NEUER STANDARD: SO LÄSST SICH KI STRUKTURIERT ENTWICKELN

Rahmenwerk für Künstliche Intelligenz | Um zu nachweislich vertrauenswürdigen Systemen mit künstlicher Intelligenz zu kommen, müssen Entwickler strukturierter arbeiten. Was der neue Standard VDE-AR-E 2842-61 dazu beiträgt, erläutert Dr. Henrik Putzer, der bei Fortiss das Kompetenzfeld Trustworthy Autonomous Systems leitet.

(Bild: Fortiss)

Dr. Henrik Putzer leitet das Kompetenzfeld Trustworthy Autonomous Systems bei der Fortiss GmbH, dem Landesforschungsinstitut des Freistaats Bayern für softwareintensive Systeme. Darüber hinaus ist er Geschäftsführer des Beratungsunternehmens Cogitron GmbH. Beim DKE ist Putzer im Vorsitz der Arbeitsgruppe 801.0.8, welche die VDE-AR-E 2842-61 erarbeitet

IHR STICHWORT

- Künstliche Intelligenz in Systemen
- Beispiel Bildauswertung in der Medizin
- Entwickler muss Rahmen abstecken
- Standard hilft bei Entwicklung der KI
- Unternehmen für Projekte willkommen

Herr Dr. Putzer, in der Anwendungsregel VDE-AR-E 2842-61 geht es um vertrauenswürdige autonom/kognitive Systeme. Was verstehen Sie darunter?
Ein autonom/kognitives System besteht aus vielen Komponenten. Es gibt darin eine Reihe etablierter technischer Bestandteile, aber auch eine künstliche Intelligenz, die in ganz unterschiedlicher Form vertreten sein kann. Um die bekannten technischen Komponenten wie auch Hardware und Software zu entwerfen, gibt es bestehende Normen und Vorgaben. Für eine künstliche Intelligenz hingegen ist der nachweislich korrekte Entwurf noch schwierig, obwohl sie die Vertrauenswürdigkeit des gesamten Systems und damit eben auch seine Risiken stark beeinflussen kann. An dieser Stelle setzen wir mit dem neuen Standard an.

Künstliche Intelligenz soll Daten analysieren, die für das Gehirn zu komplex sind. Was kann ein Standard bewirken?
Künstliche Intelligenz ist ein Thema mit sehr vielen Unbekannten. Die Forschung hat bisher keine umfassenden Erkenntnisse dazu, wie zum Beispiel ein Neuronales Netz im Einzelnen zu seinen erlernten Verhalten und inneren Wissen kommt. Von daher sind wir nicht soweit, dass wir mit einem Standard jedwede Art künstlicher Intelligenz vor allen Fehlern bewahren könnten. Aber wir können mit einem Standard die Entwicklung solcher Systeme in strukturierte Bahnen lenken. Wichtig ist, dass dabei alle relevanten Aspekte und Risiken betrachtet werden. Damit kommen wir einen großen Schritt voran, denn bisher wird künstliche Intelligenz oft sehr kreativ ohne solche Vorgaben entwickelt. KI ist aber nichts Magisches, sondern ein technisches System, für das die Entwickler den Rahmen abstecken können und müssen. Da sehe ich Parallelen zu den Anfängen der Softwareentwicklung. Inzwischen haben wir genaue Vorgaben, wie ‚gute' Software zu entwickeln ist. Für KI werden wir auch dahin kommen. Der Standard ist ein Anfang.

Welche Risiken gibt es denn, wenn eine KI in einem System mitwirkt?
Zunächst muss man sich bei der Begrifflichkeit klarmachen, dass KI und Neuronale Netze (NN) oft fälschlicherweise synonym verwendet wird. KI ist ein Oberbegriff, der den Bereich des maschinellen Lernens einschließt. Darin wiederum ist die KI-Technologie der Neuronalen Netze enthalten. Dabei hängt es speziell von der genutzten KI-Technologie ab, welche Risiken entste-

hen. Nehmen wir als Beispiel die Bildauswertung in der Diagnostik. Wenn ein Neuronales Netz bestimmte Formen von Krebs erkennen soll, wird es mit relevanten Datensätzen trainiert. Es soll vielleicht erkennen, ob es in einer Aufnahme gar keine Indizien für einen Krebs gibt oder ob sich ein gutartiger von einem bösartigen Tumor unterscheiden lässt. Nun können die Unterschiede zwischen Krebsformen minimal sein. Um das abzudecken, muss ich auch seltene Formen in das Training des Neuronalen Netzes einbeziehen – und zwar anteilig in gleichem Maß wie die häufigen Ausprägungen. Tue ich das nicht, leite ich das NN im Lernprozess in die Irre. Ob das Training erfolgreich war, ist dann an Testdatensätzen zu erken-

Testdatensätze zeigen: Schon ein Pixel kann das Ergebnis beeinflussen

nen, in denen die seltenen Formen in geringer Anzahl vorkommen. Erkennt das System dann zuverlässig die häufigen Formen? Wann macht es Fehler?

Was passiert dann bei Fehlern?
Analysiert man die Bilddaten, aus denen das Neuronale Netz die ‚falschen' Schlüsse gezogen hat, muss man anhand der Muster prüfen, was genau zu dem Missgriff geführt hat. Teilweise geht es dabei um die Schattierung eines einzelnen Pixels an einer wichtigen Stelle. Das muss der Entwickler wissen und belegen, dass er solche Überlegungen angestellt hat. Sonst helfen Ergebnisse eines Neuronalen Netzes später nichts.

Welche konkrete Hilfestellung gibt die Anwendungsregel dem Entwickler?
Lassen Sie mich das Ziel des Standards vereinfacht formulieren: Erkläre, warum Du glaubst, dass das, was mit dem System und der KI implementiert wurde, sicher ist. Man muss den Zweck benennen und mit Evidenzen belegen, wie das System funktioniert. Das fordert zum Beispiel auch die FDA im Zusammenhang mit einer KI. Für ein Medizinprodukt wird diese Information in der Sicherheitsakte hinterlegt. Um die Anforderung zu erfüllen, muss man zum Beispiel die Architektur des Neuronalen Netzes herleiten und Datensätze für das Training und die Validierung beschreiben. Wir haben in der Anwendungsregel für jede Phase der Entwicklung zusammengestellt, welche Methoden und Maßnahmen in Frage kommen. Daraus wählt der Entwickler für sein Projekt die relevanten Aspekte aus und begründet, warum er sich im jeweiligen Fall für die eine oder andere Alternative entscheidet. Das betrifft alle Elemente im System, also nicht nur die KI, aber eben auch diese.

Eine KI kann sich durch Training weiterentwickeln. Welche Sicherheiten bietet da eine Entwicklung gemäß den Vorgaben des Standards?
Für eine KI oder ein Neuronales Netz, die im trainierten Zustand eingesetzt werden und sich nicht weiterentwickeln, gelten andere Vorgaben als für ein System, das im Einsatz lernen darf und sich damit verändern wird. Daher sind die beiden Fälle als unterschiedliche Typen von Produkten zu sehen. Für beide Fälle gibt es Lösungsansätze, um zu vertrauenswürdigen Ergebnissen zu kommen. Im Fall der lernenden KI kann das zum Beispiel ein regelmäßiger Check mit einem Testdatensatz sein, dessen Ergebnisse von einem Menschen bewertet werden. Mit den richtigen Verfahren kann eine solche Re-Qualifizierung auch online geschehen. Oder es wird ausgeschlossen, dass exotische Datensätze im Alltag als Basis für eine Weiterentwicklung des KI-Verhaltens verwendet werden können. Diesen Rahmen muss der Entwickler abstecken.

Treffen Entwickler beim Anwenden der Regel auf vertraute Mechanismen?
Grundsätzlich ja. Ein autonom/kognitives System enthält Bestandteile, für die es Normen gibt. Da kann eine Ausfallrate eine Rolle spielen, menschliche Fehler, systematische Fehler, zufällige Fehler, die Alterung von Bauteilen. Das alles ist bekannt. Wichtig zu wissen ist, dass durch künstliche Intelligenz eine neue Art von Fehlern auftauchen kann: so genannte „unsicherheitsbezogene" Fehler, die uncertainty-related faults. Diese entstehen beipielsweise bei Neuronalen Netzen durch die Unsicherheit, wie das NN die Daten komprimiert, welche Features es sich heraussucht und wie es diese zu Ergebnissen aggregiert. Dieses Detail ist zurzeit nicht einmal mathematisch-wissenschaftlich klar darstellbar.

Was bedeutet das Vorliegen der Anwendungsregel für Anbieter, die jetzt schon Produkte mit KI auf den Markt gebracht haben oder neue entwickeln?
Der Standard bietet die Möglichkeit, zu überprüfen, ob alle Aspekte berücksichtig wurden, und eventuell das System zu optimieren – auch nachträglich. Rechtlich verbindlich ist die Anwendung des Standards derzeit nicht. Aber ich denke, er bietet Entwicklern Sicherheit und ist für Unternehmen aus finanziellen Überlegungen sinnvoll. Ein System mit KI soll hinreichend sicher in der Anwendung sein, denn das ist die Voraussetzung für eine Zulassung.

Welche anderen Ansätze gibt es, um eine KI zu bewerten?
Zu diesem Thema läuft derzeit viel. Im Dezember 2020 wurde die Deutsche Normungsroadmap KI vorgelegt. Sie enthält eine Liste aller Aktivitäten in Deutschland. Auch die FDA arbeitet am Thema. Sie berücksichtigt weniger Aspekte als unser Standard, wird dafür im Bereich Medizin aber an manchen Stellen detaillierter. International sind ebenfalls Standards in Vorbereitung, zum Beispiel im ISO/IEC JTC 1/SC 42.

Was empfehlen Sie Unternehmen, die KI nutzen oder künftig einsetzen wollen?
Sowohl Fortiss als auch Cogitron sind sehr offen für die Zusammenarbeit mit Unternehmen in weiteren Forschungs- und Industrieprojekten. Dabei wollen wir den Standard verbessern und ihn möglichst einfach auch für KMU einsetzbar machen. Daher freuen wir uns über neue Partner. Auf lange Sicht können die Überlegungen, die im Standard zusammengefasst sind, auch die Basis für Prüfvorgaben werden, mit denen VDE/DKE, der TÜV oder die FDA testet, ob ein System mit KI vertrauenswürdig ist. Von daher lohnt es sich, die Anwendungsregel schon jetzt im Blick zu haben.

Dr. Birgit Oppermann
birgit.oppermann@konradin.de

Whitepaper zum neuen Standard:
http://hier.pro/RxzAg