KÜNSTLICHE INTELLIGENZ

KOSMOS

DURCHBLICK

KÜNSTLICHE INTELLIGENZ

EINFACH ERKLÄRT, SCHNELL VERSTANDEN

BRIAN CLEGG

Aus dem Englischen übersetzt von Volkhardt Müller.
Titel der Originalausgabe: Navigating Artificial Intelligence
erschienen bei Riverside Press
einem Imprint von UniPress Books
Ltd World's End Studios
unter der ISBN 978-1-917226-07-3

BILDNACHWEIS

Mit 55 Illustrationen von Robert Fiszer.

IMPRESSUM

Umschlaggestaltung von Alexandre Coco unter Verwendung
einer Illustration von Robert Fiszer.
Die Illustration zeigt einen Automaton, der ein Gehirn hält.

Mit 55 Illustrationen

Unser gesamtes Programm finden Sie unter **kosmos.de**.
Über Neuigkeiten informieren Sie regelmäßig unsere
Newsletter, einfach anmelden unter **kosmos.de/newsletter**

Gedruckt auf chlorfrei gebleichtem Papier

ISBN 978-3-440-18374-8
Projektleitung: Heiko Fischer
Lektorat: Laila Prota
Gestaltungskonzept: Alexandre Coco
Gestaltung und Satz: TEXT & BILD Michael Grätzbach
Produktion: Markus Schärtlein
Printed in China / Imprimé en Chine

DURCHBLICK

KÜNSTLICHE INTELLIGENZ

EINFACH ERKLÄRT, SCHNELL VERSTANDEN

BRIAN CLEGG

INS DEUTSCHE ÜBERSETZT
VON VOLKHARDT MÜLLER

KOSMOS

EINLEITUNG

Es ist allerhöchste Zeit, dass wir uns mit den realen Auswirkungen künstlicher Intelligenz (KI) befassen. Bereits seit Jahrzehnten wird sie entwickelt, doch erst seit den 2020er Jahren erkennt man ihr wirkliches Potenzial. Die Einführung generativer KI, die sekundenschnell auf Abruf Texte und Bilder produzieren kann, sowie damit assoziierter Technologien wie Roboter und selbstfahrende Autos haben das Thema ins Rampenlicht gerückt. Diese Entwicklung hat das Potenzial, Ökonomien umzukrempeln und unser tägliches Leben zu verwandeln, sie könnte sogar zu einer existenziellen Bedrohung für die Menschheit werden.

Im ersten Kapitel betrachten wir Intelligenz im Allgemeinen sowie die Pioniertage der künstlichen Intelligenz. Die Entwicklung dieser Technologie erfolgte in Wellen, doch erst seitdem KIs auf bestimmte Anwendungen konzentriert werden, ist echter Fortschritt zu verzeichnen. Dem werden wir uns im zweiten Kapitel zur spezialisierten KI zuwenden.

Es war ein früher Erfolg, als es für KI möglich wurde, zumindest teilweise die englische Sprache zu interpretieren, und zwar trotz ihrer Mehrdeutigkeiten. Die Anwendung von neuronalen Netzen, die der Gehirnstruktur nachempfunden sind, hatte dazu beigetragen. Computer konnten nun auch Spiele in Angriff nehmen. Diese Enwicklungen steckten noch in den Kinderschuhen, doch begann man sich zu fragen, ob KIs auch kreativ sein könnten.

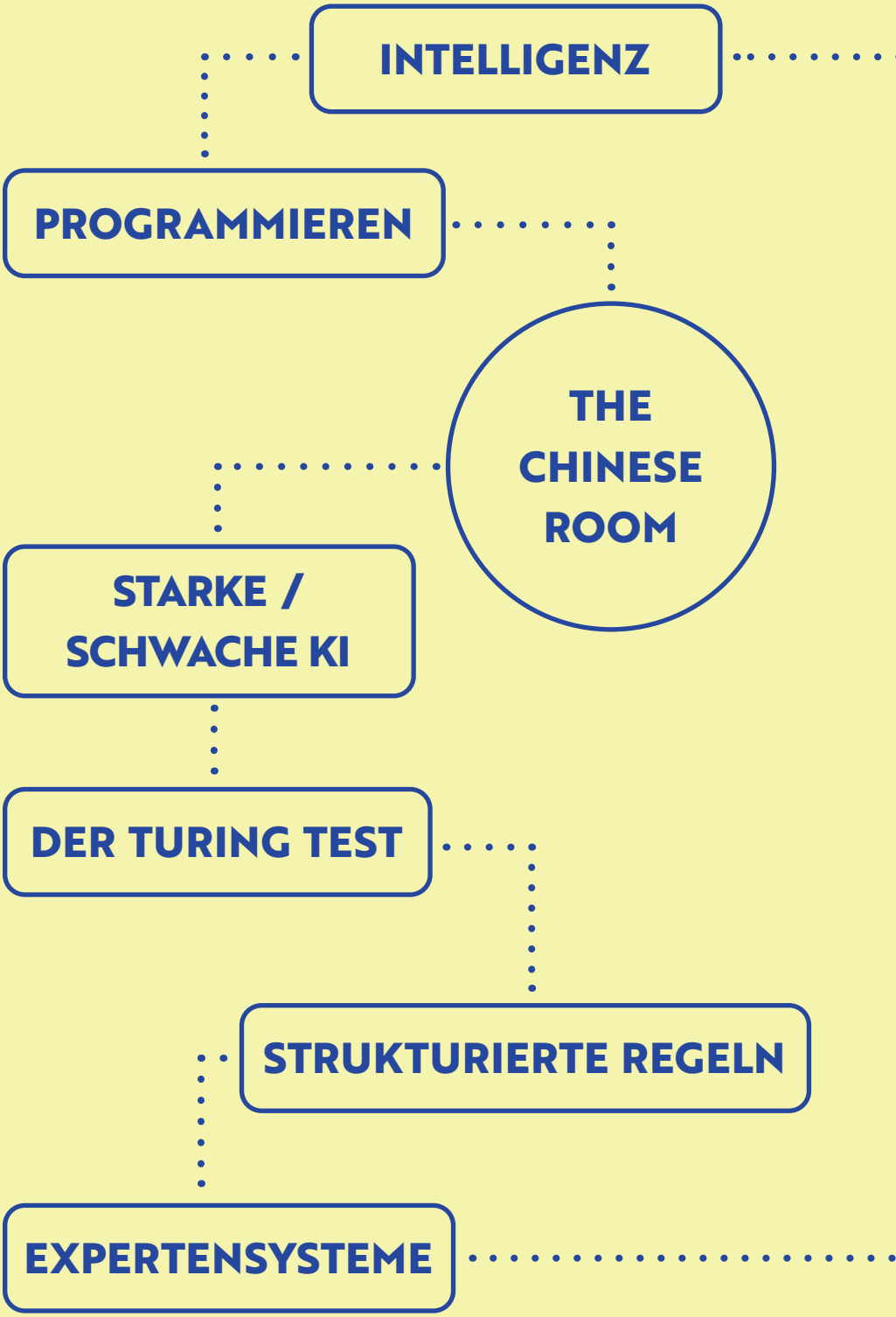
Im dritten Kapitel wenden wir uns dem maschinellen Lernen zu. Neuronale Netze wurden mit riesigen Datenmengen trainiert und dadurch auf die nächste Entwicklungsstufe gebracht.

Der Zugang zu „Big Data" über das Internet machte KIs deutlich wertvoller. Wir betrachten zwei frühe Erfolge des maschinellen Lernens: ein Programm, welches das höchst anspruchsvolle Go-Spiel meistern konnte, und ein anderes, das vorhersagen konnte, wie sich komplexe Proteinmoleküle falten, was für die Entwicklung verschiedener neuer Medikamente unerlässlich ist. Im vierten Kapitel geht es weiter zu den bekanntesten Anwendungen maschinellen Lernens, der generativen KI, die es ChatGPT ermöglicht, Essays und Gedichte zu schreiben, und die DALL-E dazu befähigt, bemerkenswert detaillierte Bilder zu erzeugen.

In Kapitel fünf und sechs erfahren wir, wie sich KI mit zwei wichtigen Hardware-Entwicklungen überschneidet: den Robotern und den autonomen Fahrzeugen. Obwohl enorme Investitionen in selbstfahrende Autos getätigt wurden, stößt diese Technologie an die Grenzen der aktuellen KI-Fähigkeiten. Schließlich widmen wir uns in den Kapiteln sieben und acht der künstlichen allgemeinen Intelligenz und den potenziellen Gefahren, die sie für die Zukunft birgt. Bei der allgemeinen KI beginnen die Fähigkeiten der Software mit denen des Menschen zu konkurrieren oder diese sogar zu übertreffen.

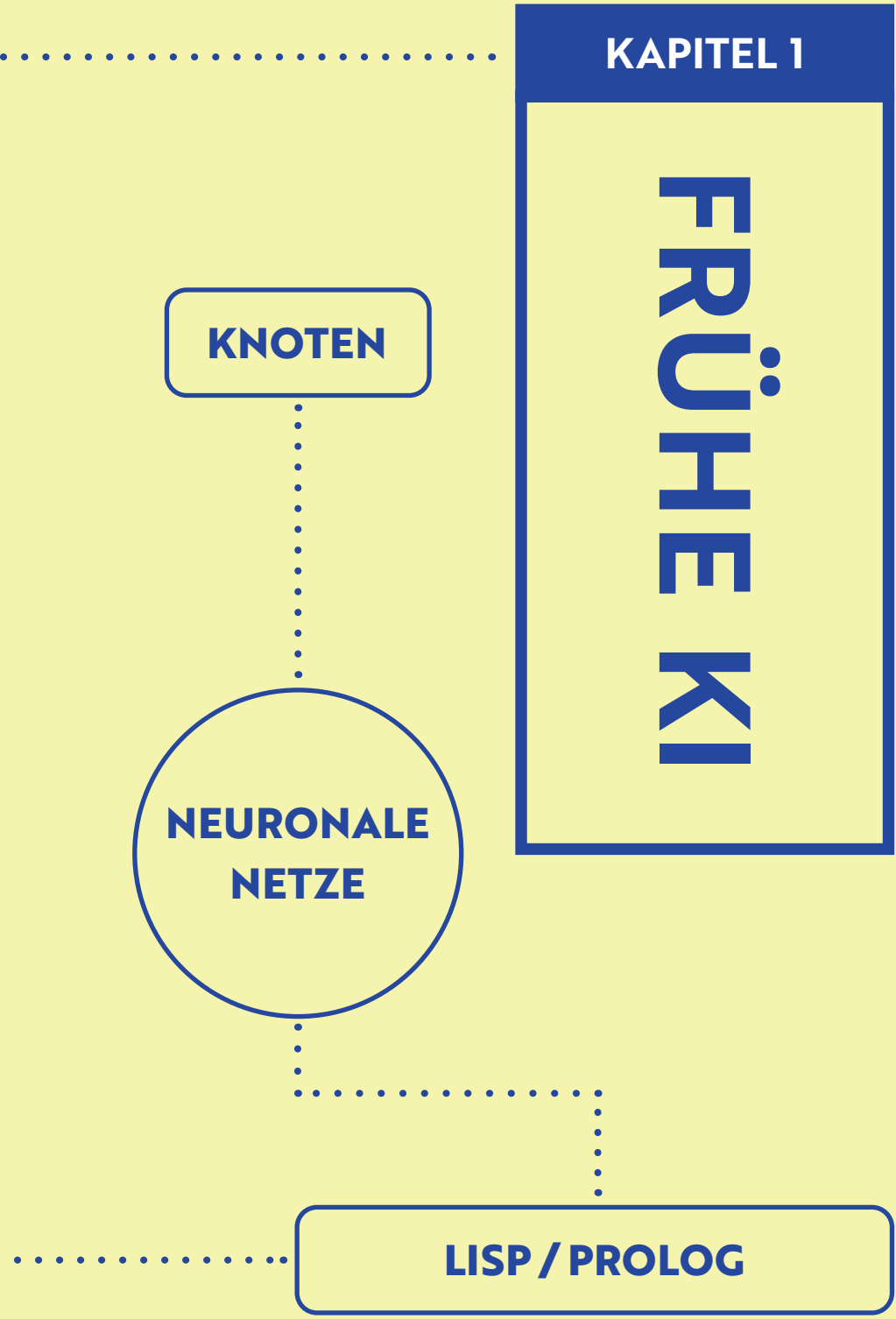
Ob diese Entwicklungen nun bedeuten, dass die Software ein Bewusstsein entwickeln oder gar die Menschheit als Gefahr für ihre Existenz betrachten könnte und dann versuchen würde, uns auszulöschen: Einige Experten sind der Meinung, dass wir sehr genau darauf achten müssen, wie sich die KI weiterentwickelt.

Erkunde mit uns die Antworten auf fünfzig provokative Fragen zur künstlichen Intelligenz!


INTELLIGENZ
PROGRAMMIEREN
THE CHINESE ROOM
STARKE / SCHWACHE KI
DER TURING TEST
STRUKTURIERTE REGELN
EXPERTENSYSTEME

KAPITEL 1

FRÜHE KI



INTELLIGENZ ist ein schwer fassbarer Begriff und darin liegt auch eines der Probleme für das Verständnis der **KÜNSTLICHEN INTELLIGENZ**. Ein Wörterbuch könnte den Begriff „Intelligenz" mit Verständnis in Verbindung bringen und einer verstandesbasierten Herangehensweise an was auch immer. In Anbetracht dessen, wie KI entwickelt wurde und momentan funktioniert, könnte man meinen, dass **VERSTÄNDNIS** und Verstand tatsächlich noch jenseits dessen liegen, was KI leistet. Tatsächlich meinen manche, dass sie künstliche Unintelligenz heißen sollte. Es ist zwar Intelligenz im Spiel, aber oft manipulieren die entsprechenden Systeme nur bereits bestehende Produkte der menschlichen Intelligenz.

Die Wesenhaftigkeit der Intelligenz ist auch ein philosophisches Problem, das zu komplexen **GEDANKEN-EXPERIMENTEN** geführt hat, so wie etwa **JOHN SEARLE**s „Chinesischem Raum".

Dieses klassische Experiment untersucht, ob es ausreicht, wenn ein gänzlich verständnisfreier Mechanismus Intelligenz vortäuschen kann. Ähnlich verhält es sich mit dem vom Computerpionier **ALAN TURING** entwickelten **TURING-TEST**, der lange als Maßstab dafür galt, ob ein System intelligent
ist oder nicht. Tatsächlich erkennt der Test aber nur eine **NACHAHMUNG** von Intelligenz, und er kann diese nicht von echter Intelligenz unterscheiden. Ein Wörterbuch ist ein Beispiel für die Nachahmung eines Aspekts von Intelligenz. Wenn man daraus ein Wort abruft, hat es die „Intelligenz", dieses Wort zu definieren und seine Etymologie zu benennen, doch niemand hält ein Wörterbuch für intelligent.

Zugegebenermaßen behaupten einige Philosophen, dass eine ausreichend gute Nachahmung echte Intelligenz ist, aber diese Ansicht ist umstritten. KI, die auf Nachahmung basiert, wird manchmal als **SCHWACHE KI** bezeichnet. Eine hypothetische KI hingegen, die verstehen und möglicherweise ein **BEWUSSTSEIN** entwickeln kann, wird als **STARKE KI** bezeichnet.

Frühe Versuche, intelligent handelnde Programme zu erstellen, stützten sich auf die Entwicklung sogenannter **EXPERTENSYSTEME**. Die Entwickler dieser frühen KI-Software versuchten, **WISSEN** als Abfolge **STRUKTURIERTER REGELN** zu codieren. Experten eines Fachgebietes wurden einer Reihe von Interviews zur Erfassung ihres Wissens unterzogen. Damit wurden dann Strukturen und Regeln entwickelt, von denen man hoffte, dass sie die Expertise ohne den Experten ausüben könnten.

Leider stieß diese Herangehensweise an Grenzen. Einerseits fiel es selbst Experten schwer, ihr Fachwissen in formatierte Regeln zu packen, andererseits waren sie wenig begeistert darüber, an ihrer eigenen Obsoleszenz zu arbeiten.

In den Anfangstagen der KI versuchten sich Computerwissenschaftler auch an einem ganz anderen Ansatz, der sich eher an der Form des menschlichen Gehirns orientierte als an unserem Verständnis vom Wissen und seiner Struktur. Solche **NEURONALEN NETZE** ersetzten die Grundeinheiten des Gehirns durch elektronische Gegenstücke und verknüpften numerische **KNOTEN** in einem **NETZWERK**. Sie erwiesen sich als interessante Forschungswerkzeuge, waren aber zunächst nur schwer praktisch umzusetzen, sodass sie jahrzehntelang weitgehend ignoriert wurden.

FRÜHE KI – EINE ÜBERSICHT

INTELLIGENZ

Fähigkeit zu verstehen und diesem Verständnis entsprechend zu handeln.

KÜNSTLICHE INTELLIGENZ (KI)
Computerbasierte KI-Technologie, die auf eine Weise handeln kann, die ähnliche Ergebnisse wie die menschliche Intelligenz hervorbringt, insbesondere wenn sie auf große Datenmengen zugreifen kann.

GEDANKENEXPERIMENT
Experiment, das nicht physisch durchgeführt wird (und möglicherweise physisch unmöglich ist), aber dazu verwendet werden kann, ein Problem zu verstehen.

PROZESSE

SCHWACHE KI
KI, die nicht verstehen kann, sondern Intelligenz nachahmt, um ähnliche Ergebnisse wie die menschliche Intelligenz zu erzielen, die jedoch fehlerhaft sein können.

STARKE KI
KI, die im gleichen Sinne wie menschliche Intelligenz verstehen kann. Dies könnte unter Umständen zu einem Bewusstsein führen.

BEWUSSTSEIN
Erkenntnis der eigenen Handlungen, Umgebung und Gedanken, welche eine unabhängige Entscheidungsfindung ermöglicht: der Zustand, der das menschliche Verständnis zu stützen scheint.

EXPERTEN-SYSTEME

Systeme, die versuchten, das menschliche Wissen bestimmter Fachgebiete mittels regelbasierter Strukturen zu reproduzieren, die ähnlich wie ein Mensch auf Fragen antworten.

WISSEN
Höchste Ebene des Trios aus: Daten (z.B. Anzahl der vorhandenen Fische), Informationen (Tabelle der Fischbewegungen) und Wissen (wo und wann gefischt werden soll).

NACHAHMUNG

Die Erzeugung ähnlicher Ergebnisse wie bei einem anderen System, ohne zwingend die gleichen Mittel dazu anzuwenden. Etwas, das sich wie etwas anderes verhält, aber nicht dasselbe ist.

REGELN

NETZWERKE

Mathematische Strukturen, die natzartig aufgebaut sind, mit als Knoten bezeichneten Punkten, die durch Linien miteinander verbunden sind.

KNOTEN
Punkte in einem Netzwerk, denen numerische Werte zugewiesen werden können. Wenn Daten in einem Netzwerk geändert werden, können die Werte an einem Knoten die Werte an verknüpften Knoten beeinflussen, je nach Gewichtung.

JOHN SEARLE, GEB. 1932
amerikanischer Philosoph, der die Natur des Bewusstseins erforscht und bestritten hat, dass eine KI bewusst sein könnte („starke KI"), nur weil sie Verständnis nachahmen kann.

STRUKTURIERTE REGELN
Kombination von Regeln – zum Beispiel: „Wenn der Kunde weiblich ist, dann …" – mit einer Struktur, die es ermöglicht, diese Regeln zur Beantwortung von Fragen oder zur Problemlösung zu einzusetzen.

NEURONALE NETZE
Computersoftware-Informationsstruktur basierend auf einem vereinfachten Modell des Gehirns, mit einer Reihe verknüpfter Werte, die als Gewichte bezeichnet werden und zur Umwandlung von Eingabedaten in Ausgabeinformationen verwendet werden.

TESTS

TURING-TEST

Ein angeblicher Test zur Erkennung menschenähnlicher KI, basierend auf einer Publikation von Turing. Die KI muss ein textbasiertes Gespräch führen, das von einem menschlichen nicht zu unterscheiden ist.

ALAN TURING, 1912–54
Englischer Mathematiker und Logiker, der im Zweiten Weltkrieg eine wichtige Rolle als Codebrecher in Bletchley Park spielte. Turing war ein IT-Pionier und leistete bedeutende Beiträge zur KI.

VERSTEHEN
Bedeutung erfassen, eine Fähigkeit, die Menschen haben, aber KI derzeit nicht. KIs beantworten Fragen durch das Abgleichen von Mustern, haben aber kein Verständnis dafür, was gefragt wird.

Was ist Intelligenz?

→ Intelligenz ist ein Konzept, das ohne ein Maß an Intelligenz nicht definiert werden kann. Ein Wörterbuch mag sie als die Fähigkeit zu verstehen beschreiben, doch das erklärt nichts. Stattdessen könnte man sagen, es handelt sich um die Fähigkeit, Informationen in einen Kontext zu setzen und sie dementsprechend zu nutzen.

Bevor wir über uns über künstliche Intelligenz den Kopf zerbrechen, sollten wir eine Vorstellung davon entwickeln, was Intelligenz überhaupt ist. Dabei kann es hilfreich sein, sich anzuschauen, was „unintelligent" bedeuten kann. Nehmen wir drei Beispiele für mangelnde Intelligenz: das sture Befolgen einer Reihe von Regeln, die nicht zum gewünschten Ergebnis führen; die Unfähigkeit, logisches Denken anzuwenden, und der Mangel an Vorstellungskraft, sich Konsequenzen auszumalen: „Was wäre wenn?" Nehmen wir an, du hättest die Regeln für die Zubereitung eines Kaffees, wonach ein Löffel Instantkaffee mit heißem Wasser übergossen wird und noch ein Spritzer Milch hineingegeben wird. Dank unserer Intelligenz können wir über diese sehr basalen Anweisungen hinaus kontextbezogene Anforderungen hinzufügen und dadurch Pannen vermeiden. Man kann sich ausmalen, wie eine computergesteuerte Kaffeemaschine ohne ein derartiges Verständnis völlig versagt.
Ohne Intelligenz würde sie mechanisch Regeln befolgen, ohne den Kontext zu verstehen. Sie wüsste weder, was sie tut, noch wozu, und ihr einziges Konzept von Erfolg wäre, jeden Schritt der Anleitung abzuschließen. Dabei wird nirgends ausdrücklich eine Tasse erwähnt. Hier kommt die Logik ins Spiel: Kaffee ist eine heiße Flüssigkeit. Daraus können wir ableiten, dass er in ein dichtes und hitzefestes Gefäß abgefüllt werden muss. Und weil wir verstehen, dass wir uns ein Getränk zubereiten, muss dieses auch zum Trinken geeignet sein (zugegebenermaßen tut sich manch menschlicher Designer schwer, das zu begreifen). Die Grundlage unserer intelligenten Herangehensweise ist die Fähigkeit, die Folgen abzuschätzen, potenzielle Risiken zu erkennen und ihnen vorzubeugen. Dies bewahrt uns beispielsweise davor, einen Becher aus gefaltetem Papier zu verwenden.
Manche sagen, dass „künstliche Intelligenz" ein irreführender Begriff sei, da KI weder künstlich noch intelligent sei. Die erste Behauptung scheint fragwürdig – aber der Punkt ist, dass KI ihre Schlussfolgerungen aus menschlichen Informationen und Programmierungen ableitet – sie arbeitet nicht eigenständig. Und wie wir feststellen werden, ist KI bisher alles andere als intelligent, gerade weil sie weder über Verständnis noch über die Fähigkeit verfügt, Kontexte einzubeziehen.

INTELLIGENZ ALS VERSTÄNDNIS

Eine KI-gesteuerte Kaffeemaschine bräuchte mehr als nur ein Rezept für die Kaffeezubereitung. Sie sollte beispielsweise auch wissen, dass eine Tasse bereitstehen muss. Man könnte sie durchaus so programmieren, dass sie dies überprüft, doch der springende Punkt ist, dass wir dank unserer Intelligenz unsere Zielsetzungen verstehen und dadurch Küchenunfälle vermeiden können.

Was verrät uns ein falscher Chinesischsprachiger über Intelligenz?

→ Computer können viele Aufgaben besser erfüllen als Menschen. Z.B. Tabellenkalkulation. Aber deshalb sind Computer noch lange nicht intelligent. Ein Experiment, in dem es darum geht, ohne jedes Verständnis der chinesischen Sprache auf Chinesich zu antworten, verdeutlicht dies.

Um 1980 ersann der amerikanische Philosoph John Searle (*1932) ein Szenario, in dem sich ein System so verhalten kann, als ob es eine Sprache verstünde, ohne dies zu tun. Damals war das noch hypothetisch, spiegelt heute aber unsere Interaktionen mit generativer KI wie ChatGPT wider.
Searle stellte sich einen Menschen vor, der in einem verschlossenen Raum sitzt. Ein Satz auf Chinesisch wird durch einen Schlitz in der Wand gereicht. Der Mensch prüft dann eine Reihe von Aktenschränken, die Anweisungen enthalten, wie auf eine beliebige Kombination von Zeichen zu reagieren ist. Die konstruierte Antwort reicht er oder sie dann durch den Schlitz zurück. (In der Praxis wäre es nicht möglich, brauchbare Anweisungen zu erstellen, aber für die Zwecke des Gedankenexperiments funktioniert es.)
Jemand außerhalb des Raums würde glauben, er hätte sich mit einer Person ausgetauscht, die Chinesisch versteht. Tatsächlich aber versteht die Person im Raum die Sprache nicht und es handelt sich um eine Funktionalität ohne Verständnis. Dies, so argumentierte Searle, sei der Unterschied zwischen menschlicher und maschineller Intelligenz. Der Computer hat keinen Verstand; er denkt nicht. Er ist darauf trainiert, auf eine bestimmte Weise auf eine Eingabe zu reagieren und eine Ausgabe zu erzeugen. Dabei weiß er nicht mehr, als ein Ofen darüber weiß, wie man aus einer Kuchenmischung Kuchen macht.
Searle verwendete dieses Beispiel, um zwischen „starken" KIs zu unterscheiden, die verstehen würden, was geschieht, und „schwachen" KIs, die nur Verständnis simulieren. Man kann davon ausgehen, dass KIs derzeit nach wie vor „schwach" sind und die meisten Fehler auf diese Schwäche zurückzuführen sind.
Searles Chinesischer Raum wurde vielfach kritisiert. So wurde etwa darauf hingewiesen, dass die Person im Raum zwar kein Chinesisch versteht, der gesamte Inhalt des Raums (einschließlich der Aktenschränke) jedoch sehr wohl versteht, denn sonst könnte kein konsistent sinnvolles Ergebnis erzielt werden. Simulationen des Verstehens können zwar zu überzeugenden Ergebnissen führen, doch fehlt es der KI an Flexibilität in ihren Antworten, die nur durch echtes Verständnis ermöglicht wird.

DAS CHINESISCHE ZIMMER

Im Gedankenexperiment kommuniziert eine Person, die kein Chinesisch versteht, schriftlich auf Chinesisch mithilfe von bereits niedergeschriebener Antworten auf verschiedene Symbolfolgen. Da kein Verständnis im Spiel ist, besteht immer die Möglichkeit eines unerwarteten (und unerwünschten) Ergebnisses. Wenn menschliche Intelligenz einige Informationen erhält, weiß sie, worauf Bezug genommen wird, und kann daher die Informationen in einen Kontext setzen und Schlussfolgerungen ziehen, die für eine KI – oder die Person im Raum – nicht möglich sind.

Warum hat Alan Turing einen Test erfunden?

→ Der englische Computerpionier Alan Turing war von der Idee einer maschinellen Intelligenz fasziniert, lange bevor sie zum realen Thema wurde. Im Jahre 1950 entwickelte er einen Test, um herauszufinden, ob Maschinen genauso denken können wie wir.

Alan Turing (1912–54) leistete einen enormen Beitrag zur Informatik, von seiner Arbeit als Codebrecher während des Zweiten Weltkriegs bis hin zur Entwicklung eines konzeptionellen „Universalcomputers". Im Jahr 1950, vier Jahre vor seinem Vergiftungstod (es ist umstritten, ob es sich dabei um einen Unfall oder Selbstmord handelte), schrieb Turing einen Aufsatz mit dem Titel „Können Maschinen denken?".
Er eröffnete mit der Überlegung, ob diese Frage angesichts der problematischen Interpretation der Wörter „Maschine" und „denken" überhaupt angemessen sei, und wählte dann eine alternative Formulierung. Diese ging aus dem sogenannten Imitationsspiel hervor, einem spekulativen Zeitvertreib.
Beim ursprünglichen Spiel kommunizieren ein Mann und eine Frau mit einem Vernehmer, der in einem anderen Raum sitzt. Letzterer versucht nun mittels vorzugsweise schriftlich übermittelter Fragen herauszufinden, wer der Mann und wer die Frau ist. Turing ersetzte eine der beiden verborgenen Personen durch einen Computer. Er wird hierzu zitiert: „Wir stellen nun die Frage: ‚Was geschieht, wenn eine Maschine die Rolle [eines Spielers] in diesem Spiel übernimmt?' Wird der Vernehmer dann genauso oft falsch entscheiden wie bei einem Spiel mit einem Mann und einer Frau?"
Später im Aufsatz fasste Turing wie folgt zusammen: „Sind digitale Computer vorstellbar, die das Imitationsspiel gut beherrschen würden?" Dies wird heute als Turing-Test bezeichnet. Ab den 1950er Jahren gab es Versuche, den Test zu bestehen. Die ersten davon basierten auf Programmen, die dem Computer eine bestimmte Reihe von Regeln für eine Antwort vorgaben. Um wenigstens minimale Erfolgsaussichten zu haben, wurde geschummelt, entweder indem nach Art eines Analysten zurückgefragt oder indem Fragen eingeschränkt wurden. So wurde behauptet, der Chatbot Eugene Goostman aus dem Jahr 2001 habe den Turing-Test bestanden, aber dies gelang nur, indem er vorgab, ein 13-Jähriger mit begrenzten Englischkenntnissen zu sein, und dadurch die Fragen einschränkte.
Solche Versuche wurden später von großen Sprachmodellen (siehe Seite 70) übertroffen – dennoch verfehlt dies das ursprüngliche Ziel. Zweifellos gibt es digitale Computer, die gut im Imitationsspiel abschneiden, doch die Frage „Können Maschinen denken?" wird dadurch nicht beantwortet.

KI BESTEHT DEN TEST

Zweifellos kann heutige KI den Turing-Test mit Textantworten bestehen, die menschlichen Antworten täuschend gleichen, und zwar weit besser, als selbst Turing sich das vorgestellt hatte. Er verglich den Test mit einer mündlichen Prüfung, die etablieren soll, ob etwas wirklich verstanden oder nur auswendig gelernt wurde. Als Beispiel nannte er die Kritik an einem Sonett. Generative KI ist dieser Aufgabe heute problemlos gewachsen. Tatsächlich erfüllt eine generative KI wie Gemini die Aufgabe besser als dereinst Turing selbst.

Folgt Intelligenz strukturierten Regeln?

Die Ersten, die versuchten, eine KI zu entwickeln, glaubten daran – und zwar deshalb, weil sie Computerprogrammierer waren. Programmiersprachen verwendeten Sammlungen strukturierter Regeln, und sie tun es bei herkömmlichen Sprachen immer noch. Von diesen wurde angenommen, dass sie wiederum KI-Systeme bilden könnten.

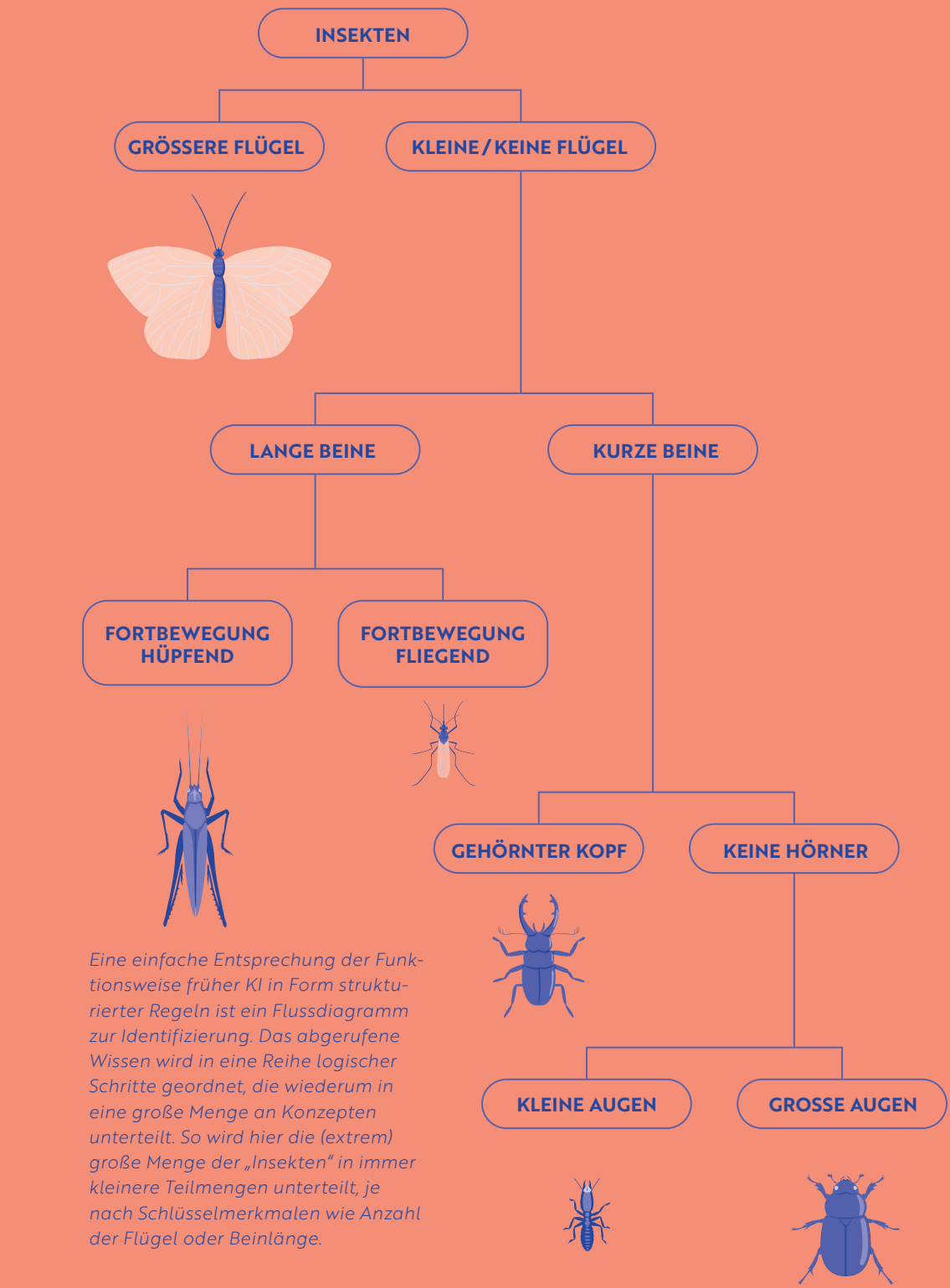
Wenn Software-Ingenieure Programme in nicht-fachspezifischen Sprachen wie C oder BASIC schreiben, besteht ihr Code aus einer Reihe strukturierter, logischer Regeln. Ein gängiges Format in den meisten Programmiersprachen wäre so etwas wie „WENN ... DANN [TUE ETWAS], SONST [TUE ETWAS ANDERES]". Wenn also etwas zutrifft, wird ein Teil des Codes ausgeführt; wenn nicht, ein anderer. Man kann sich das leicht als eine Form von Wissen vorstellen. Man könnte beispielsweise sagen: „Wenn ein Organismus sechs Beine hat, dann ist es ein Insekt; sonst (d. h. wenn es eine andere Anzahl Beine hat) ist es kein Insekt." So entstand die Versuchung, Wissen als eine Reihe von strukturierten Regeln abzufassen.

Zwei KI-Sprachen dominierten das frühe KI-Feld: LISP (von „list processing") in den 1960er Jahren und Prolog ab den frühen 1970er Jahren. LISP wurde entwickelt, um Informationen über Listen zu strukturieren und verwandte Daten über Verzweigungsstrukturen zu verknüpfen. Prolog verwendete formale Logikstrukturen, um Fakten zu identifizieren, und Regeln, um diese Fakten zu verknüpfen. Beide Sprachen waren im Ansatz unterschiedlich, aber im Ergebnis ähnlich.

Vorausgesetzt, die Quelldaten des Programms enthielten die benötigten Informationen, konnte man mithilfe einer sogenannten „Inferenz- oder Schlussfolgerungsmaschine" Fragen stellen und Antworten erhalten, welche die codierte „Intelligenz" widerspiegelten.

Leider hatte dieser Ansatz drei wesentliche Einschränkungen. Erstens ist es alles andere als trivial, die Fakten und Regeln für ein bestimmtes Fachgebiet festzulegen (dies wird deutlicher, wenn wir auf Seite 22 die Frage beantworten: „Kann ein Computersystem Experte sein?" Zweitens müssen solche Systeme für jede Art von Problem komplett neu konstruiert werden, denn das Wissen ist in die nicht übertragbare Struktur eingebaut. Und schließlich passen viele Aspekte von Intelligenz nicht in eine grob vereinfachende, regelbasierte Struktur. Selbst ein Gebiet wie das Rechtswesen, das einem durch und durch kodifizierten Fachgebiet am nächsten kommt, weist Mehrdeutigkeiten und Schlupflöcher auf. Interessanterweise entspricht das Format dieses Buches und der anderen in der Reihe ungefähr einer „Fakten und Fragen"-KI-Struktur: Man hofft, dass die Autoren intelligent sind; die Bücher selbst sind es nicht.

STRUKTURIERTE REGELN BEFOLGEN



Eine einfache Entsprechung der Funktionsweise früher KI in Form strukturierter Regeln ist ein Flussdiagramm zur Identifizierung. Das abgerufene Wissen wird in eine Reihe logischer Schritte geordnet, die wiederum in eine große Menge an Konzepten unterteilt. So wird hier die (extrem) große Menge der „Insekten" in immer kleinere Teilmengen unterteilt, je nach Schlüsselmerkmalen wie Anzahl der Flügel oder Beinlänge.

Kann ein Computersystem Experte sein?

→ Fachwissen kann, wie Intelligenz, schwer zu definieren sein. Grob gesagt bedeutet es, so viel wie möglich über ein Thema zu wissen und dieses Wissen sinnvoll anwenden zu können. Die Antwort ist also ein bedingtes *Ja*, doch erhalten wir einen merkwürdig ungelenken Experten.

Als KI-Systeme erstmals realistisch in Angriff genommen wurden, lag der Schwerpunkt hauptsächlich auf dem, was als „Expertensysteme" bekannt werden sollte. Diese sollten menschliche Entscheidungsfähigkeiten nachahmen. Das Gebiet wurde Ende der 1970er Jahre umfassend untersucht, als die japanische Regierung eine dreijährige Studie über das „Computing der fünften Generation" in Auftrag gab, wie es damals genannt wurde. Das inspirierte staatliche Forschungsvorhaben auf der ganzen Welt. Dabei gingen, wie scheinbar so oft in der Informationstechnologie, alle echten Entwicklungen von der Industrie aus, mit einer Ausnahme: dem Aufbau des Internets.

Mithilfe des Ansatzes der strukturierten Regeln sollten Expertensysteme die Fähigkeiten von Spezialisten eines Fachbereiches nachahmen – egal, ob es sich dabei um ein esoterisches akademisches Fach, Medizin, das Recht oder die Lösung von Problemen mit dem eigenen Computer handelte. Die ersten beiden Schritte für das letzte Fachgebiet bestanden in der Regel in den Fragen: 1. Ist er eingesteckt? und 2. Haben Sie versucht, ihn aus- und wieder einzuschalten?

Die Implementierung von Expertensystemen umfasste drei Hauptphasen. Zuerst kam die Wissenserfassung, gefolgt vom Aufbau einer Wissensbasis und schließlich der Feinabstimmung durch den Abgleich mit realen Beispielen. Jede Phase barg potenzielle Probleme.

An Wissen zu kommen, mag relativ simpel erscheinen, wenn ein menschlicher Experte verfügbar ist, doch die Realität war nuancierter. Das lag einerseits an der verständlichen Zurückhaltung der Experten, die befürchteten, sich selbst um ihre Arbeitsplätze zu bringen. Andererseits fiel es vielen Experten schwer, ihr Wissen in eine geeignete Struktur unterzubringen. Obwohl sie Fachautoritäten waren, kämpften sie damit, ihr Wissen detailliert darzulegen.

Wie wir gesehen haben, war der Aufbau des Systems selbst nicht einfach und seine Effektivität war nicht immer gegeben. Tests ermittelten Probleme am System, jedoch war oft unklar, wie man diese beheben konnte. Eine Reihe von Expertensystemen wurde entwickelt, doch meist beschränkten sich diese auf hochstrukturierte Nischenanwendungen, wie z. B. die Konfiguration neuer Computer, und es wurde sehr still in diesem Bereich – bis zur Entwicklung der maschinellen Intelligenz.

WISSENSERFASSUNG

In der Science-Fiction würde den menschlichen Experten ihr Wissen mittels einer Art Hirnsonde ausgelesen werden. Da sie nicht über derartige Mittel verfügten, Fachwissen elektronisch zu extrahieren, gingen frühe KI-Designer – vielleicht etwas blauäuigig – davon aus, dass Experten ihre Fähigkeiten bereitwillig teilen würden und einen Entwicklungsplan bereitstellen könnten, damit dieses Fachwissen in eine Informationsbank integriert werden kann. Bald sollten sie feststellen, dass die Realität anders aussah.

Was macht ein neuronales Netzwerk?

→ Obwohl der Schwerpunkt der frühen KI auf Expertensystemen lag, versuchten einige Software-Ingenieure, den Mechanismus des Gehirns zu simulieren. Anfangs war ihr Ansatz noch zu strukturiert, denn schließlich handelte es sich ja um Programmierer. Erst als die neuronalen Netze mehr Lernstoff bekamen, wurden sie nützlich.

Die Bausteine eines künstlichen neuronalen Netzes basieren auf den Grundbausteinen des Gehirns: Zellen, Neuronen genannt, die durch elektrochemische Kreuzungen, sogenannte Synapsen, verbunden sind. Die Wechselbeziehung, die neuronale KI-Netze inspirierte, braucht ein „Aktionspotenzial". Inputs in ein Neuron, in Form von geladenen Teilchen, erhöhen oder senken die Spannung an der Synapse. Erreicht sie einen Auslösewert (normalerweise etwa -50 mV), „feuert" das Neuron und setzt Neurotransmitter frei, Chemikalien, die die Synapse zum nächsten Neuron im Netzwerk durchqueren und dessen Spannung beeinflussen.

Das neuronale Netz einer KI simuliert diese Interaktion, jedoch rein digital. Neuronen werden durch „Knoten" ersetzt, die innerhalb eines Netzwerks verbunden sind. Werte von einer virtuellen Schicht dieser Knoten werden an eine andere Schicht weitergeleitet usw. Ursprünglich handelte es sich dabei eher um Spielzeugmodelle mit geringem praktischem Nutzen, die nur zwei Schichten hatten, eine für die Eingabe und eine für die Ausgabe. Deren Effektivität wurde erhöht, indem zwischen den Eingabe- und Ausgabeschichten „verdeckte Schichten" eingeführt wurden, die mehr Möglichkeiten zur Feinabstimmung boten. Eine wesentliche und notwendige Verbesserung, um diese zu nutzen, war die Einführung von „Rückpropagierung" bzw. „Fehlerrückführung" und „Verstärkung".

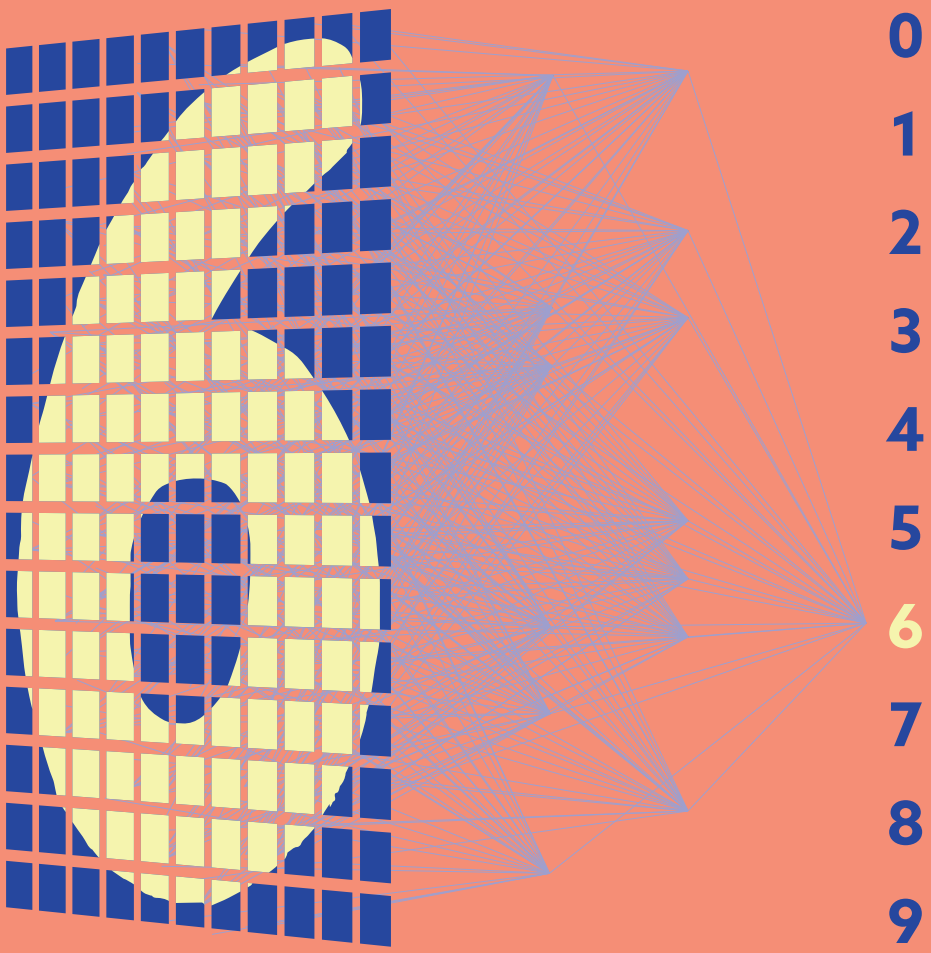
Jede Verbindung zwischen Knoten hat eine Gewichtung, die darstellt, wie stark sie den Wert aus vorigen Schichten weitergibt. Wenn das neuronale Netz Fehler macht, ändert die erstmals in den 1970ern eingesetzte Fehlerrückführung diese Gewichtungen, um das gewünschte Ergebnis besser darzustellen – eine Form der Fehlerkorrektur.

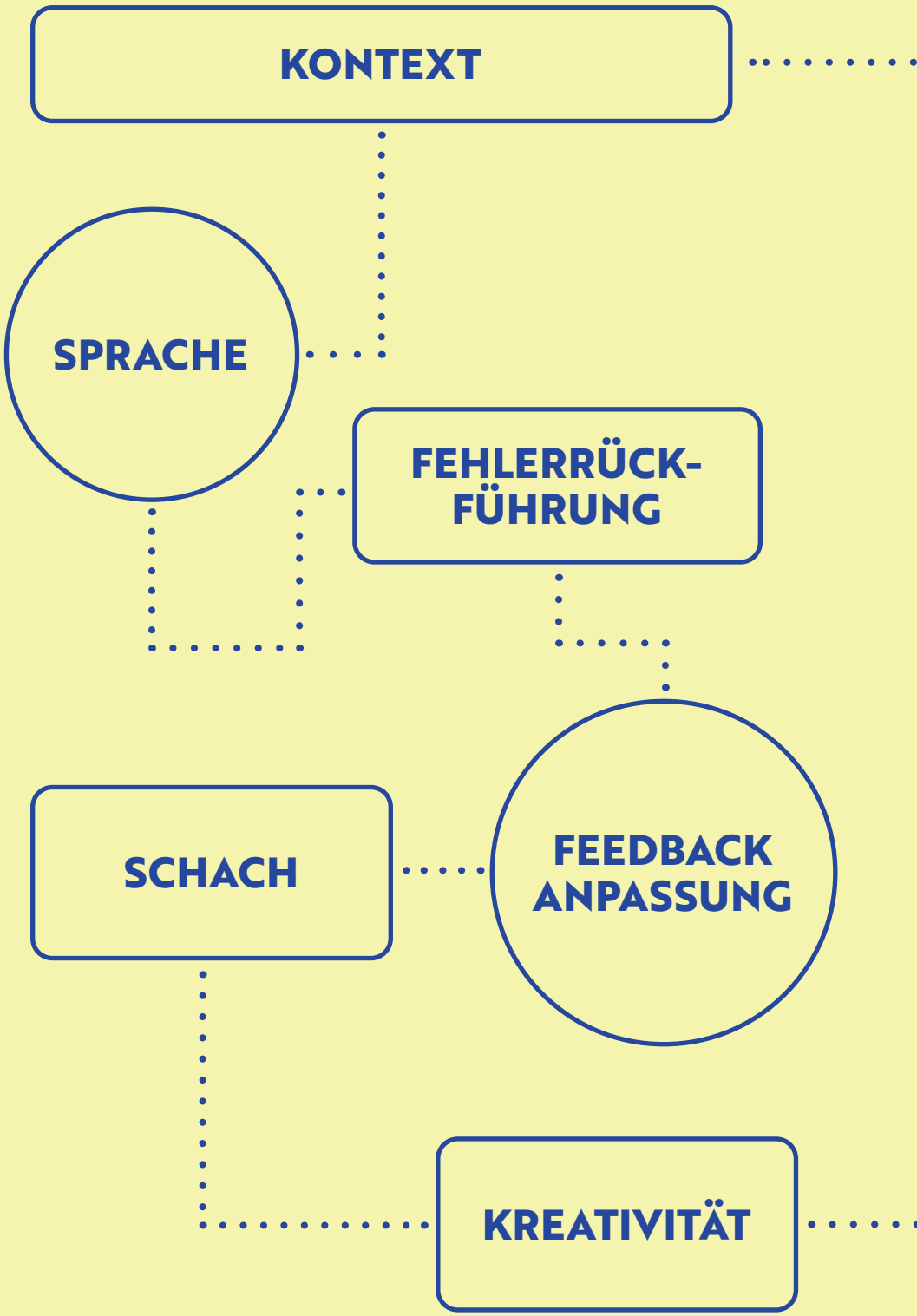
Der Name kommt daher, dass die Gewichtungsänderungen von der Ausgabeseite her über das Netzwerk zurückgeführt werden. Verstärkung ist stattdessen das positive Äquivalent der Fehlerrückführung, bei der die Gewichtungen erhöht werden, wo korrekte Ergebnisse gefunden werden.

Trotz dieser Prozesse war die zweite Generation neuronaler Netze selten nützlich. Erst mit der Entwicklung von maschinellen Lerntechniken unter Verwendung großer Datenmengen, insbesondere durch den Zugang im Internet, konnten große neuronale Netze für eine Vielzahl von Aufgaben trainiert und für bestimmte Aufgaben, wie Bilderkennung, optimiert werden. Zunächst erfolgte dies durch manuelles Akzeptieren oder Ablehnen von Ergebnissen.

TRAINING NEURONALER NETZE

Eine der ersten Aufgaben für einfache neuronale Netze war die Erkennung einer handgeschriebenen Ziffer. Diese wird in ein Feld aus Zellen aufgeteilt, von denen jede den Wert für einen Eingabeknoten festlegt. Nach dem Durchlaufen einer oder mehrerer verdeckter Schichten wird das Netz trainiert, indem ihm eine Vielzahl von Zahlen gezeigt wird, wodurch jene Gewichtungen verstärkt werden, die die richtige Ziffer identifizieren.




KONTEXT
SPRACHE
FEHLERRÜCK-
FÜHRUNG
SCHACH
FEEDBACK
ANPASSUNG
KREATIVITÄT

KAPITEL 2

SPEZIALISIERTE KI

FLIEHKRAFT-REGLER

INTELLIGENTE AGENTEN

SPIEL-THEORIE

Der erste Teilbereich, in dem KI eine breite Anwendung fand, war die **MASCHINENELLE ÜBERSETZUNG**, also die Fähigkeit, einen Text in eine andere Sprache umzuwandeln. Dabei handelt es sich um **SPEZIALISIERTE KI**, also Software, die sich bestmöglich auf eine Aufgabe konzentriert, anstatt eine breite Palette von Funktionen bereitzustellen. Die Pioniere der maschinellen Übersetzung verfolgten den gleichen struktur- und regelbasierten Ansatz wie die Entwickler der Expertensysteme. Dabei waren Experten, die sich der **ERHEBUNG** ihres Wissens widersetzten, das geringere Problem im Vergleich zur Mehrdeutigkeit. Das klassische Beispiel aus dem Englischen sind die scheinbar identischen Strukturen der Sätze „Time flies like an arrow" (Zeit fliegt wie ein Pfeil) und „Fruit flies like an apple" (Fruchtfliegen lieben Äpfel). Die englischen Wortpaare „time flies" und „fruit flies" sowie „like" sind je nach Kontext mehrdeutig. Erst als die maschinelle Übersetzung begann, statt Wörterbüchern und Grammatik **BIG DATA** (Massendaten) zu verwenden, die bereits Unternehmen wie Google zur Verfügung standen, verbesserte sich die Qualität der Übersetzung.

Gleichzeitig begann das Prinzip der neuronalen Netze an Boden zu gewinnen, das die Technik der Fehlerrückführung einsetzte. Dadurch konnte die Software **FEEDBACK** zur Abstimmung ihrer Antworten nutzen und dann durch **VERSTÄRKENDES LERNEN** die Qualität verbessern, auch wenn die Technik noch nicht praxistauglich war. Stattdessen lagen die ersten kleinen KI-Erfolge nicht bei der Maschinenübersetzung, sondern im Bereich der Spiele.

Schachspielende Software faszinierte Computeringenieure so sehr, dass **IBM** den spezialisierten Schachcomputer

DEEP BLUE entwickelte, der gegen Großmeister antrat und den Weltmeister **GARRY KASPAROV** besiegte. Trotz dieses Erfolges war Deep Blue jedoch keine nützliche KI. Er paarte eine große Datenmenge aus früheren Spielen mit der Kapazität, rasend schnell massenhaft Positionen auszurechnen; Fähigkeiten, die nicht einmal Intelligenz nachahmten. Sein totaler Fokus war nicht lediglich auf Spiele beschränkt, sondern galt einzig und allein einem Spiel.

Damals gab es bereits eine als **SPIELTHEORIE** bekannte mathematische Disziplin. Viel davon war dem Universalgelehrten **JOHN VON NEUMANN** zu verdanken, und man könnte glauben, sie sei der ideale Ausgangspunkt für eine Spiele spielende KI. Aber trotz **JACOB BRONOWSKI**s falscher Annahme, dass die Spieltheorie auf Schach anwendbar wäre, geht es darin nicht um „Spiele" im herkömmlichen Sinne, sondern um Situationen, in denen Menschen Entscheidungen treffen, basierend auf ihrem Verständnis oder ihrer Erwartung des Verhaltens anderer. Die Spieltheorie untersucht einen Teilaspekt der Natur der menschlichen Intelligenz, hier war sie jedoch keine Hilfe.

Die menschliche **KREATIVITÄT** wurde in dieser Phase noch nicht ernsthaft herausgefordert. Wie wir noch sehen werden, können Maschinen kreativ sein, doch damals ging man davon aus, dass Kreativität auf lebende Gehirne beschränkt ist. Ein Teil des Problems beim Verständnis von KI ist unsere Tendenz, Intelligenz in anthropomorphen Kategorien zu denken: KI kann am besten als eine Erweiterung von **INTELLIGENTEN AGENTEN** betrachtet werden, die von **ZIELFUNKTIONEN** angetrieben und gesteuert werden, ein Bereich, der seinen Ursprung in den **FLIEHKRAFTREGLERN** des Dampfzeitalters hat.

SPEZIALISIERTE KI – EINE ÜBERSICHT

LÖSUNGEN

INTELLIGENTER AGENT

Einheit (lebendig, mechanisch oder elektronisch), die auf Eingaben reagieren und zielführende Änderungen vornehmen kann, ein hochentwickelter Agent kann aus den Veränderungen in seiner Umgebung lernen.

FLIEHKRAFTREGLER
Mechanische Vorrichtungen, die bei Dampfmaschinen eingesetzt wurden, um die Drehzahl eines Motors automatisch zu regeln und Feedback geben, damit der Motor nicht durchdreht. Die einfachste Form eines intelligenten Agenten.

FEEDBACK (RÜCKKOPPLUNG)
Verwendung eines Systemoutputs zur Beeinflussung oder Steuerung eines Systeminputs, z. B. Mikrofon-Feedback, bei dem der Ton aus den Lautsprechern zu einer Eingabe in das Mikrofon wird.

ZIELFUNKTION
Zielsetzung für intelligente Agenten, um seine Outputs zu steuern und somit etwas Bestimmtes zu erreichen oder zu maximieren.

VERSTÄRKUNG (VERSTÄRKENDES LERNEN)
oder positives Feedback. Im Falle intelligenter Agenten die Nutzung von Feedback (Fehlerrückführung bei KI), um korrekte Ergebnisse zu verstärken, damit sie wahrscheinlicher werden.

SPIELABLAUF

SPEZIALISIERTE KI

KI, die nur ein bestimmtes Problem oder eine Problemklasse lösen kann, anstatt universell einsetzbar zu sein.

MASCHINELLE ÜBERSETZUNG
Verwendung von Computertechnologie zur automatischen Übersetzung von Texten von einer Sprache in eine andere.

MASSENDATEN
Große Datenmengen, die entweder aus proprietären Systemen oder dem Internet stammen und verwendet werden können, um zu versuchen, menschliches Denken nachzuahmen oder Entscheidungen zu treffen.

EINGABE

ERHEBUNG

Bei Expertensystemen handelt es sich um Wissensentnahme von einem menschlichen Subjekt und die Übersetzung dieses Wissens in eine Form, die durch strukturierte Regeln beschrieben werden kann.

KREATIVITÄT
Die ursprünglich ausschließlich Menschen zugeschriebene Fähigkeit, neue Ideen zu entwickeln und Probleme mittels Erfahrungswerten zu lösen.

SPIEL-THEORIE

Mathematische Theorie, die entwickelt wurde, um „Spiele" besser zu verstehen, und zwar im Sinne einfach strukturierter Interaktionen zwischen mindestens zwei Menschen, wo es darum geht, jeweils die Handlungen der anderen vorherzusehen.

DEEP BLUE
Von IBM entwickelter Schachcomputer, der zwischen 1985 und 1997 verschiedene Entwicklungsstufen durchlief und schließlich den Weltmeister Garry Kasparov besiegte.

GARRY KASPAROV (*1963)
Russischer Schachgroßmeister, der zwischen 1985 und 1993 bzw. umstrittenermaßen 2000 Weltmeister war und von DEEP BLUE besiegt wurde.

JOHANN (JOHN) VON NEUMANN (1903–57)
Ungarisch-amerikanischer Mathematiker und Physiker, der maßgeblich zur Entwicklung der Spieltheorie beitrug sowie zur Quantenphysik und dem Atombombenprogramm.

JACOB BRONOWSKI (1908–74)
Polnisch-britischer Mathematiker und Philosoph, der das Feld der Operationsforschung (Operational Research) mitentwickelte und die bahnbrechende TV-Serie *The Ascent of Man* (Der Aufstieg des Menschen) präsentierte.

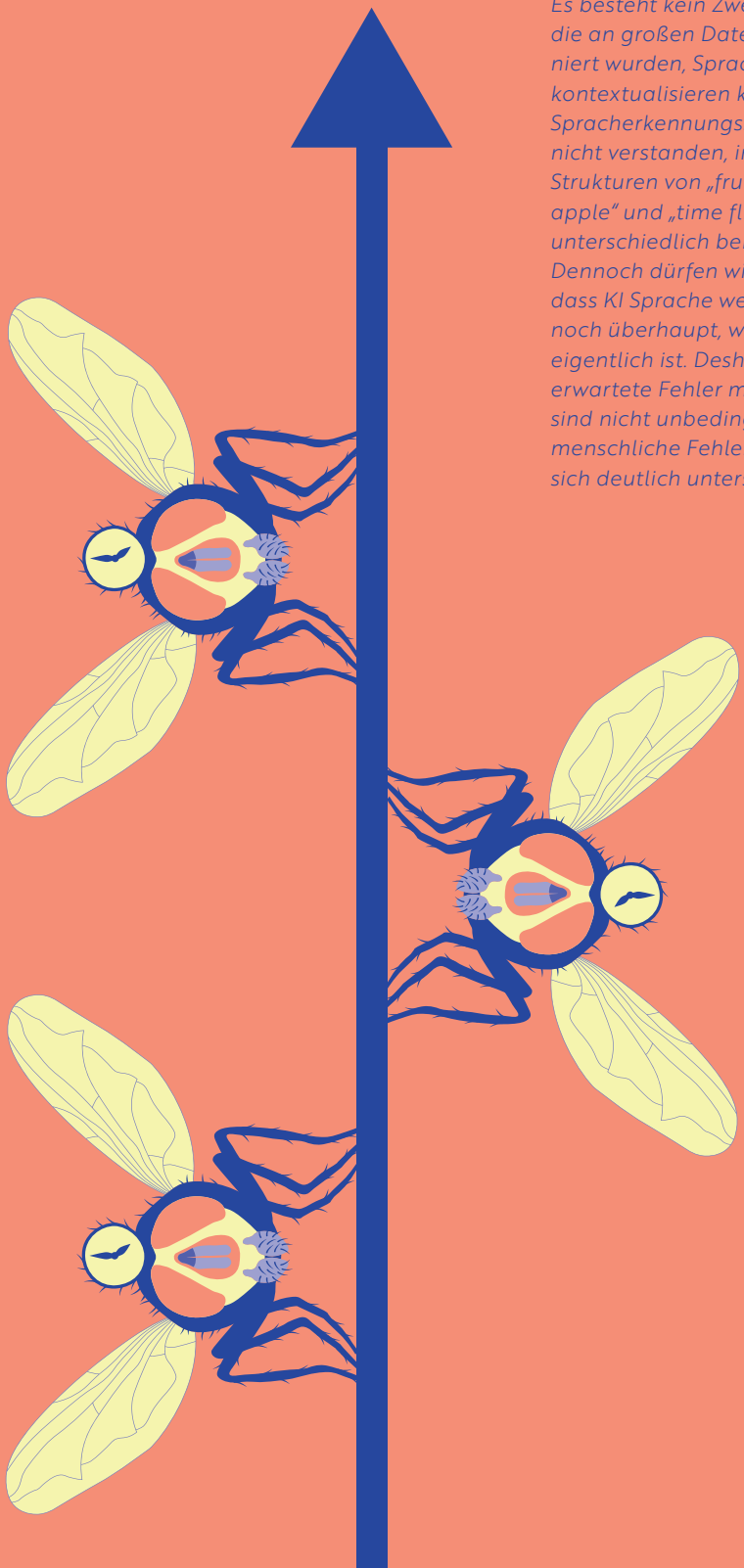
Wodurch unterscheidet sich „time flies like an arrow" von „fruit flies like an apple"?

→ Eines der Hauptprobleme mit KI ist ihre Unfähigkeit, Informationen in einen Kontext zu setzen. Bereits den frühesten Computern versuchte man ein Verständnis der englischen Sprache beizubringen, insbesondere in Bezug auf Mehrdeutigkeiten.

Als KI-Forscher begannen, sich mit dem Sprachverständnis von Computern zu befassen (meist ging es dabei um Englisch), verwendeten sie oft Satzpaare, um die entstehenden Schwierigkeiten zu verdeutlichen. Die Aussagen „time flies like an arrow" (wörtlich: Zeit fliegt wie ein Pfeil) und „Fruit flies like an apple" (Fruchtfliegen lieben einen Apfel) scheinen identische Strukturen zu haben. Das Problem liegt bei den Wörtern „flies" und „like", die in beiden Sätzen jeweils eine andere Bedeutung haben.
Englischsprachige Menschen können diese Sätze dahingehend interpretieren, dass „flies" im ersten Satz ein Verb ist (fliegen) und im zweiten Satz ein Hauptwort (die Fliegen), während „like" im zweiten Satz ein Verb ist und es im ersten Satz eine adverbiale Funktion hat.
Mit dem regelbasierten Ansatz früher KIs war es äußerst schwierig, einer solchen Anomalie Herr zu werden. Sie hätte entweder explizit codiert oder als unlösbar aufgegeben werden müssen.
Erst als man KIs mit großen Datensätzen trainierte, wurden derartige Fälle weniger problematisch. In diesem speziellen Fall gibt es zwar Insekten, die „Fruchtfliegen" heißen, aber keine „Zeitfliegen". Die metaphorische Botschaft des ersten Satzes über das Vergehen der Zeit würde aus dem Sprachgebrauch heraus verstanden werden und nicht als Erklärung dafür, dass „Zeitfliegen" genannte Insekten eine Vorliebe für Pfeile haben.
Der zweite Satz ist problematischer, weil Obst fliegen kann, ob es nun von verärgertem Publikum geworfen oder per Flugzeug aus einem anderen Land eingeflogen wird. Für eine KI, die versucht, diesen Satz zu entschlüsseln, könnte die Rettung darin liegen, dass irgendwelche aerodynamischen Eigenschaften gemeinhin nicht mit „wie ein Apfel" beschrieben werden. Wenn der Satz „wie ein Ziegelstein" gelautet hätte, wäre es problematischer gewesen, da dieser Ausdruck im Englischen oft schlechte aerodynamische Eigenschaften beschreibt. Doch glücklicherweise sind Fruchtfliegen nicht dafür bekannt, dass sie Ziegelsteine lieben.

SPRACHE IM KONTEXT

Es besteht kein Zweifel, dass KIs, die an großen Datensätzen trainiert wurden, Sprache weit besser kontextualisieren können als frühe Spracherkennungssysteme, die nicht verstanden, inwiefern sie die Strukturen von „fruit flies like an apple" und „time flies like an arrow" unterschiedlich behandeln sollten. Dennoch dürfen wir nicht vergessen, dass KI Sprache weder versteht – noch überhaupt, was Sprache eigentlich ist. Deshalb kann KI unerwartete Fehler machen. Diese sind nicht unbedingt schlimmer als menschliche Fehler, aber sie können sich deutlich unterscheiden.



Warum ist Fehlerrückführung so wichtig?

→ Die Fehlerrückführung war der Schlüssel, um die erste große Einschränkung von KIs zu überwinden. Frühe KIs versuchten, Wissen zu kodifizieren, doch die Fehlerrückführung ermöglichte es dem System, zu lernen und sich zu verändern – eine wesentliche Voraussetzung für intelligentes Verhalten.

↘ Auf Seite 24 haben wir gesehen, dass neuronale Netze erfolgreich arbeiteten, nachdem sie begannen, die Fehlerrückführung einzusetzen. Diese war unerlässlich, um die Art von Flexibilität zu erreichen, die Vorraussetzung für auch nur annähernd intelligente Fähigkeiten ist. Wir sehen das Äquivalent der Fehlerrückführung in der gesamten menschlichen Bildung, beginnend mit der grundlegenden Verhaltensschulung für Kleinkinder. Viele KI-Pioniere glaubten, man müsse einfach nur Wissen in ein System gießen – aber wir tun etwas ganz anderes für unsere Kinder. Wir geben ihnen Feedback auf ihre Handlungen, nachdem sie den ersten Input erhalten haben. Wenn sich beispielsweise ein kleines Kind an etwas Heißem verletzt, verstärken wir den Lerneffekt, indem wir darauf hinweisen, wie es in Zukunft solche Schmerzen vermeiden kann. Ebenso geben wir in Schulen nicht nur Informationen weiter, sondern testen auch das Verständnis der Schüler. Obwohl praktische Übungen und Tests manchmal lediglich zur Bewertung der Schüler eingesetzt werden, liegt ihr Wert in der Korrektur von Fehlern und Lücken. Wie effektiv diese ist, hängt von der Qualität der Benotung ab. Eine Richtig/Falsch-Angabe ist nutzlos. Ein konstruktives Feedback hingegen ist wichtiger Bestandteil des Lernprozesses und ermöglicht es dem Schüler, seine Antworten zu verbessern. Diesem Prozess entspricht etwa die Fehlerrückführung im KI-System. Der Ansatz wird meist explizit mit neuronalen Netzen in Verbindung gebracht, deren Ausgabequalität dazu verwendet wird, früheren Netzwerkschichten Feedback zu geben, wodurch wiederum an jedem Knoten die Gewichtungen optimiert werden. Tatsächlich braucht jedes System unabhängig von seiner Struktur eine Art der Fehlerrückführung, um sich an die Realität oder an Umweltveränderungen anzupassen.

Heute mag es merkwürdig erscheinen, dass Lernfähigkeit nicht schon früher als unabdingbar für KI erkannt wurde. Einerseits ist die alte Binsenweisheit, dass das Gehirn wie ein Computer arbeitet, viel zu stark vereinfacht. Sowohl Gehirne als auch Computer verarbeiten Daten, doch ihre Operationsweise und Struktur unterscheiden sich radikal, und man hätte man eigentlich von Anfang an erkennen müssen, dass es unmöglich ist, einem nicht lernfähigen System wirkliche Intelligenz einzuhauchen.