
Advertorial

**Qualitätssicherung von AI**

# Qualitätsmerkmale von KI und Robotik testen

*von Humayun Shaukat*

Künstliche Intelligenz und Roboter, die von KI angetrieben werden, werden die ganze Zeit um uns herum sein, ob wir uns dessen bewusst sind oder nicht. KI wird in vielen Anwendungen verwendet, in denen ein Fehler schwerwiegende Folgen haben kann (z. B. den Verlust von Leben oder Eigentum). Solche Systeme müssen systematisch auf Herz und Nieren geprüft werden. Das Testen von Qualitätsmerkmalen eines Systems gehört somit zum professionellen Software-Qualitätssicherungsprozess.

In diesem Artikel werden Sie mehr über die Qualitätsmerkmale der Künstlichen Intelligenz (KI) und Testmethodik erfahren. KI und maschinelles Lernen revolutionieren unsere Welt. Personalisierte Produkte, natürliche Sprachverarbeitung und Gesichtserkennung haben ihren Weg in unseren Alltag gefunden. Das Hauptziel des Artikels ist die Umgrenzung der Metriken zur Messung der Softwarequalität in der Architektur eines künstlichen Intelligenzsystems. Dieser Artikel erleichtert Ihnen den Einstieg in die Qualitätssicherung für KI-betriebene Systeme.

Unsere erstes Whitepaper „Testing of Artificial Intelligence: AI quality engineering skills – an introduction” [AI1] beschreibt, wie man KI testen kann und welche Fähigkeiten Tester brauchen, um KI-betriebene Systeme zu testen.

Im zweiten Whitepaper „Machine Learning Quality Characteristics: How to measure the quality of Artificial  Intelligence and robotics" [AI2] erweitern wir diese Softwarequalitätsmerkmale und stellen die Verbindungen zwischen diesen Eigenschaften dar. Die Kernbotschaft beider Whitepapers ist, dass vorhandene Software-Qualitätsmerkmale nicht ausreichend sind für die Qualitätssicherung von KI-betriebenen Systemen.

## Das Ziel

Ein Verbesserungsziel des Systemengineerings besteht darin, bessere Systeme zu entwerfen und zu entwickeln mit weniger Aufwand. Was heißt bessere Systeme? Dies bezieht sich auf relevante Qualitätsmerkmale. Anhand dieser Eigenschaften können Ingenieure Produkte bewerten hinsichtlich ihrer Stärken und Schwächen. Systemqualität ist definitionsgemäß der Grad, zu dem Systeme eine gewünschte Kombination von Attributen aufweisen. Heutzutage können KI und Robotik verwendet werden, um intelligente Systeme zu entwickeln.

Aber die Verwendung von KI und Robotik führt zu neuen Qualitätsrisiken. Daher ist es wichtig, geeignete Eigenschaften von der bestehenden Liste sowie neue Eigenschaften, die für KI- und Robotiksysteme relevant sind, auszuwählen.

# Die Künstliche Intelligenz

Ich möchte hier zwei Definitionen vorstellen:

> Artificial Intelligence (AI) ist ein Teilgebiet der Informatik mit dem Ziel der Entwicklung von Computern, die Aufgaben ausführen können, die normalerweise von Menschen ausgeführt werden. Insbesondere die Aufgaben, die mit Intelligenz zu tun haben. Künstliche Intelligenz ist einfach ausgedrückt die Fähigkeit von Maschinen, Aufgaben und Aktivitäten auszuführen, die wir als „intelligent" bezeichnen würden. [AI1]

# Robotik

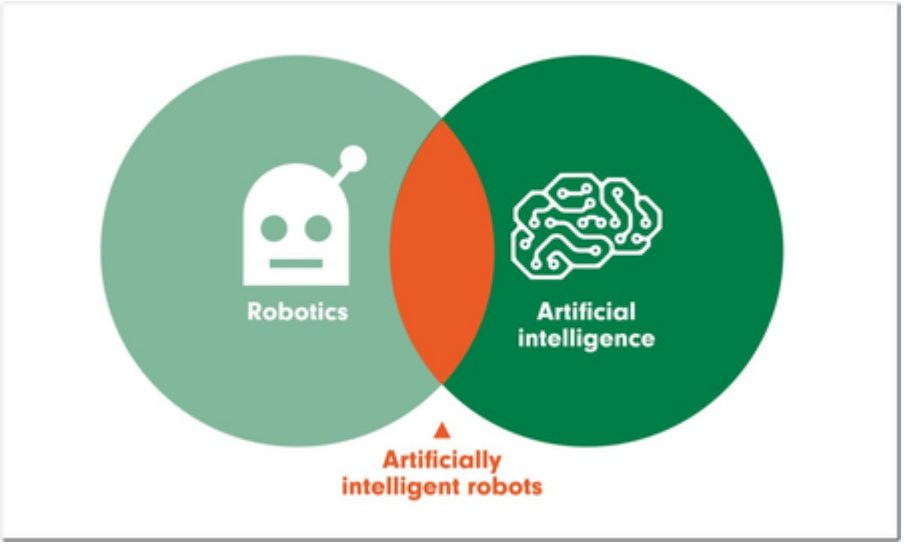


Abb. 1: Die Schnittmenge stellt den künstlich intelligenten Roboter dar.

Robotik ist ein Teil der Technologie, der sich mit Robotern beschäftigt [Sog]. Was aber ist ein Roboter? Er ist eine Maschine, die Informationen über ihre Umgebung mithilfe zum Beispiel von Sensoren sammelt und analysiert und in der Lage ist, ihr Verhalten dementsprechend anzupassen. Kombiniert mit Machine Learning werden die Reaktionen des Roboters im Laufe der Zeit adäquater. Die Nutzung von Internet der Dinge (IoT), Big Data Analytics und Cloud-Technologie macht einen Roboter vielseitig. Roboter gibt es in vielen verschiedenen Formen. Es kann ein Digitaler Agent wie ein Chatbot oder ein selbstfahrendes Auto sein.

# Neue Qualitätsmerkmale für intelligente Maschinen

Die neuen Merkmale und Untereigenschaften, die ich für die Verwendung mit AI und Robotik in Betracht ziehe, werden in der erweiterten ISO25010-Grafik in **Abbildung 2** dargestellt.

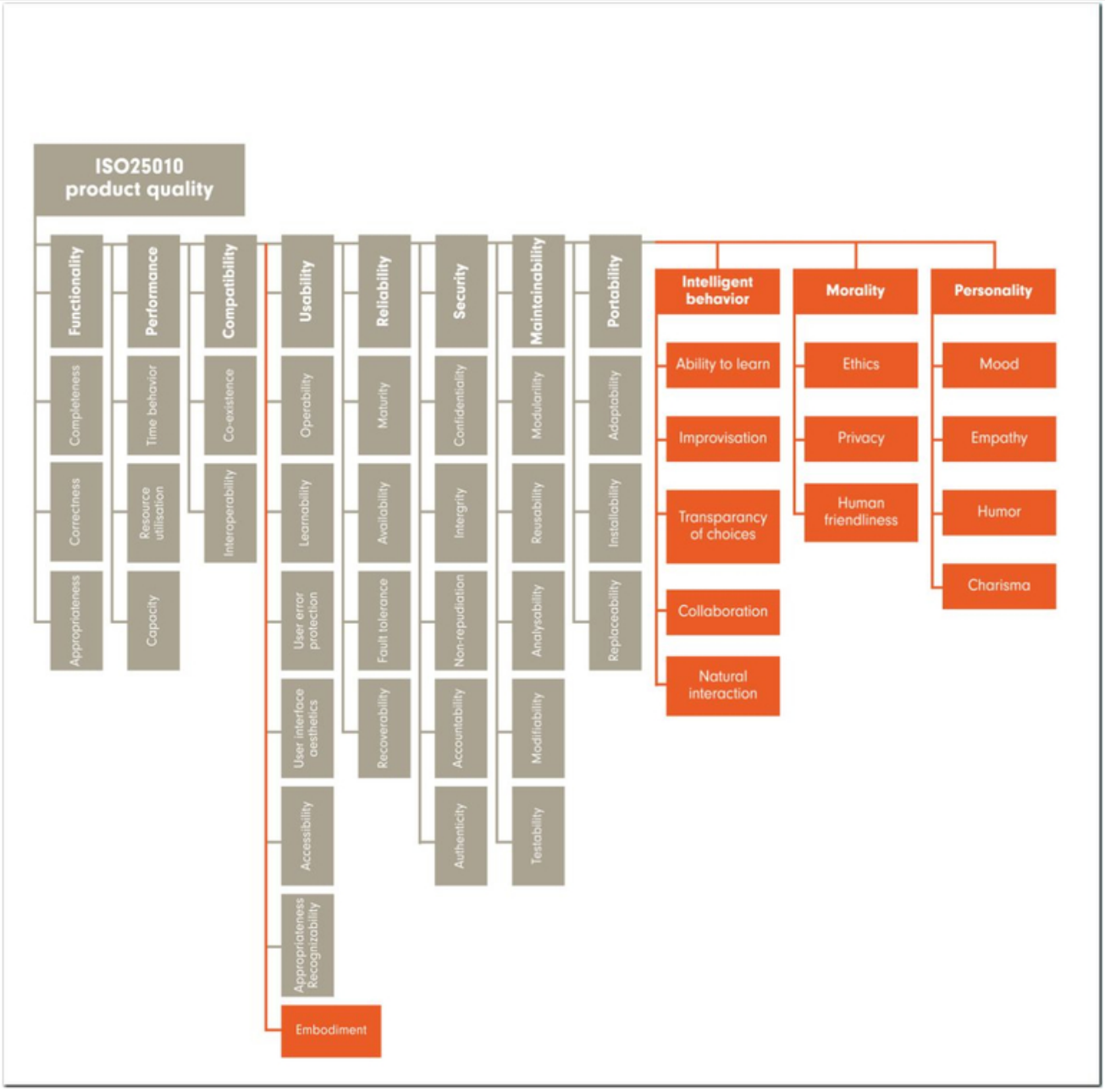

Abb. 2: Die neuen Merkmale nach ISO25010

# Intelligentes Verhalten

Intelligentes Verhalten ist die Fähigkeit zu verstehen [Sog]. Es ist im Grunde eine Kombination aus logischem Denken, Gedächtnis, Fantasie und Beurteilung. Jede dieser Fähigkeiten stützt sich auf die anderen. Intelligenz ist eine Kombination aus kognitiven Fähigkeiten und Wissen, die durch anpassungsfähiges Verhalten deutlicher wird.

**Fähigkeit zu lernen**

Tom Mitchell beschreibt in seinem Buch „Machine Learning“ [Mit13] die Definition für das Erlernen von Fähigkeiten so: "*A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience*".

Die Fähigkeit ist sicher eine messbare Eigenschaft, und wenn man eine Machine Learning-Lösung baut, dann wird vorher überlegt, wie und was die Lösung sein soll, zum Beispiel selbstfahrende Autos: Um den Autos beizubringen, selbstständig zu fahren, benötigt man Machine Learning. Sie müssen unterschiedlichste Szenarien durchspielen und lernen, richtig zu reagieren.

**Nachahmung**

Passt sich das Verhalten an neue Situationen an? Nachahmung ist die Kraft des intelligenten Systems, um in neuen Situationen

die richtigen Entscheidungen zu treffen. Die derzeit erfolgreichste und populärste Methode im Bereich der maschinellen Lernverfahren ist das sogenannte Deep Learning. Dabei handelt es sich um eine spezielle Methode der Informationsverarbeitung.

Deep Learning ist ein Teilbereich des Machine Learning und nutzt neuronale Netze. Zur Herstellung künstlicher Intelligenz werden Trainingsmethoden genutzt, die große Datenmengen heranziehen und analysieren. Das kann man eigentlich auch eine Frage der Perspektive nennen. Nachahmung ist die Stärke des intelligenten Systems, in neuen Situationen richtige Entscheidungen zu treffen. Situationen, die es nie gegeben haben mag, erfordern eine schnelle Interpretation neuer Informationen und die Fähigkeit, bestehendes Verhalten anzupassen.

**Transparenz der Wahlmöglichkeiten**

Kann ein Mensch verstehen, wie eine Maschine zu ihren Entscheidungen kommt? Ein System künstlicher Intelligenz arbeitet rund um die Uhr und trifft viele Entscheidungen. Daher muss es Transparenz darüber geben, wie ein KI-System diese Entscheidungen trifft. Damit könnten Systemanalytiker die Entscheidungen der Maschine analysieren und letztendlich das System verbessern.

**Zusammenarbeit**

Wie gut arbeitet der Roboter neben Menschen? Versteht er das erwartete und unerwartete menschliche Verhalten? Wie Kommunikation innerhalb dieses Teams funktioniert, ist sehr wichtig. Ein Roboter muss sich der Teammitglieder bewusst sein. Mithilfe natürlicher Interaktion muss der Roboter Aufmerksamkeit auf sich ziehen können. Teamarbeit ist besonders wichtig in der industriellen Automatisierung, wo Roboter und Menschen in einer Fabrik nebeneinander arbeiten.

**Natürliche Interaktion**

Natürliche Interaktion ist im Wesentlichen Usability und für viele Lösungen ein wichtiger Punkt. Wie natürlich ist die Sprache eines Chatbots, wie Alexa oder Siri? Natürliche Interaktion ist wichtig, sowohl in verbaler als auch nonverbaler Kommunikation. Vor allem bei sozialen Robotern ist es wichtig, dass die Art und Weise, wie Menschen mit einem Roboter interagieren, natürlich ist und reflektiert, wie sie mit Menschen interagieren.

# Moral

Moral bezeichnet zumeist die faktischen Handlungsmuster, -konventionen, -regeln oder -prinzipien bestimmter Individuen, Gruppen oder Kulturen (Wikipedia). Können wir Maschinen Moral lehren? Eine schwierige, aber nicht unmögliche Aufgabe. Maschinen Moral zu lehren ist schwierig, weil Menschen Moral nicht objektiv in messbaren Metriken vermitteln können, die es einem Computer leicht machen, sie zu verarbeiten.

Bevor wir Maschinen Moral beibringen können, müssen wir zuerst die Moral so definieren, dass Computer sie verarbeiten können. Ingenieure müssen genügend Daten über explizite ethische Maßnahmen sammeln, um AI-Algorithmen entsprechend zu trainieren.

**Ethik**

Bei der Ethik geht es darum, nach verschiedenen Prinzipien zu handeln. Wichtige Grundsätze sind Gesetze, Regeln und Vorschriften. Ethik wird verschiedene Herausforderungen verursachen. Beispielsweise ist es nicht zu schwierig, einer KI beizubringen (mithilfe von Machine Learning), Menschen anhand von Gesichts- oder anderen Körperteilmerkmalen zu unterscheiden. In den meisten Ländern wäre dies aber nicht ethisch. Also müssen Tester Akzeptanzkriterien dafür haben und etwas damit machen.

**Privatsphäre**

*„Privatsphäre bezeichnet den nichtöffentlichen Bereich, in dem ein Mensch unbehelligt von äußeren Einflüssen sein Recht auf freie Entfaltung der Persönlichkeit wahrnimmt“* (Quelle: Wikipedia). Erfüllt die intelligente Maschine die Datenschutzgesetze? Der Brennstoff der Maschinenlernalgorithmen sind Daten. Sie bestimmen, was die Lösung am Ende bewirken kann und wird.

Es ist wichtig sicherzustellen, dass die gesammelten Daten und die aus diesen Daten gewonnenen Erkenntnisse mit den Geschäftszielen übereinstimmen. Es gibt auch rechtliche Beschränkungen, die von nationalen und internationalen Gesetzen abhängen.

**Menschliche Freundlichkeit**

Menschliche Freundlichkeit bezieht sich darauf, dass die intelligente Maschine Menschen oder Lebewesen nicht schädigen darf. Es wird oft befürchtet, dass Roboter und andere intelligente Maschinen die gesamte menschliche Arbeit übernehmen werden. In der Tat wurde im Laufe der Jahre viel menschliche Arbeit automatisiert, aber jedes Mal entstehen neue herausfordernde Aufgaben, die menschliches Eingreifen erfordern.

## Persönlichkeit

Eine Persönlichkeit ist die Kombination von Eigenschaften oder Qualitäten, die den individuellen Charakter eines Individuums ausmachen.

**Stimmung**

Stimmung ist ein vorübergehender Geisteszustand oder Gefühl. Wird eine intelligente Maschine immer in der gleichen Stimmung sein? Eine Maschine weiß nichts über Stimmung, sie erfüllt ihre Aufgabe auf die gleiche Weise wieder und wieder. Aber durch Hinzufügen von Intelligenz kann die Maschine ihr Verhalten auf verschiedene Situationen abstimmen oder auf verschiedene Tageszeiten.

**Empathie**

Empathie ist die Fähigkeit, die Gefühle eines anderen zu verstehen und zu teilen. Maschinen können keine Empathie empfinden. Sie sollten in der Lage sein, menschliche Emotionen zu erkennen und darauf zu reagieren. Dies ist besonders wichtig bei Robotern, die zum Beispiel in Krankenhäusern arbeiten.

**Humor**

Humor könnte als Kunst definiert werden. Witzig zu sein oder die Fähigkeit etwas Lustiges zu finden. Wie werden Roboter diese sehr menschlichen Verhaltensweisen erkennen? Das ist der nächste Schritt in der KI, in dem Roboter mit dieser Fähigkeit programmiert werden. Es gibt einen ganzen Wissenschaftszweig, der sich der Forschung und Entwicklung auf diesem Gebiet widmet.

**Charisma**

Das Wort Charisma bezieht sich auf eine seltene Eigenschaft, die in bestimmten menschlichen Persönlichkeiten gefunden wird. Es ist wie eine „magische" Qualität der Persönlichkeit. Mögen Leute die intelligente Maschine? Lieben die Menschen die intelligente Maschine? Also, das Charisma eines Produktes ist schon wichtig. Ist Charisma ein Zeichen der Intelligenz? Ja, ist es. Es ist alles gelerntes Verhalten.

## Verkörperungen

Das ist ein großes Thema im KI-Umfeld. Die Idee, dass Intelligenz einen Körper braucht. Ein wichtiger Punkt hier ist, dass das Aussehen eines Roboters seine Funktionen reflektiert. Menschen werden zuerst das Aussehen eines Roboters betrachten und ihn bewerten. Ein weiterer relevanter Aspekt der Ausführungsform ist der Grad, in dem ein Roboter einem Menschen ähnelt. Studien belegen, dass die Verkörperung eines Roboters oder Ähnlichkeit zu einem Menschen eine große Rolle spielt.

## KI systematisch testen

Das Testen von Software beansprucht 25 bis 40 Prozent des Entwicklungsaufwands, bei kritischen Systemen liegt der Anteil noch erheblich darüber. Entwickler und Tester stehen häufig vor dieser Herausforderung. Das Testen von KI beziehungsweise KI betriebener Systeme sollte wie folgt erfolgen:

**Dynamisches Testen**

Im Fall herkömmlicher Software: Die effektivsten dynamischen Teststrategien sind Stichproben, funktional Testen und Testen von

Pfaden. In dieser (absteigenden) Reihenfolge. Alle drei Methoden ergänzen jedoch einander und jeder sollte Teil einer umfassenden und intensiven Prüfstrategie sein. Funktionsprüfung erfordert detaillierte Anforderungen und Spezifikationsdokumente und diese sind möglicherweise nicht für das gewünschte AI-System verfügbar.

Die Kernbotschaft hier ist, die Erwartungen an KI-Systeme und Anforderungen sollen partitioniert werden in Serviceanforderungen und minimale und gewünschte Kompetenzanforderungen.

**Einfluss des Konflikts – Lösungsstrategie**

Konfliktlösungsstrategien werden in Produktionssystemen für künstliche Intelligenz eingesetzt.

Die Konfliktlösung wird von einer Vielzahl von Disziplinen, wie Biologie, Wirtschaft, Mathematik, Sozialwissenschaften, Recht und Dialogtheorie, schon untersucht. Informatik und künstliche Intelligenz waren inspiriert von Theorien und Techniken aus diesen Disziplinen. Das hat zu einer Vielzahl von Rechenmodellen und Ansätzen geführt, wie automatisierte Verhandlungen, Gruppenentscheidungen und Mensch-Maschine-Interaktion.

Konflikte treten auf bei der Verkettung von zwei Regeln, die zur gleichen Schlussfolgerung führen. Solche Konflikte können wir mithilfe folgender Strategien lösen:

*Kontextbegrenzung:* Verringerung der Konfliktwahrscheinlichkeit, indem man die Regeln in Gruppen unterteilt, von denen nur einige jederzeit aktiv sind.
*Regeln:* Anordnung aller Regeln in einer priorisierten Liste.
*Daten:* Anordnung aller möglichen Aussagen in einer priorisierten Liste.

**Sensitivitätsanalyse**

Die Sensitivitätsanalyse untersucht, wie sich die verschiedenen Werte einer Gruppe von unabhängigen Variablen unter bestimmten Bedingungen auf eine bestimmte abhängige Variable auswirken. Im vorherigen Abschnitt wurde empfohlen, zu überprüfen, ob die gleiche Eingabe die gleiche Ausgabe produziert. Es ist ebenso wichtig zu bestimmen, ob sehr ähnliche Eingaben sehr unterschiedliche Ausgaben erzeugen können. Wir stellen fest, dass die Sensitivitätsanalyse ein guter Weg ist, um zusätzliches Vertrauen in Software zu gewinnen.

**Statistische Analyse**

Künstliche Intelligenz (KI) ist von Natur aus datengesteuert. Sie erfordert die Anwendung statistischer Konzepte durch Mensch-Maschine-Kollaboration während der Datengenerierung, der Entwicklung von Algorithmen und der Auswertung der Ergebnisse. Datengesteuerte Entscheidungen sind der Kern von KI. Im Informationszeitalter sind Daten nicht mehr knapp. So eine Datenmenge zu durchsuchen oder eine bestimmte Information zu suchen, ist nicht einfach. Da kommen statistische Analyse und Werkzeuge ins Spiel. Durch ständige Datenanalyse und Überwachung werden Fehler in den Daten entdeckt.

**Regressionstests und automatische Testunterstützung**

KI-Systeme können erheblich modifiziert werden während des Tests und der Verfeinerungsphase in der Entwicklung. Es ist daher wichtig, Regressionstests durchzuführen, um sicherzustellen, dass Modifikationen nicht nachteilig sind. Eine automatisierte Unterstützung für das Ausführen und Überprüfen dieser Testfälle ist eindeutig wünschenswert.

# Statisches Testen

Statische Testtechniken dienen hauptsächlich dem Prüfen von Artefakten, wie Anforderungen oder Quelltext, ohne diese auf einem Rechner auszuführen. Wir haben es hier aber mit einem System zu tun, was Künstliche Intelligenz besitzt. Anomalie-Erkennung, mathematische Verifikation, strukturierte Walk-Throughs und Verständnishilfen werden bei KI-Systemen wichtig zu prüfen.

**Erkennung der Anomalien**

Die Erkennung von Anomalien hängt entscheidend von der Redundanz in Programmen ab. Eine Anomalie in einem Programm ist nichts anderes als ein scheinbarer Konflikt zwischen einem Hinweis auf Absicht oder Zweck. Beispielsweise deklariert ein Programmierer eine Variable als ganze Zahl, weist ihm jedoch einen String-Wert zu.

Die Wirksamkeit der Anomalie-Erkennung wird im Wesentlichen durch den Grad der Redundanz und die Struktur in der verwendeten Programmiersprache bestimmt. Wenn KI-Software in einer herkömmlichen, imperativen Programmiersprache (z. B. C oder Ada) geschrieben wird, können die Techniken der Anomalie-Erkennung genauso wie bei herkömmlichen Programmen eingesetzt werden.

**Mathematische Verifikation**

Man kann sagen, Mathematik und künstliche Intelligenz sind zwei Zweige desselben Baums.

Maschinelles Lernen und KI sind aufgebaut mithilfe von mathematischen Prinzipien von beispielsweise Differenzial und Integralrechnung, Lineare Algebra, Wahrscheinlichkeit, Statistik und Optimierung. Ohne diese Prinzipien wäre es nicht möglich, ein KI-System zu programmieren. Daher es ist wichtig, zum Beispiel eine mathematische Verifikation der Algorithmen durchzuführen.

**Strukturierte Walk-Throughs**

Walk-Throughs beinhalten eine informelle, aber sehr detaillierte, manuelle Prüfung des Programmcodes, der Spezifikationen oder Anforderungen. Die Teilnehmer an einem Durchlauf der KI-Software sollten anders sein als für herkömmliche Software. Anstatt vom Moderator, Designer, Implementierer und Tester können wir vom Moderator, Domänenexperten, Knowledge Engineer und Tester sprechen. Die wirklichen Vorteile von Walk-Throughs können erreicht werden, wenn man ein Modell untersucht.

**Verständnishilfen**

Leseverständnis ist eine Fähigkeit, die misst, wie eine Person fähig ist zu begreifen und zu verstehen, was sie liest. Dies zu verbessern, beinhaltet Lesen und jede Menge Übung, um unterschiedliche Verwendung und Wortkontexte zu messen und zu unterscheiden. Dasselbe gilt für künstliche Intelligenz. Beispielsweise lehrt Facebook seine KI, Bücher und andere Kurzgeschichten zu lesen und zu verarbeiten, um das Leseverständnis zu verbessern und logische Verbindungen herzustellen.

## Fazit

KI-Software unterscheidet sich von konventioneller Software in zweierlei Hinsicht: Sie adressiert im Allgemeinen verschiedene Arten von Problemen und funktioniert im Allgemeinen anders als herkömmliche Software. Auf der anderen Seite hat KI-Software viel Gemeinsamkeit mit konventioneller Software. Daher können viele schon bekannte Testtechniken und Strategien von konventioneller Software auch hier angewendet werden.

Aber die sind nicht ausreichend. Sie müssen erweitert werden. Für Qualitätssicherung der KI-Software benötigt man auch besondere Fähigkeiten. Die Fähigkeiten habe ich schon in meinem Whitepaper „Testing of Artificial Intelligence: AI quality engineering skills – an introduction" [AI1] beschrieben.

## Literatur

**[AI1]** Whitepaper Testing of Artificial Intelligence: AI quality engineering skills – an introduction, siehe: https://www.sogeti.com/explore/blog/testing-of-artificial-intelligence/

**[AI2]** Whitepaper Machine Learning Quality Characteristics: How to measure the quality of Artificial Intelligence and robotics, siehe: https://www.sogeti.com/explore/reports/machine-intelligence-quality-characteristics/

**[Mit13]** T. Mitchell, Machine Learning, McGraw-Hill, 2013

**[Sog]** Buchveröffentlichung: „Testing in the Digital Age: AI makes the difference" von T. van de Ven, R. Marselis, H. Shaukat, ICT Book, 2018, siehe: https://www.sogeti.com/explore/newsroom/testing-in-the-digital-age/

## Humayun Shaukat

arbeitet bei der Sogeti als Senior Berater für Digital Testing und KI-Themen. Die Technologie „KI“ ist an sich schon spannend genug und es ist aktuell sein Treibstoff. Das Interesse an neuen Technologien treibt Humayun Shaukat an.

E-Mail: humayun.shaukat(at)sogeti.com

---

### Bildnachweise:

Sogeti