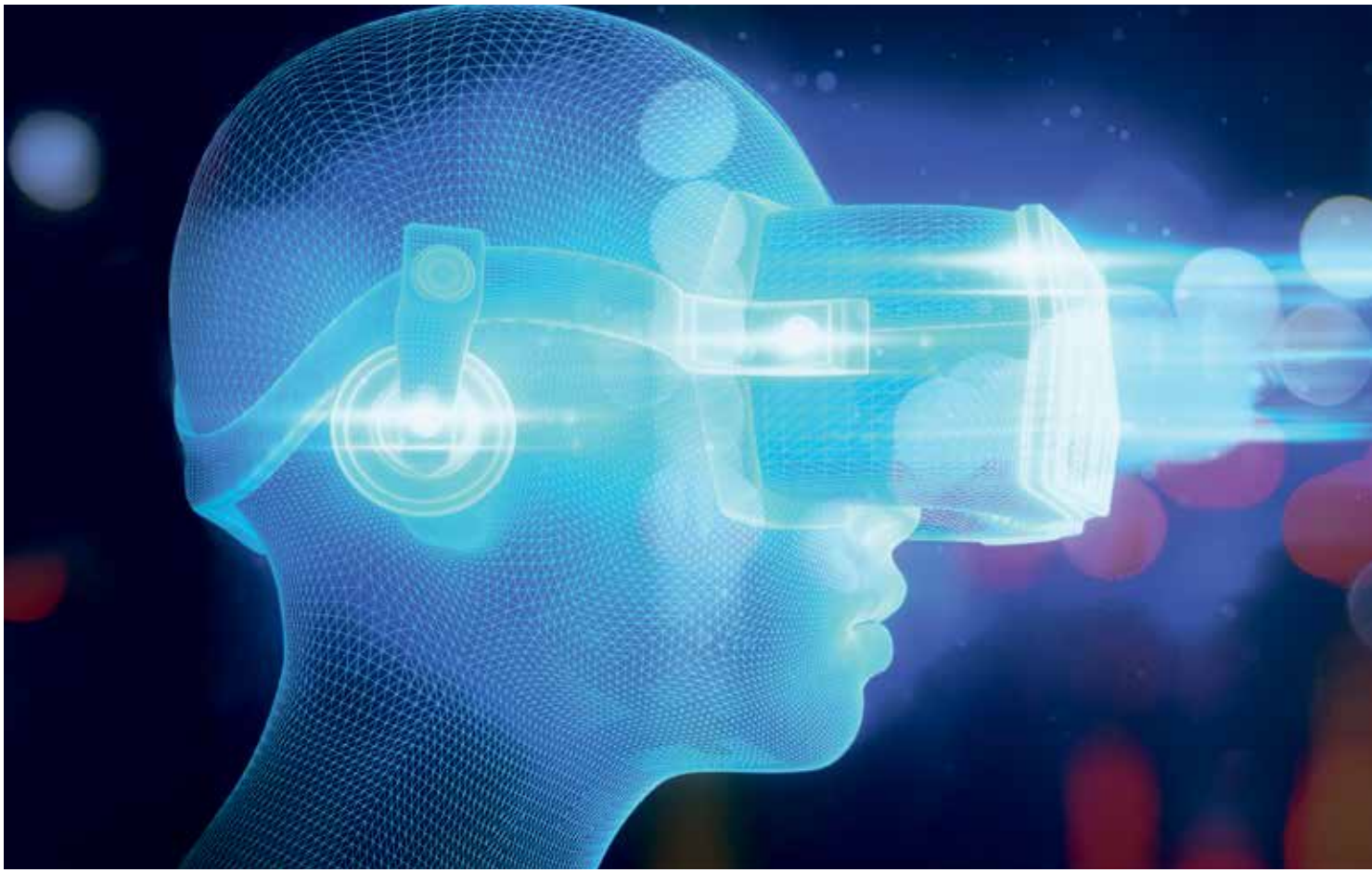
- Systeme der zweiten KI-Welle benötigen große Mengen an annotierten Trainingsdaten, weshalb solche KI-Systeme oft als »datenhungrig« bezeichnet werden.

Die wichtigsten Fragen der KI

- Wie werden die den Entscheidungen des Systems zugrunde liegenden, elementaren Berechnungen ausgewählt?
- Wie kann man diese anordnen, um eine bestimmte Aufgabe in einem gegebenen Kontext zu erfüllen?
- Welche alternativen Architekturen für KI-Systeme sind denkbar und möglich?
- Wie kann man Entscheidungen unter Berücksichtigung des entsprechenden Kontextes treffen?
- Wie kann das KI-System selbst diagnostizieren, dass eine bestimmte Berechnung nicht durchzuführen ist?
- Wie können wir sicherstellen, dass ein KI-System in einer definierten Welt mit festgelegten Grenzen sicher funktioniert?

Entwurfsmethodik für sichere intelligente Systeme





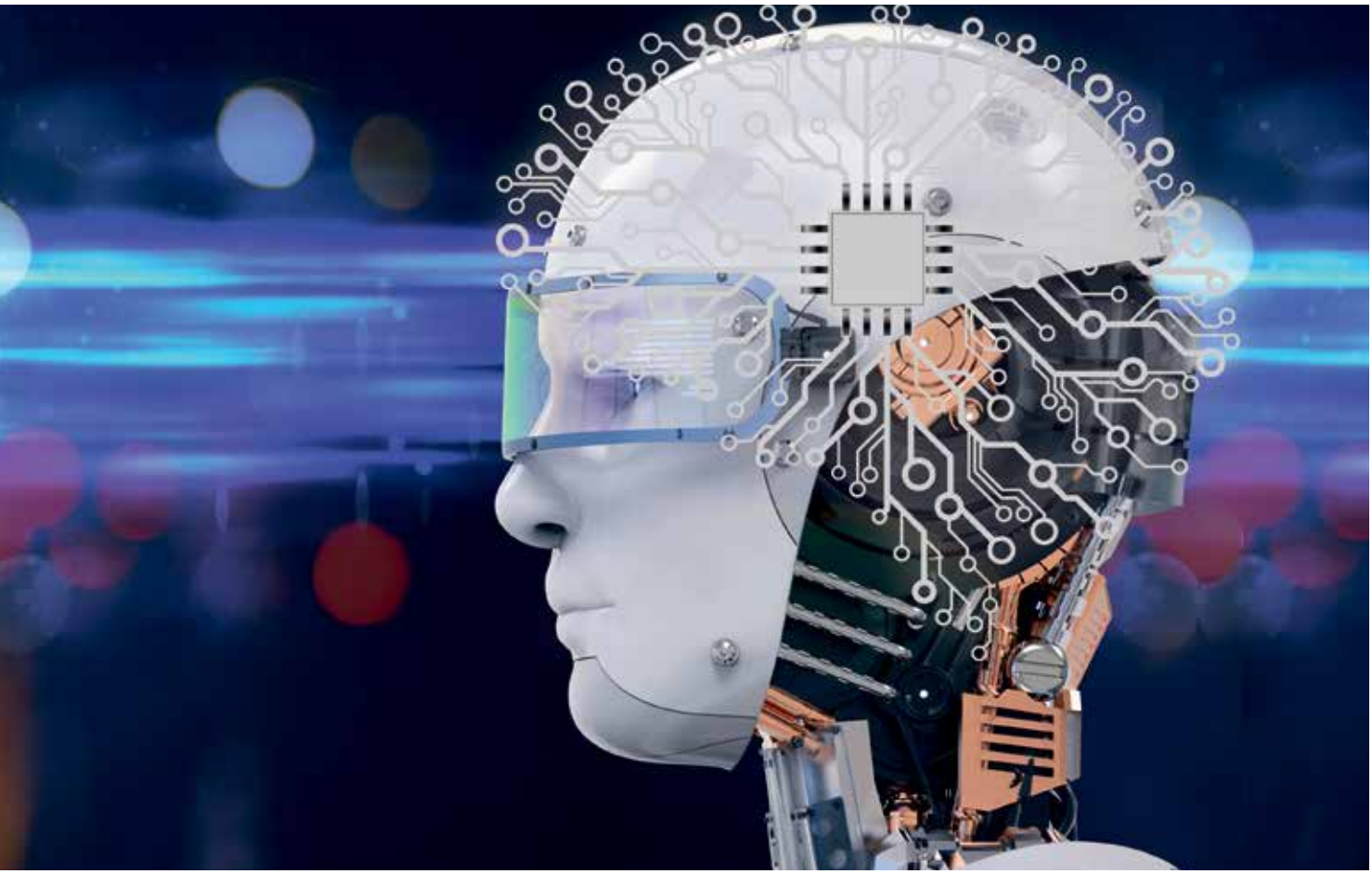
- Systematische Fehler in der Erhebung der Trainingsdaten führen zu Verzerrungen im KI-System.
- Wie und warum ein KI-System zu den Ergebnissen gekommen ist, kann oft kaum erklärt werden.
- Verallgemeinerung, d. h. die Übertragung erlernter Lösungswege von einem Problemfeld auf ein anderes, verwandtes Feld, ist schwierig.

Als logische Konsequenz treten in der dritten Welle, der »kontextuellen Adaption«, die drei Eigenschaften Erklärbarkeit, Argumentation und Abstraktion in den Vordergrund. Gerade bei sicherheitsrelevanten Anwendungen, wie dem autonomen Fahren, mit schweren Konsequenzen im Falle eines Fehlverhaltens, ist es von enormer Bedeutung, genau verstehen und nachvollziehen zu können, warum ein Algorithmus zu einer bestimmten Entscheidung gelangt ist. Das System muss seine eigene Verlässlichkeit selbst einschätzen können, um nötigenfalls auf eine sichere Alternative umsteigen zu können. Die zunehmenden Berührungspunkte zwischen künstlicher und menschlicher Intelligenz (z. B. in Assistenzsystemen, der Mensch-Roboter-Interaktion oder beim autonomen Fahren) erfordern zudem, dass das System mit dem Menschen kommunizieren und seine Annahmen, Motivation und Argumentation nahtlos begründen kann.

Ein Intelligentes Sehsystem – Beispiel für die zweite KI-Welle

Blicken wir zurück ins Jahr 1999. Ich war technischer Manager und leitete das Programm für Echtzeit-Bildverarbeitung bei Siemens Corporate Research in Princeton, New Jersey. Zu dieser Zeit bestand unser kleines Team aus Doktoranden und einigen wenigen wissenschaftlichen Mitarbeitern. Einer unserer großen Kunden hatte ein herausforderndes Projekt: Der Standstreifen einer Autobahn sollte in einem Abschnitt von ca. zwei Kilometern mithilfe eines kamerabasierten Systems überwacht werden. Sobald das System z.B. einen Fußgänger oder ein stehendes bzw. langsames Fahrzeug auf dem Standstreifen identifizierte, sollte ein Kontrollzentrum alarmiert werden. Damit sollte sichergestellt werden, dass der Standstreifen im Falle eines Staus für den Verkehr geöffnet werden konnte.

Das System sollte eine möglichst perfekte Erkennungsrate bei einer minimalen Anzahl an Fehlalarmen aufweisen, unter allen Wetterbedingungen außer schwerem Schneefall und Regen funktionieren und Selbstdiagnosen über die eigene Nichtverfügbarkeit im Fall von extremen Bedingungen (z. B. starke Blendung) stellen. Die Anforderung an das System, diese



AUF DEN PUNKT GEBRACHT

- Die Anwendung von KI-Systemen ist im letzten Jahrzehnt durch verbesserte Rechenleistung, Big-Data, Trainingsalgorithmen und weltweit verfügbare Computernetzwerke nahezu explodiert.
- Wie sicher KI-Systeme sind, hängt davon ab, ob sie für den Menschen nachvollziehbare Entscheidungen treffen und ihre Verlässlichkeit selbst einschätzen können.
- Zukünftige Herausforderung ist die »kontextuelle Adaption« von KI-Systemen. Sie profitiert von transdisziplinären Methoden aus den Ingenieurwissenschaften und der Hirnforschung.

Aufgaben in nahezu Echtzeit mit bis zu vier Kameras pro Computer zu bewältigen, stellte eine weitere Rahmenbedingung dar. Auch wenn das beschriebene System bereits einige Elemente der dritten Welle beinhaltet, würde es heute wahrscheinlich als ANI-System bezeichnet werden. Denn seine Komplexität könnte von einem Kind bewältigt werden, das eine spezifische Aufgabe unter klar umrissenen Umgebungsbedingungen löst.

Meine Aufgabe war es, die Anforderungen unseres Kunden in einen Systementwurf zu übersetzen. Glücklicherweise wurde ich seit den späten 1980er Jahren durch meine Mentoren, Prof. Robert Haralick und Prof. Thomas Binford, inspiriert. Sie untersuchten, wie künstliche Sehsysteme systematisch und prinzipientreu entworfen und analysiert werden können. Zusammen mit Michael Greiffenhagen, einem meiner damaligen Doktoranden, untersuchte ich, wie ich Kontext, Aufgabenstellung(en) und Performanz eines intelligenten Sehsystems so modellieren und analysieren konnte, dass die Anforderungen in einem implementierten System umgesetzt werden können.

Komplexe Systeme sicher entwerfen

Ein entscheidendes Kriterium beim Entwurf komplexer und sicherheitsrelevanter Systeme ist die explizite Aufspaltung in verschiedene Bereiche (Benutzeranforderungen, Modellierung, Implementierung, Validierung) und die Definition der Schnittstellen zwischen ihnen. Dazu werden kontextuelle Modelle sowie die zu erfüllenden Aufgaben und Anforderungen an die Performanz explizit auf algorithmische Strukturen abgebildet. So kann man sicher sein, dass Annahmen transparent kommuniziert und behandelt werden und das System in und zwischen Betrachtungsebenen erklärbar und konsistent ist. Durch die Aufspaltung in einzelne

4 Die Kommunikation zwischen Mensch und Maschine ist entscheidend für die zukünftige Entwicklung der KI. Menschen müssen nachvollziehen können, wie Maschinen Entscheidungen treffen.

Literatur

1 McCarthy, J. et al.: A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955, in: AI Magazine, Jg. 27, Nr. 4, 2006.

2 Binford, T.O. et al.: Bayesian Inference in Model-Based Machine Vision, https://arxiv.org/abs/1304.2720, (Stand: 19.09.2018)

3 Haralick, R. et al.: »DIALOG: Performance Characterization in Computer Vision«, Computer Vision, Graphics, and Image Processing, in: Image Understanding, Jg. 60, Nr. 2, 1994, S. 245-249.

4 Ramesh, Visvanathan: Performance Characterization of Image Understanding Algorithms, Ph.D Dissertation, Department of Electrical Engineering, University of Washington, 1995.

5 Greiffenhagen, M. et al.: Design, analysis, and engineering of video monitoring systems: an approach and a case study, in: Proceedings of IEEE, Jg. 89, Nr. 10, 2001.

6 Kahneman, D.: Thinking, Fast and Slow, Penguin Books, London 2012.

7 Lecun, Y., Bengio, Y., Hinton, G.: Deep Learning, in: Nature, Jg. 521, 2015, S. 436-444.

8 Jahangiri, E., Yörük, E., Vidal, R., Younes, L., Geman, D.: Information Pursuit: A Bayesian Framework for Sequential Scene Parsing, in: CoRR abs/1701.02343 (2017).

9 von der Malsburg, C.: A Vision Architecture, https://arxiv.org/abs/1407.1642, (Stand: 19.09.2018).

10 Veeravasarapu, V. S. R., Rothkopf, C. A., Ramesh, V.: Model-Driven Simulations for Computer Vision, in: Proceedings of the IEEE WACV, 2017, S.1063-1071.

11 Veeravasarapu, V. S. R., Rothkopf, C. A., Ramesh, V.: Adversarially Tuned Scene Generation, in: IEEE CVPR, 2017, S. 6441-6449.

12 Weis, T., Mundt, M., Harding, P., Ramesh, V.: Anomaly detection for automotive visual signal transition estimation, in: IEEE ITSC, 2017, S. 1-8.

5 Beispielhafte 3d-Simulation von Verkehrssituationen. Gegebene Situationen können beliebig oft mit variierenden Umgebungs-Charakteristiken (Beleuchtung, Wetter, Position und Geschwindigkeit der Objekte, etc.) synthetisiert werden. Durch die automatische Generierung von Referenz- oder Ground-truth-Daten können diese synthetischen Sequenzen zum Training und zur Evaluierung von Algorithmen benutzt werden. **a–c**: verschiedene Rendering-Methoden, **e–f**: simulierte Beleuchtungs- und Wetter-Situationen, **d+h**: zugehörige automatisch erstellte semantische Annotationen als Referenzdaten für Evaluierungszwecke.

statistische Tests, die über ein Modell fest mit physikalischen Gegebenheiten in der realen Welt verknüpft sind, wird die Aufteilung eines Systems in seine Subkomponenten gefördert. Dies wiederum ermöglicht, das System zu skalieren, zu adaptieren und zu erweitern. Außerdem bietet es den Vorteil, dass man die System-Performanz nachvollziehen und überprüfen kann (»Safety by design«).

Man analysiert die Systemunsicherheiten auf einer Meta-Ebene: Dabei wird modelliert, wie sich eine Sequenz statistischer Tests als Funktion von Eingabedaten, möglichen Störeinflüssen und Kontroll-Parametern verhält. Dies ermöglicht die Auswahl geeigneter Verfahren für das Lernen und die Optimierung von System-Parametern im jeweiligen Kontext. Diese Meta-Analyse ist eine notwendige Voraussetzung für die Selbstdiagnose und die automatische Adaption von Parametern und ermöglicht die gezielte Auswahl und Ausführung von Explorations- oder Lernverhalten in unbeschränkten Umgebungen.

Ein Schlüsselelement der modellbasierten Entwurfsmethodik ist die detaillierte Modell-Validierung durch empirische Daten. Im obigen Beispiel wurden Videodaten von vielfältigen Szenarien systematisch gesammelt und ein detaillierter Plan für Experimente aufgestellt. Unsere modellbasierte Entwurfsmethodik ermöglichte uns, Systemparameter auch mit geringen Datenmengen zu lernen. Jedoch erlaubte uns erst die große Datensammlung mit vielfältigen Szenarien eine umfassende Validierung. Die Validierung stellt nach wie vor für sehr komplexe intelligente Systeme eine Herausforderung dar. Geeignete Werkzeuge dafür sowie für die Simulation und Modellierung sind unabdingbar.

Unser System hatte am Ende eine Detektionsrate von 98 Prozent, selbst bei sehr geringen Kontrastunterschieden zwischen Fußgängern und Hintergrund, bei nur einem einzigen Fehlalarm pro Kamera und Tag. Und das im Jahr 1999! Auf einem Computer, der in etwa 10 000-mal weniger Leistung hatte als die Rechner, die heutzutage für ähnliche Aufgaben verwendet werden.

Was KI aus der Hirnforschung lernt

Eine große Herausforderung stellt die Skalierung der System-Komplexität dar, die dem System erlaubt, eine breitere Palette von Aufgaben in diversen Kontexten mit menschenähnlicher Performanz oder darüber hinaus auszuführen. Hierzu integrieren wir im Rahmen unserer Methodik zusätzlich zu verwandten Forschungsbereichen (Informatik, Mathematik und Statistik) auch Erkenntnisse und Inspirationen aus unmittelbar fachfremden Forschungsgebieten, die sich jedoch auf verschiedenen Ebenen mit der Analyse menschlicher Fähigkeiten beschäftigen (Neurowissenschaften, Psychologie und Kognitionswissenschaft). Das Hauptaugenmerk liegt hierbei natürlich auf dem menschlichen Gehirn, das als entwickeltes System mit einer flexiblen Lernarchitektur durch die Natur so geschaffen wurde, dass es eine breite Palette von Aufgaben in ähnlichen Umgebungen und Situationen löst.

Obwohl die Sichtweisen vom ingenieurtechnischen und neurowissenschaftlichen Standpunkt vereinbar sind, ist unserer Ansicht nach Skalierung hauptsächlich durch Fortschritte in kognitiven Architekturen, KI-Systemtheorie und Software-Plattformen zu erreichen, die rapide Entwürfe und Validierung dieser Systeme zulassen.

Auf der dritten KI-Welle surfen

Unsere Forschung im Rahmen der Projekte »Bernstein Fokus: Neurotechnologie«, dem EU-

Projekt »AEROBI« sowie diverser Fallstudien dreht sich um die Fortentwicklung und ein besseres Verständnis von Systems-Engineering für visuelle intelligente Systeme. Diese versuchen wir in automatisierte Werkzeuge zu integrieren, die den Entwurf, die Validierung und Zertifizierung von KI-Systemen erleichtern. Wir untersuchen die Rolle grafischer Computersimulationen und der Modellierung beim Entwurf und der Analyse kognitiver Seh-Systeme. Simulationen können eine entscheidende Rolle dabei spielen, das Verhalten alternativer Implementierungen zu validieren und systematische Performanz-Charakterisierung vorzunehmen. Dieser systemorientierte Ansatz ist integrativ und vereint alle Aspekte der drei Wellen.

Des Weiteren generieren wir mit Simulationen synthetische Daten für Methoden des maschinellen Lernens. Im Rahmen verschiedener kognitiver Architekturen kombinieren wir modellbasiertes und datengetriebenes Lernen. In Fallstudien demonstrieren wir, wie Erwartungsmodelle im Kontext benutzt werden können, um mithilfe statistischer Tests kognitive Aufgabenstellungen zu lösen, komplexe Situationen einzuschätzen und Anomalien zu detektieren. Diese Anwendungen beinhalten: Videoüberwachung, Bremslicht-Transitions-Detektion im automotiven Umfeld, die Klassifizierung von Defekten an Bauten wie Brücken und sogar fußballspielende Roboter.

Auf dem Weg zu intelligenten Systemen der Zukunft

Wie können wir den Anforderungen an immer komplexere Systeme und deren Integration in Industrie und Gesellschaft gerecht werden? Auf theoretischer Seite sehen wir die Integration und Harmonisierung von modellbasiertem Systementwurf und modernsten Techniken des maschinellen Lernens als essenziell an. Durch die Entwicklung einer umfassenden Datenbank, die immer größere Bereiche kontextueller Modelle, Anforderungen und Kriterien sowie deren Abbildung auf algorithmische Strukturen umfasst, könnte eine Enzyklopädie von Algorithmen geschaffen werden, die die Basis für intelligente Systeme der Zukunft darstellt.

Die Entwicklung künstlicher intelligenter Systeme wird sich künftig drastisch beschleunigen dank der Ausweitung bestehender Open-Source-Plattformen und Open-Data-Initiativen zusammen mit integrierten Netzwerken in Forschung und Entwicklung. Firmen und Länder müssen die existierenden akademisch-wirtschaftlichen Schnittstellen erweitern. Geeignete Plattformen müssen geschaffen werden, um die Ergebnisse wissenschaftlicher Forschung zeitnah, sicher und verlässlich in die Industrie und Gesellschaft zu integrieren. Ultimativ ist KI eine Disziplin, die aus den vier Ks »Kreativität«, »Kooperation«, »Kommunikation« und »kritischem Denken« sowie Problemlösungsfähigkeiten besteht – also inhärent menschlichen Qualitäten. Zusammen mit »System-Denken« sind dies Schlüsselelemente, die wir als Lehrende an die nächste Generation weitergeben müssen, um sie für ihre Rolle in der neuen Welt der allgegenwärtigen KI zu wappnen. ●

13 Launchbury, J.: A DARPA Perspective on Artificial Intelligence, https://www.darpa.mil/attachments/AIFull.pdf (Stand: 19.09.2018).

14 Pearl, J., Mackenzie, D.: The Book of Why: The New Science of Cause and Effect. Hachette Book Group, New York, 2018.

Der Autor

Prof. Dr. Visvanathan Ramesh, Jahrgang 1962, ist seit 2011 Professor für Software Engineering mit dem Schwerpunkt »Biologisch inspirierte Sehsysteme«. Von 2011 bis 2016 koordinierte er den Bernstein Focus: Neurotechnology an der Goethe-Universität und am Frankfurt Institute of Advanced Studies (FIAS). Bevor er nach Frankfurt kam, war er Bereichsleiter bei Siemens Corporate Research in New Jersey, USA.
Bei seiner Forschung lernt er Unsicherheit als fundamentale Eigenschaft von Datenquellen kennen und weiß, dass sie auch jedem unserer Modelle über uns selbst und die Welt innewohnt. Sich dessen bewusst zu sein, ermöglicht Prof. Ramesh zu würdigen, wie wir etwas Wesentliches herausarbeiten können und dennoch nicht gänzlich sicher sein können über unsere eigene fundamentale Natur und unseren Platz im Universum.

ramesh@fias.uni-frankfurt.de