FÜR SICHERHEIT.
JEDEN TAG.

DGZfP-Fachausschuss ZfP 4.0
Arbeitsgruppe Artificial Intelligence
im Unterausschuss Schnittstellen, Dokumentation, Datenformate

Merkblatt ZfP 4.0 – 02

Künstliche Intelligenz für die Zerstörungsfreie Prüfung

Oktober 2024

WER IST DIE DGZfP?

Als technisch-wissenschaftlicher Verein verfolgt die DGZfP das Ziel der Erforschung, Anwendung und Verbreitung der zerstörungsfreien Prüfverfahren. Zu ihren rund 1.600 Mitgliedern gehören große Konzerne und mittelständische Unternehmen, die Zerstörungsfreie Prüfung (ZfP) einsetzen, ebenso Forschungseinrichtungen, Universitäten und Behörden, aber auch einzelne Personen, die sich mit der ZfP beschäftigen.

Die DGZfP organisiert die Kommunikation und den Erfahrungsaustausch zwischen Forschungs- und Entwicklungsinstituten und ZfP-Anwendern, Geräteherstellern und Dienstleistern und informiert über neueste ZfP-Entwicklungen in der Gerätetechnik und den ZfP-Anwendungen.

Zur Lösung spezieller technischer Probleme hat die DGZfP Fachausschüsse eingerichtet. Der praxisnahe Erfahrungsaustausch und die kostenfreie fachliche Weiterbildung finden in regionalen Arbeitskreisen statt.

Die DGZfP veranstaltet regelmäßig Konferenzen und Seminare. Höhepunkt ist die jährliche Jahrestagung mit über 500 Teilnehmenden. Die DGZfP ist Mitglied des EFNDT und ICNDT, der europäischen und der weltweiten Dachorganisation der nationalen ZfP-Gesellschaften.

Weitere Arbeitsgebiete sind die Qualifizierung und Zertifizierung von Prüfpersonal in der Zerstörungsfreien Prüfung. Das Angebot umfasst Schulungen und Qualifizierungsprüfungen nach DIN EN ISO 9712, in allen Produktsektoren, in zehn ZfP-Verfahren und drei Qualifizierungsstufen sowie behördlich anerkannte Ausbildungen im Strahlenschutz entsprechend den Fachkunde-Richtlinien Technik.

Mit Veröffentlichung der europäischen Norm DIN EN 473 im Jahre 1993, ersetzt durch die DIN EN ISO 9712 im Jahr 2013, wurde eine unabhängige DGZfP-Personal-Zertifizierungsstelle (DPZ) eingerichtet.

Die DPZ ist von der Zentralstelle der Länder für Sicherheitstechnik (ZLS) als anerkannte unabhängige Prüfstelle nach der europäischen Richtlinie 2014/68/EU (Druckgeräterichtlinie „DGR“) anerkannt und für den nicht geregelten Bereich von der Deutschen Akkreditierungsstelle GmbH (DAkkS) als Zertifizierungsstelle für Personal der Zerstörungsfreien Prüfung akkreditiert. Weiterhin garantieren Verträge mit vielen Ländern Europas und Ländern in Übersee, sowie die Mitgliedschaft im Multilateral Recognition Agreement des EFNDT, dem DGZfP-Zertifikat internationale Anerkennung.

Der Inhalt von DGZfP-Richtlinien und -Merkblättern ist ein von Experten formulierter Stand der Technik, dessen Anwendung empfohlen wird.

Besuchen Sie uns:

Herausgeber:

Max-Planck-Str. 6, 12489 Berlin
Tel.: +49 30 67807-0 | E-Mail: mail@dgzfp.de

ISBN 978-3-947971-37-4

WEITERE DGZfP-RICHTLINIEN UND MERKBLÄTTER:

(Stand: 10/2024)

A 01	Richtlinie über die Qualifizierung von Prüfwerkern der Zerstörungsfreien Prüfung
A 05	Richtlinie für Ausbildungsstätten – Mindestanforderungen an Ausstattung und Organisation
AT Kom.	Kompendium Schallemissionsprüfung – Grundlagen, Verfahren und praktische Anwendung
B 02	Zerstörungsfreien Betondeckungsmessung und Bewehrungsortung an Stahl- und Spannbetonbauteilen
B 03	Elektrochemische Potentialmessungen zur Detektion von Bewehrungsstahlkorrosion
B 04	Ultraschallverfahren zur zerstörungsfreien Prüfung im Bauwesen
B 05	Merkblatt über das aktive Thermographieverfahren zur Zerstörungsfreien Prüfung im Bauwesen
B 06	Merkblatt über die Sichtprüfung und Endoskopie als optische Verfahren zur Zerstörungsfreien Prüfung im Bauwesen
B 09	Dauerüberwachung von Ingenieurbauwerken
B 10	Merkblatt über das Radarverfahren zur Zerstörungsfreien Prüfung im Bauwesen
B 11	Merkblatt über die Anwendung des Impakt-Echo-Verfahrens zur Zerstörungsfreien Prüfung von Betonbauteilen
B 12	Korrosionsmonitoring bei Stahl- und Spannbetonbauwerken
B 14	Quantifizierung von Chlorid in Beton mit der laserinduzierten Plasmaspektroskopie (LIBS)
Bruchortung	Positionspapier – Magnetische Verfahren zur Spannstahlbruchortung
B-LF 01	Leitfaden zur Erstellung von Prüfanweisungen für die Zerstörungsfreie Prüfung im Bauwesen (ZfP Bau)
D 01	Messung der optischen Dichte in Durchstrahlungsaufnahmen
D 02	Dunkelkammerverarbeitung von Industrie-Röntgenfilmen
D 03	Schweißnahtvermessung bei Zerstörungsfreier Prüfung und Filmkennzeichnung bei Durchstrahlungsprüfungen
D 04	Ermittlung der Prüfbereichsabmessung für die Durchstrahlungsprüfung von Gussstücken
D 05	Vergleichs-Durchstrahlungsbilder für Gussstücke aus Gusseisen mit Lamellen- und Kugelgraphit
D 06	Anforderungen und Rahmenbedingungen für den Einsatz der Röntgencomputertomographie in der Industrie
DP 01	Richtlinie über die Auswahl eines geeigneten Prüfgases für die Dichtheitsprüfung nach DIN EN 13185
DP 02	Richtlinie zur Umrechnung der mit Prüfgasen gemessenen Leckageraten in andere Medien
DP 03	Merkblatt zur Charakterisierung von Prüfgas-Nachweissystemen für Dichtheitsprüfungen
DP 04	Arten von Gasprüflecks und ihre Verwendung bei Dichtheitsprüfverfahren
DP 05	Messunsicherheit und Messmittelfähigkeit bei der Dichtheitsprüfung
EM 06	Merkblatt über Betrachtungsplätze für die fluoreszierende Prüfung mit dem Magnetpulver- und Eindringverfahren – Ausrüstung und Schutzmaßnahmen bei Arbeiten mit UV-Strahlung
HB PA	Handbuch für die Materialprüfung mit Ultraschall Phased Array
ISB 02	Zustand der Eisenbahnfahrzeuge – Verfahren zur Bestätigung der Kompetenz einer ZfP-Prüfstelle nach DIN 27201-7 durch eine dritte Seite
ISB 03	Zustand der Eisenbahnfahrzeuge – Validierung und Überwachung von mechanisierten bzw. automatisierten Prüfanlagen in ZfP-Prüfstellen
MC 01	Richtlinie für Kriterien zur Auswahl von Härteprüfverfahren mit mobilen Geräten
MR 01	Metrologische Rückführbarkeit von Hilfsgeräten für die Eindring- und Magnetpulverprüfung
MTHz 01	Mikrowellenprüfung: Grundlagen und Anwendungen
OV 01	Merkblatt über Optische Verfahren – Auswahl und Erstprüfung von optischen Sichtprüfgeräten; Überprüfung von optischen Sichtprüfgeräten durch den Anwender
SE 02	Verifizierung von Schallemissionssensoren und ihrer Ankopplung im Labor
SE 05	Detektion von Spanndrahtbrüchen mit Schallemissionsanalyse
SHM 01	Strukturprüfung mit geführten Wellen als Sonderform des Ultraschalls
US 06	Akustische Resonanzverfahren zur Zerstörungsfreien Prüfung
US 07	Richtlinie zur Festlegung des Prüfrasters bei der automatisierten Ultraschallprüfung großer Schmiedestücke
US 08	Charakterisierung und Verifizierung der luftgekoppelten Ultraschallprüfköpfe
US 08 E	Characterization and verification of air-coupled ultrasonic probes
ZfP 01	Richtlinie Zerstörungsfreie Prüfung entsprechend ASME Boiler and Pressure Vessel Code
ZfP 4.0 – 01	DICONDE in der ZfP

Informationen unter: www.dgzfp.de

INHALTSVERZEICHNIS

Vorwort

Die Künstliche Intelligenz (KI) treibt eine tiefgreifende Transformation voran, die nahezu alle Industriezweige erfasst hat. Die zerstörungsfreie Prüfung (ZfP) bildet hierbei keine Ausnahme. In einer zunehmend digitalisierten Welt eröffnet die Integration von KI in die ZfP bahnbrechende Möglichkeiten. Diese Technologie hat das Potenzial, unsere Prüfverfahren nicht nur effizienter und präziser zu gestalten, sondern auch völlig neue Ansätze zu erschließen, die bisher undenkbar waren.

Dieses Merkblatt, verfasst von der Arbeitsgruppe für KI der Deutschen Gesellschaft für Zerstörungsfreie Prüfung (DGZfP), soll Ihnen einen umfassenden Einblick in die Anwendung von KI in der ZfP geben. Unser Ziel ist es, Ihnen die grundlegenden Prinzipien dieser Technologie näherzubringen und die Potenziale für Ihre tägliche Arbeit verständlich zu machen.

Die Bedeutung von KI in der ZfP kann nicht genug betont werden. KI-gestützte Systeme bieten die Fähigkeit, enorme Datenmengen in kürzester Zeit und mit höchster Präzision zu analysieren. Dies eröffnet uns neue Dimensionen in der Fehlererkennung und Qualitätskontrolle. KI kann Muster identifizieren, die für das menschliche Auge unsichtbar bleiben, und dadurch die Zuverlässigkeit unserer Prüfverfahren signifikant steigern. Diese Entwicklung führt nicht nur zu einer Erhöhung der Produktqualität und -sicherheit, sondern auch zu einer spürbaren Effizienzsteigerung in den Prüfprozessen.

Darüber hinaus ermöglicht der Einsatz von KI in der ZfP eine umfassendere Überwachung und Steuerung des gesamten Lebenszyklus eines Produkts. Von der ersten Entwurfsphase bis zur finalen Auslieferung können KI-Systeme wertvolle Einblicke und Prognosen liefern, die zu optimierten Prozessen und einer kontinuierlichen Verbesserung der Produktqualität beitragen.

Es ist jedoch entscheidend, dass wir die Weichen richtig stellen, um das volle Potenzial dieser Technologie auszuschöpfen. Dies erfordert eine tiefgehende Auseinandersetzung mit den technischen, ethischen und rechtlichen Fragestellungen, die mit der Nutzung von KI einhergehen. Nur so können wir sicherstellen, dass wir diese mächtige Technologie verantwortungsbewusst und wirkungsvoll einsetzen.

Dieses Merkblatt, verfasst von der Arbeitsgruppe für KI der Deutschen Gesellschaft für Zerstörungsfreie Prüfung (DGZfP), soll Ihnen einen umfassenden Einblick in die Anwendung von KI in der ZfP geben. Unser Ziel ist es, Ihnen die grundlegenden Prinzipien dieser Technologie näherzubringen und die Potenziale für Ihre tägliche Arbeit verständlich zu machen. Ebenso soll es Sie dazu inspirieren, sich aktiv mit den vielfältigen Möglichkeiten der KI in der ZfP auseinanderzusetzen. Es soll Ihnen als Leitfaden dienen, um erste Schritte in diesem spannenden und dynamischen Feld zu unternehmen und die Vorteile der KI für Ihre Arbeit zu erkennen und zu nutzen.

Wir laden Sie herzlich ein, dieses Dokument aufmerksam zu lesen und die Möglichkeiten, die die KI bietet, mit Offenheit und Interesse zu betrachten. Lassen Sie uns gemeinsam die Zukunft der zerstörungsfreien Prüfung gestalten und von den vielfältigen Vorteilen profitieren, die diese Technologie für uns bereithält.

Dieses Vorwort wurde durch ChatGPT erstellt und von uns, den Mitgliedern der Arbeitsgruppe KI, gegengelesen und korrigiert. Wir wünschen Ihnen viel Spaß beim Lesen des Merkblattes (das im Weiteren gänzlich ohne ChatGPT erstellt wurde).

1 Grundlagen Künstliche Intelligenz und Maschinelles Lernen

Dieses Merkblatt bietet einen Überblick, wie die Entwicklung eines KI-Systems ablaufen kann, zeigt die Grenzen und Möglichkeiten der Verfahren auf und legt möglichst anschaulich dar, worauf für die erfolgreiche Abwicklung eines KI-Projekts zu achten ist. Bevor es aber konkret um die Entwicklung von KI-Systemen und das Training von KI-Modellen geht, werden zunächst einige grundlegende Begriffe erläutert, um ein einheitliches Verständnis für das folgende Merkblatt aufzubauen. Sämtliche Begriffe können jederzeit im angefügten Glossar nachgeschlagen werden (siehe Glossar).

KÜNSTLICHE INTELLIGENZ, MASCHINELLES LERNEN UND DEEP LEARNING

Die Begriffe „Künstliche Intelligenz“, „Maschinelles Lernen“ und „Deep Learning“ (oft übersetzt als tiefes oder tiefgreifendes Lernen), werden öfters synonym verwendet, eigentlich stellen sie aber jeweils eine Untermenge des vorhergehenden Begriffs dar (siehe Abbildung 1). „Künstliche Intelligenz“[1] – der Oberbegriff – beschreibt Algorithmen zur Informationsgewinnung und Entscheidungsfindung, die versuchen menschliches Verhalten nachzubilden: von einfachen Wenn-Dann-Entscheidungen über mathematische Funktionen und statistische Verfahren bis hin zu komplexen Kaskaden verschiedener Verfahren des maschinellen Lernens. Das „Maschinelle Lernen“ bezeichnet eine konkrete Klasse von Algorithmen, die auf der Basis von aufbereiteten Trainingsdaten und von Experten ausgesuchten Merkmalen versuchen, ein Muster für ungesehene Daten herzuleiten und nicht explizit programmiert werden müssen. Zu beachten ist hier, dass die Definition eines „Merkmals“ in der ZfP und in der KI voneinander abweichen. Während in der ZfP ein Merkmal bereits eine semantische Bedeutung wie „Pore“ hat, ist dies in der KI nicht der Fall. So sind z. B. beliebige Kanten und Ecken als Merkmal anzusehen. „Deep Learning“ wiederum beschreibt ein Untergebiet des maschinellen Lernens, das sich speziell mit tiefen neuronalen Netzen befasst, die mithilfe ihres mehrschichtigen Aufbaus neben den Mustern auch die nötigen Merkmale aus den Daten extrahieren – dafür aber deutlich mehr aufbereitete Trainingsdaten benötigen.

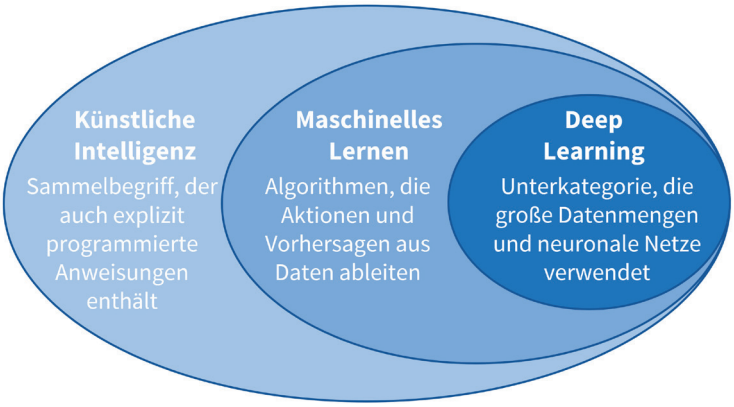

Abb. 1: Oft synonym verwendet, sind die Begriffe "Künstliche Intelligenz", "Maschinelles Lernen" und "Deep Learning" eigentlich Untermengen des jeweils vorhergehenden Begriffs. Künstliche Intelligenz stellt daher einen Sammelbegriff aller Systeme dar, die automatisiert zu Entscheidungen kommen, während beim maschinellen Lernen stets allgemeine Algorithmen auf Basis von Daten zu Entscheidungen kommen. Das Deep Learning fokussiert sich dabei auf eine spezielle Kategorie von Algorithmen, den tiefen neuronalen Netzen.

Im Umfeld des maschinellen Lernens fallen häufig die Begriffe „Data Science“, „Data Engineering“ und „Big Data“. „Data Science“ beschreibt dabei die Tätigkeit der Aufbereitung der Daten für das maschinelle Lernen sowie allgemein die Extraktion von Informationen aus den Daten und setzte eine Mischung aus Domänenwissen, statistischen Methoden und Kenntnissen in der Datenverarbeitung voraus. Das „Data Engineering“ befasst sich mit der Erstellung automatisierter Abläufe zur Aufbereitung neuer Daten, die aus dem produktiven Einsatz zurück in die Entwicklung gespielt werden.

„Big Data“ bezieht sich im Bereich des Deep Learning sowohl auf die großen Datenmengen, die für das Training verwaltet werden müssen, als auch auf die Daten, die während des laufenden Betriebs anfallen und verarbeitet werden müssen.

Den Begriff der künstlichen Intelligenz in seiner heutigen Form gibt es bereits seit den 1950er Jahren. Auch die Anfänge der tiefen neuronalen Netze fallen in diese Zeit. Jedoch waren die Ansätze und Methoden in ihrer praktischen Anwendung lange sehr beschränkt. Den Durchbruch Anfang der 2010er Jahre verdankt das maschinelle Lernen und insbesondere das Deep Learning drei wesentlichen Faktoren: (1) die effiziente Implementierung eines funktionierenden Trainingsalgorithmus; (2) das Aufkommen großer, frei verfügbarer Datensätze; und (3) die Entwicklung leistungsfähiger Hardwarebeschleuniger.

1 Dabei kommt es im Deutschen häufig zu einem falschen Verständnis der Begrifflichkeit durch die Übersetzung aus dem Englischen “Artificial Intelligence”. Das englische “Intelligence” wird im angelsächsischen auch verwendet, um die Möglichkeiten der Datenauswertung zu beschreiben, wie bei der Bezeichnung verschiedener amerikanischer Geheimdienste zu sehen (z. B. CIA – Central Intelligence Agency). Es geht hier nicht um die Beschreibung eines kognitiv, dem Menschen gleichen Systems wie es das deutsche “Intelligenz” näher legt.

Dies führte dazu, dass sich die Entwicklung von KI-Systemen mehr und mehr hin zu datengetriebenen Ansätzen verlagerte und daraufhin immer mehr aufbereitete Trainingsdaten benötigt wurden – genauer gesagt, gelabelte Daten, sprich Daten für die das zu erwartende Ergebnis bekannt ist.

VOR- UND NACHTEILE DATENGETRIEBENER ANSÄTZE

Im Allgemeinen haben die datengetriebenen Ansätze des maschinellen Lernens den Vorteil, dass sie eine höhere Genauigkeit als regelbasierte Systeme ermöglichen. Diese Algorithmen sind besser in der Lage verborgene Muster in komplexen Daten zu erkennen und Fehler und Verzerrungen durch menschliche Einflüsse zu vermeiden – sofern diese nicht bereits fest in den Daten verankert sind (mehr dazu in Abschnitt 3.3). Damit verschiebt sich der Aufwand von der Merkmalskonfiguration hin zur Pflege der Daten, denn neuronale Netze können nur lernen, was in den Trainingsdaten enthalten ist. Dies macht es auch schwerer, Aussagen über den Produktiveinsatz des neuronalen Netzes zu treffen. Das KI-Modell muss daher, ebenfalls datengetrieben, mittels eines Testdatensatzes validiert werden (siehe Abschnitt 2.4).

AUFGABEN FÜR DIE KÜNSTLICHE INTELLIGENZ

Prinzipiell lassen sich mit den aktuellen Methoden der künstlichen Intelligenz viele sehr komplexe Fragestellungen lösen, für die nach heutigem Stand noch menschliche Eingaben und Überprüfungen nötig sind. So lassen sich theoretisch KI-Systeme erstellen, die analoge Anzeigen ablesen, daraus Vorhersagen über das erstellte Bauteil ableiten und automatisch die Prozessparameter anpassen, um den Produktionsprozess zu optimieren; KI-Systeme, die aus einfachen Kameraaufnahmen alter Industrieanlagen automatisiert Bau- und Leitungspläne ableiten; oder KI-Systeme zur prädiktiven Diagnose und Überwachung von Instrumenten auf Verschleißanzeigen. Die Hürde, an denen solche Systeme in der Realität oft scheitern, ist der enorme Aufwand bei der Beschaffung passender Trainingsdaten. Fragestellungen, die sich mit überschaubarem Aufwand lösen lassen, sind vor allem die Anomalieerkennung, die Klassifizierung, die Lokalisierung und die Segmentierung (siehe Abbildung 2).

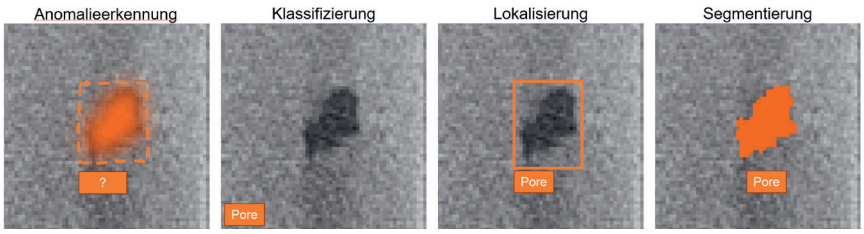

Abb. 2: Unterschiedliche Aufgabenstellungen, die ein KI-Modell in der Bildverarbeitung lösen kann. Bei der Anomalieerkennung werden Abweichungen vom Soll-Zustand gesucht, ohne diese näher zu beschreiben; bei der Klassifizierung wird das Bild als Ganzes kategorisiert; die Lokalisierung findet Bereiche im Bild und kategorisiert diese; und die semantische Segmentierung identifiziert gesuchte Objekte pixelgenau im Bild. Mit steigender Komplexität der Aufgabe, steigt auch der Aufwand für die Erstellung eines geeigneten Trainingsdatensatzes.

Bei der Anomalieerkennung werden in den Eingangsdaten Abweichungen vom Normalzustand erkannt. Die Ursache für solche Abweichungen wird nicht beurteilt. Bei der Klassifizierung werden die Eingangsdaten in eine von mehreren vorgegebenen Kategorien einsortiert. Die Lokalisierung kombiniert die Klassifizierung mit einer Zuordnung einer Position innerhalb der Eingangsdaten. Die Segmentierung ermöglicht eine präzise Unterteilung der Eingangsdaten.

Während Abbildung 2 ein Beispiel aus der Bildverarbeitung zeigt, lassen sich die Aufgaben auch auf beliebigen anderen Eingangsdaten lösen, z. B. auf Audiosignalen einer Klangprüfung, A-Scans einer Ultraschallprüfung in Kombination mit C-Bildern oder Signalen einer Wirbelstromprüfung.

DER UNTERSCHIED ZWISCHEN KI-SYSTEM UND KI-MODELL

Im bisherigen Kapitel wurden bereits die Begriffe KI-System und KI-Modell verwendet, deren Unterschiede an dieser Stelle herausgearbeitet werden. Den Kern des KI-Systems bildet das KI-Modell, z. B. das trainierte neuronale Netz, welches später die gestellte Aufgabe, wie das Segmentieren von Auffälligkeiten in Röntgenbildern, übernimmt. Das KI-System beinhaltet zusätzlich noch die Komponenten zur Datenerzeugung, z. B. Kameras, Röntgendetektoren oder Ultraschallsensorik, sowie die Prüfaufgabe, beispielsweise das Auffinden unzulässiger Defekte ab einer bestimmten Größe in den untersuchten Bauteilen. Die Unterscheidung zwischen KI-System und KI-Modell wird insbesondere bei der Auswertung und Beurteilung des Prüf- und Auswerteverfahrens wichtig. Das KI-Modell operiert grundsätzlich auf einer diskretisierten, digitalen Abbildung, während das

KI-System im physikalischen Raum ausgewertet und auf reale kontinuierliche Informationen angewendet wird (mehr dazu in Abschnitt 2.4).

In den nachfolgenden Abschnitten wird zunächst das typische Vorgehen bei der Entwicklung eines KI-Systems im Allgemeinen und des KI-Modells im Speziellen vorgestellt (siehe Kapitel 2). Anschließend wird betrachtet, worauf bei den Vorbereitungen eines KI-Projekts und beim Einsatz eines KI-Systems zu achten ist (siehe Kapitel 3). Kapitel 2 beleuchtet die Entwicklung dabei aus Sicht technischer Überlegungen und Kapitel 3 aus Sicht des anwendungsbezogenen Rahmenwerks. Durch die Betrachtung aus verschiedenen Blickwinkeln kommt es daher zu thematischen Überlappungen zwischen den Abschnitten.

2 Vorgehen bei einer KI-Entwicklung

In diesem Abschnitt wird das typische Vorgehen bei der Umsetzung einer KI-Entwicklung erläutert und Einblicke in die Arbeit des Entwicklungsteams gegeben. Dies unterscheidet sich geringfügig von der klassischen Softwareentwicklung durch die im Zentrum stehenden Daten. Abbildung 3 zeigt eine Übersicht der einzelnen Schritte der KI-Entwicklung. Der Fokus liegt dabei auf der datengetriebenen Entwicklung eines Deep-Learning-Modells, die Entwicklung herkömmlicher Verfahren des maschinellen Lernens, wie Entscheidungsbäume oder Regressionsmodellen, verläuft jedoch analog [1]. Den Anfang macht die Zieldefinition (siehe Abschnitt 2.1): Welche Fragestellung soll durch das Projekt adressiert werden und was muss für einen erfolgreichen Abschluss erfüllt sein? Dabei ist zu berücksichtigen, dass auch eine künstliche Intelligenz keine 100 % Erkennungsrate erreichen wird. Nach der Festlegung des Zieles geht es daran die nötigen Daten für den Trainingsprozess zu sammeln und aufzubereiten. Wichtig ist, dass die Trainingsdaten genauso aufgenommen werden wie später die Daten aus dem Fertigungsprozess (siehe Abschnitt 2.2). In Abschnitt 2.3 wird die Aufteilung der gesammelten Daten in Trainings- und Validierungsdaten erläutert und der Trainingsprozess beschrieben. Die Validierungsdaten werden erst nach dem Training verwendet, um die Übereinstimmung mit der Zieldefinition zu überprüfen (siehe Abschnitt 2.4). Nach erfolgreicher Validierung kann das trainierte Modell in die Produktivumgebung überführt und dort final ausgewertet werden (siehe Abschnitt 2.5). Dort muss das KI-System automatisiert überwacht werden – sollten sich Änderungen in den Bildmodalitäten oder gar der Zieldefinition ergeben, beginnt mit den in der Produktivumgebung neu gesammelten Daten die nächste Iteration im Projektzyklus. Jede Iteration kann dabei in einem eigenen Projekt abgehandelt werden und zwischen den Projekten kann beliebig viel Zeit vergehen.

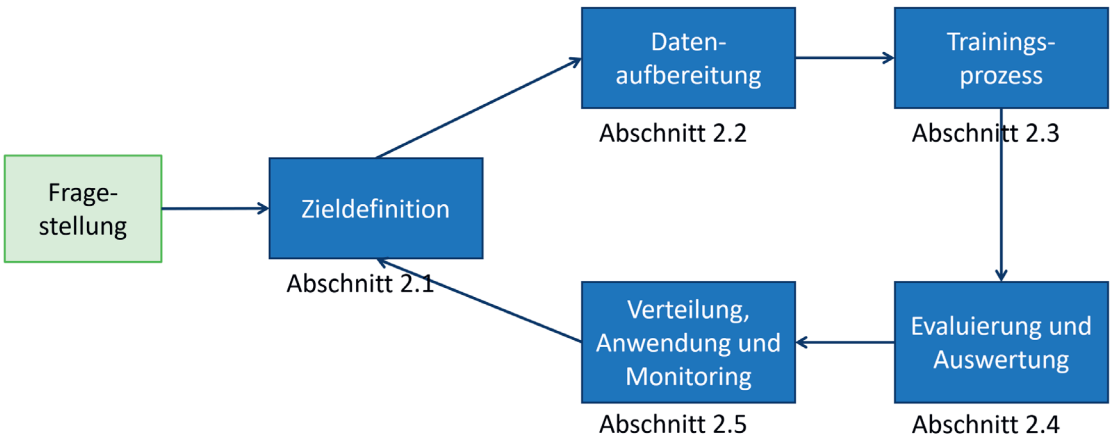

Abb. 3: Der Projektzyklus zur Entwicklung eines KI-Modells: Initiiert von einer Fragestellung werden zunächst ein realistisches Ziel formuliert und die Kriterien zu dessen Erreichung festgelegt. Anschließend werden die Trainingsdaten gesammelt und aufbereitet, bevor es daran geht das konkrete KI-Modell zur Lösung der definierten Aufgabenstellung zu trainieren. Das fertige KI-Modell wird dann mit Hilfe eines Testdatensatzes validiert. Erfüllen die Validierungsergebnisse die Kriterien der Zielsetzung kann das KI-Modell produktiv geschaltet und ins KI-System eingebettet werden. Treten während des Betriebs Änderungen am KI-System auf, beginnt der Zyklus von neuem.

Die Entwicklung sollte dabei stets unter ständiger Einbeziehung des Anwenders ablaufen, damit dieser Rückmeldung zum aktuellen Stand des Projekts geben kann und für unvorhergesehene Detailfragen zur Verfügung steht. Die Umsetzung eines KI-Projektes profitiert stark von einer agilen Entwicklung, die sich durch kleinschrittige Iterationen und einen starken Einbezug aller Interessensgruppen auszeichnet.

2.1 Die Zieldefinition

Die kritische Erfolgsgröße eines KI-Systems liegt in der Erfüllung der definierten Ziele und Anforderungen der Anwendung, für die das System entwickelt wurde. Ein solches Ziel kann beispielsweise im Erkennen von Merkmalen auf Bilddaten liegen, in der Prognose zukünftiger Zustände eines Systems (Ausfallszenario einer Maschine) oder schlicht in der Beschleunigung von Prüfverfahren. Bereits vor dem Start der Entwicklung eines KI-Systems sind daher klare Zielsetzungen in Form objektiver Metriken zu definieren, die bei einer erfolgreichen Umsetzung erreicht werden müssen. Zur besseren Definition des Zieles und um den Erfolg der Entwicklung nachweisen zu können, sollte ein bereits vorhandenes, meist manuell, vom Menschen durchgeführtes Prüf- und Auswerteverfahren, nach den genau gleichen Metriken bewertet werden. In diesem Abschnitt wird auf die grundlegenden Metriken zur Bewertung eines KI-Modells sowie des gesamten KI-Systems eingegangen. Wichtig ist zu verstehen, dass KI-Systeme nicht per se einem spezifischen Anwendungsfall angepasst sind, sondern im Entstehungsprozess daraufhin entwickelt und konfiguriert werden müssen.

UNTERSCHIEDE IN DER QUALITATIVEN ZIELDEFINITION EINES KI-SYSTEMS

Am Beispiel aus der Bilderkennung: KI-Technologie ist in der Lage Bildinterpretationen durchzuführen und kann damit in der Kontrolle von Bauteiloberflächen, Röntgenbildern und anderen Pixelrepräsentationen eingesetzt werden. Ein solches KI-Modell kann beispielsweise darauf trainiert werden, Poren in Röntgenbildern zu finden, zu lokalisieren und zu markieren. Diese Erkennung findet zunächst ohne Einbezug weiteren Kontextes statt. Damit gemeint ist das Hinzuziehen weiterer Bewertungskriterien wie einer Spezifikation, die gefundene Merkmale als Defekte oder eben keine Defekte klassifiziert. Diese zusätzlichen Informationen sind häufig in ZfP-Dateninterpretationen wichtig, müssen aber während der Entwicklung eines KI-Systems aktiv eingeplant werden. So kann eine einfache Zieldefinition in der Schweißnahtprüfung lauten: Finde alle Anzeigen auf dem Röntgenbild, die laut ISO 5817 einer Unregelmäßigkeit entsprechen und ordne das Ergebnis der Zulässigkeitsgrenze 2 nach ISO 10675 zu. Ebenso kann als alternative, pragmatische Zieldefinition für den gleichen Anwendungsfall der Schweißnahtprüfung gewählt werden: Erkenne Bilder mit Unregelmäßigkeiten und sortiere sie aus.

In der Praxis werden solche Unterschiede im Verständnis der Anforderungen unzureichend beschrieben und häufig erst im Laufe der Entwicklung des KI-Systems deutlich. Als Konsequenz entstehen erhöhte Kosten und Aufwände im laufenden Projekt. Notwendig ist die Definition einer solchen Zielsetzung bereits bei Start der Entwicklung, da diese z. B. die Anforderungen an das Labeling der Trainingsdaten bestimmen.

DIE QUANTIFIZIERUNG DER ZIELDEFINITION

In vielen Anwendungsfällen ist zusätzlich zur qualitativen Zieldefinition eine quantitative Bewertungskomponente hinzuzufügen. So kann eine Zieldefinition ebenfalls lauten: Finde 95 % aller Anzeigen auf dem Röntgenbild, die laut ISO 5817 einer Unregelmäßigkeit entsprechen und bewerte die Anzeigen mit der Zulässigkeitsgrenze 2 nach ISO 10675. Diese quantitative Zieldefinition erlaubt eine genaue Evaluierung der Ergebnisse eines KI-Systems auf Basis des konkreten Anwendungsfalles. In der Praxis stellt sich allerdings häufig die Frage nach einem Leistungsvergleich, dem Benchmark: Welche “Genauigkeit” der Erkennung muss erreicht werden und auf welchen Daten soll diese festgestellt werden?

In der ZfP gibt es je nach Branche und Anwendung unterschiedliche Anforderungen an die Genauigkeit einer Dateninterpretation in der Qualitätskontrolle. Durch die Einführung der Detektionswahrscheinlichkeit (engl. „probability of detection“, POD), die vor allem in der Luft- und Raumfahrttechnik Anwendung findet, bietet die ZfP ein Maß für die Qualifizierung von Auswertesystemen. Ein Beispiel hierfür wäre die Detektion relevanter Fehlergrößen, z. B. größer 0,3 mm, mit einer Wahrscheinlichkeit von 90 % und einer Konfidenz von 95 %. Auch die Grenzwertoptimierungskurve (engl. „receiver-operating-characteristic“, ROC) wurde bereits in verschiedenen Forschungsarbeiten auf ZfP-Prüfsysteme erfolgreich angewendet. Die ZfP-Metriken arbeiten anzeigenorientiert und sind üblicherweise abhängig von der Anzeigengröße.

BEWERTUNGSMETRIKEN AUS DEM FACHBEREICH DES MASCHINELLEN LERNENS

Auch die Datenwissenschaften bieten verschiedene Bewertungsmetriken für KI-Modelle an. Diese arbeiten im Allgemeinen pixelbasiert (wie viele Pixel wurden richtig ausgewertet) und sind dadurch unabhängig von der Anzeigengröße, die sonst für die Auswertung der ZfP-Ergebnisse eine zentrale Rolle einnimmt. Die Grundlage für viele dieser Metriken bildet die Konfusionsmatrix (engl. „confusion matrix“, siehe Abbildung 4), die eine detaillierte Aufschlüsselung darüber bietet, welche Klassen das KI-Modell wie oft fehlerhaft zuordnet. In ihrer einfachsten Form, der binären Klassifikation, unterscheidet die Konfusionsmatrix richtig vorhergesagte positive Ergebnisse (engl. „true positives“), richtig vorhergesagte negative Ergebnisse (engl. „true negatives“), falsch vorhergesagte positive Ergebnisse, die tatsächlich negativ wären (Pseudos, Typ-I-Fehler, engl. „false positives“), sowie falsch vorhergesagte negative Ergebnisse, die tatsächlich positiv wären (Schlupf, Typ-II-Fehler, engl. „false negatives“). Aus diesen vier Werten lassen sich unter anderem die Bewertungsmetriken Genauigkeit (engl. „precision“), Trefferquote (engl. „recall“), Richtigkeit (engl. „accuracy“) und der Jaccard-Index oder Intersection-over-union (IoU) ableiten [2]. Einige

dieser Maße lassen sich wiederum weiter zusammenfassen. Das F-Maß (engl. „F-score“) bspw. kombiniert Genauigkeit und Trefferquote in einer einzigen Metrik. Dies ist besonders nützlich, wenn ein ausgeglichenes Verhältnis der kombinierten Metriken benötigt wird. Wichtig ist zu beachten, je abstrakter die Metrik desto schwieriger die Interpretation.

Wie Abbildung 4 zeigt, beschreibt die Genauigkeit, wie viele der als positiv vorhergesagten Ergebnisse tatsächlich positiv, also relevant, sind. Die Trefferquote beschreibt, wie viele aller positiven Elemente tatsächlich als positiv erkannt werden, also die Fähigkeit alle relevanten Ergebnisse zu finden. Die Richtigkeit liefert eine Aussage darüber, wie viele Vorhersagen des KI-Systems (positiv wie negativ) tatsächlich richtig waren. Das Maß berücksichtigt alle korrekten Vorhersagen – positive wie negative. Die IoU dagegen betrachtet lediglich wie viele relevante Ereignisse vorhergesagt und wie viel Fehler dabei gemacht wurden. Die IoU unterscheidet sich von der Richtigkeit dadurch, dass sie die richtig negativ vorhergesagten Ergebnisse ignoriert und sich dadurch für die Evaluierung von Aufgaben eignet, bei denen es eine starke Asymmetrie oder Unausgewogenheit zwischen den Klassen gibt. Dies tritt vor allem in der Objekterkennung und Segmentierung von Anzeigen auf, wenn es darum geht wenige kleine Anzeigen in großen sonst guten Daten zu erkennen.

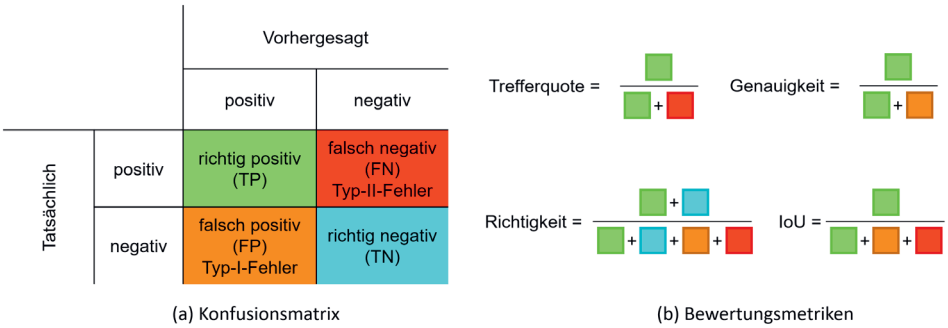

Abb. 4: Die Tabelle auf der linken Seite zeigt eine Konfusionsmatrix. Die Formeln auf der rechten Seite zeigen, wie die Werte aus der Konfusionsmatrix zu aussagekräftigen Werten kombiniert werden. So bildet sich die Trefferquote beispielsweise aus dem Quotienten aus richtig als positiv erkannten Merkmalen und allen tatsächlich positiven Merkmalen zusammen. Deutlich ist auch der Unterschied zwischen Richtigkeit und IoU: Die IoU fokussiert sich auf die relevante Klasse und eignet sich daher auch für unausgeglichene Datensätze, bei denen es deutlich mehr negative als positive Merkmale gibt.

In Abhängigkeit der grundlegenden Prädiktionslogik erfolgt die Berechnung der Metriken leicht unterschiedlich. In der Klassifizierung wird die Zahl der klassifizierten Ereignisse zur Berechnung herangezogen, in der Segmentierung von Bildern die Zahl der klassifizierten Bildelemente (z. B. Pixel) und in der Objekterkennung die Fläche der minimal umgebenden Rechtecke (engl. „bounding boxes“, vgl. Abbildung 5).

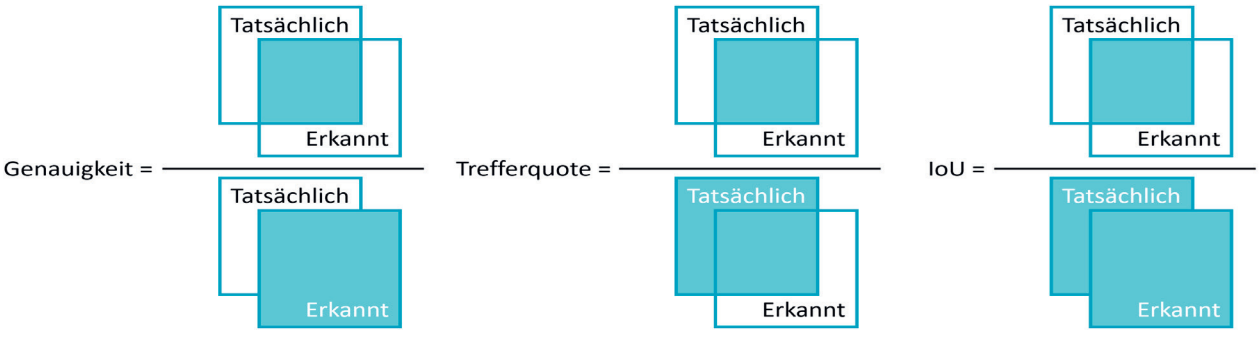

Abb. 5: Eine alternative Betrachtung der Bewertungsmetriken ist die Berechnung aus den Inhalten der minimal umgebenden Rechtecke: Der richtig positive Anteil der Vorhersage ist beispielsweise die Schnittmenge des tatsächlichen und des vorhergesagten minimal umgebenden Rechtecks. Die Kombination aus richtig positiv und falsch positiv ist die Menge aller vorhergesagten Elemente und entspricht daher dem vorhergesagten minimal umgebenden Rechteck. Der Quotient ergibt dann die Genauigkeit.

ABWÄGUNGEN BEI DER ZIELSETZUNG UND NOTWENDIGKEIT VON REFERENZDATEN

In der industriellen Praxis, vor allem der ZfP, gilt es in erster Linie Schlupf zu vermeiden. Dies steht häufig dem Wunsch gegenüber, falsch positive Ergebnisse ebenso gering zu halten, um unnötigen Ausschuss in der Produktion zu vermeiden. Genau an dieser Stelle ist die klare Darstellung der Zieldefinitionen unterstützend, um Ergebnisse vergleichbar und transparent kommunizierbar zu halten. Im Zuge eines Ergebnisvergleichs ist die Bewertung anhand von Referenzdaten, einer sogenannten Ground-Truth oder einem Benchmark von höchster Bedeutung. Ein KI-System kann seine Güte im besten Falle gegenüber einem Vergleichswert, z. B. dem Menschen, nachweisen und wird dadurch vergleichbar. Neben der Verständigung auf Metriken zur Ermittlung der quantitativen Zieldefinition für das KI-System, gilt es diese Metriken für die Ermittlung der Ground-Truth anzulegen. Ein Beispiel aus der Schweißnahtprüfung: In der ROC-Studie [3] konnte nachgewiesen werden, dass der erfahrenste menschliche Experte dieser Stichprobe im besten Fall ca. 88,0 % der Anzeigen in den Röntgenaufnahmen erkennt. Dies tat er bei einer Falsch-Alarm-Rate von 12,9 %. Diese Information kann beispielsweise für ein Benchmark für die Entwicklung eines KI-Systems für den gleichen Anwendungsfall verwendet werden. Die Ermittlung von Benchmarks und Ground-Truths stellt in der Praxis einen umfassenden Diskussionspunkt dar und sollte elementarer Bestandteil der Zieldefinition sein.

Besonders hervorgehoben wird an dieser Stelle die Notwendigkeit von Referenzdaten, auf denen die Benchmarks nach den mit dem Anwender definierten Bewertungsmetriken durchgeführt werden. Nachdem sich auf einen Referenzdatensatz geeinigt wurde, muss dieser stets getrennt von den Trainingsdaten gehalten werden, sodass das KI-Modell immer Daten auswertet wird, die es noch nicht „gesehen“ hat.

STANDARDS ALS TEIL DER ZIELDEFINITION

Die Validierung anhand von Referenzdaten ist in der Praxis häufig nicht ausreichend, um KI-Systeme für den industriellen Serienbetrieb zu qualifizieren. Unternehmen und Branchen schreiben hier meist andere Rahmenbedingungen einer Qualitätserhebung für KI-Systeme fest. So müssen KI-Systeme über einen gewissen Zeitraum (z. B. mehrere Arbeitsschichten) auf konstantem Niveau funktionieren. Dies muss anhand einer gewissen Zahl von Stichproben (z. B. kontrollierte Teile in der Qualitätssicherung) nachweislich durchgeführt werden. Diese Erhebungen müssen dann auf Basis entsprechender Metriken ausgewertet und mit den angelegten Benchmarks verglichen werden. Auch das Zusammenbringen dieser Metriken, Benchmarks und Erhebungs- bzw. Qualifizierungsmethoden für KI-Systeme sollte während der Zieldefinition ausführlich diskutiert werden, um das Bewusstsein für solche Vorgaben als Erfolgskriterium bei den Entwicklungs-Teams zu schärfen.

Während es in der ZfP zahlreiche Standards gibt, fehlen diese im Bereich des maschinellen Lernens noch oft. Bereits laufende Forschungsprojekte befassen sich mit der Fragestellung von Standards für die Qualifizierung von KI-Systemen, z. B. in der industriellen Qualitätskontrolle (siehe z. B. AIQualify, https://www.aiqualify.de/). Diese werden helfen, die Kommunikation zu vereinfachen und die Realisierung von industriellen KI-Systemen deutlich zu beschleunigen.

2.2 Die Datenaufbereitung

Der große Unterschied von Deep Learning zu herkömmlichem maschinellen Lernen ist die Bewegung weg von der manuellen Merkmalskonfiguration hin zur automatisierten Merkmalsextraktion, dem sog. „end-to-end“-Training. Während bei der Merkmalskonfiguration ein Mensch überlegt, welche Merkmale, wie Kanten, Ecken oder Formen, zu guten Entscheidungen des KI-Modells führen könnten, werden bei der automatischen Merkmalsextraktion genau diese Merkmale im Training. Das hat den großen Vorteil, dass das System so optimiert werden kann, dass die Daten optimal verarbeitet werden. Der Nachteil: Es werden deutlich mehr gelabelte Referenzdaten benötigt, um zu einem robusten Ergebnis zu gelangen. Wie viele Datenpunkte genau benötigt werden, kann dabei nicht pauschal gesagt werden und ist stets von der konkreten Fragestellung abhängig. In diesem Abschnitt wird daher darauf eingegangen, worauf es bei der Erstellung eines Datensatzes zum Training ankommt und welches die häufigsten Fehlerquellen sind.

Ein Datensatz ist eine Sammlung an Datenpunkten, der die gesamte Spanne an möglichen Datenpunkten darstellt, d.h. den Raum aller möglichen Datenpunkte abdeckt (siehe Abbildung 6). In der Regel ist es so, dass ein trainiertes KI-Modell nur das erkennt, was durch die Trainingsdaten abgebildet wird. Dies schließt sowohl die Anzeigen ein, die das Modell erkennen soll, als auch die Aufnahmemodalitäten, unter denen die Daten erzeugt werden. Im Beispiel der Porositätsanalyse sind dies z. B. die verschiedenen Merkmalstypen einer Schweißnaht und die Kontrasteinstellungen des Röntgenbildes oder möglicherweise auftretende Artefakte. Deshalb ist es wichtig darauf zu achten, dass die Datenpunkte eines Trainings- bzw. Testdatensatzes immer aus der tatsächlichen Spanne der Daten stammen, die später im produktiven Einsatz auftritt und dass diese möglichst vollständig abgebildet wird. Am besten kommen die Daten, die zum Training verwendet werden, nicht aus einer Vorstudie, sondern direkt aus dem Produktivsystem.

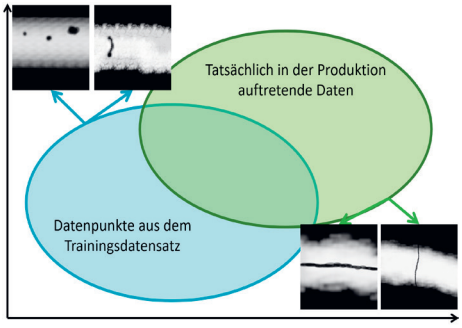

Abb. 6: Bei der Erstellung eines Trainingsdatensatzes ist es wichtig, dass die Trainingsdaten die gesamte Spanne, der an tatsächlich im produktiven Einsatz auftretenden Zustände abbildet. Das heißt, es müssen die gleichen Aufnahme-Modalitäten verwendet werden und die gleichen Unregelmäßigkeiten enthalten sein. Im obigen Beispiel wäre es kaum möglich, dass das trainierte KI-Modell Risse erkennt.

KLASSIFIZIERUNG, LOKALISIERUNG UND SEGMENTIERUNG

Je nach Aufgabenstellung werden unterschiedliche Arten von Label benötigt, die unterschiedlich schwierig zu erstellen sind. Für eine Klassifizierung wird ein Label pro Datenpunkt, sprich z. B. pro Bild-, Ton- oder Textdatei. Bei einer Lokalisierung geht es darum, eine Auffälligkeit oder ein Merkmal innerhalb eines Datenpunktes zu finden. Entsprechend wird für eine Lokalisierung mehr Information benötigt; neben dem „was?“ wird auch ein „wo?“ benötigt. Für Bilder kann dies in Form eines umgebenden Rechtecks oder eines Mittelpunkts geschehen, für Audiodaten beispielsweise ein Zeitstempel. Die genausten Informationen werden für eine Segmentierung benötigt – das Labeling ist entsprechend aufwendig. Es muss jeder Bildpunkt in jedem Datenpunkt gelabelt werden.

SYSTEMATISCHE FEHLER IN DEN DATEN

In den Trainingsdaten spiegeln sich verschiedene Arten systematischer Fehler (engl. „bias“) wider. Allen voran die Stichprobenverzerrung: Es müssen alle Repräsentationen von Unregelmäßigkeiten in den Trainingsdaten vorhanden sein. Was das Modell im Training nicht sieht, kann später nicht erkannt werden (Abdeckungsfehler, engl. „coverage bias“). Nicht nur das Vorkommen aller möglichen Repräsentationen von Unregelmäßigkeiten in den Trainingsdaten ist entscheidend, sondern auch deren Häufigkeit (engl. „sampling bias“). Ist eine Unregelmäßigkeit in den Daten über- oder unterrepräsentiert muss darauf im Trainingsprozess Rücksicht genommen werden.

Eine weitere Quelle systematischer Fehler sind eigene Annahmen, die sich in den Daten wiederfinden. Werden bspw. absichtlich Parameter des Fertigungsprozesses geändert, um Anzeigenbilder zu erzeugen, kann es sein, dass nur bekannte Parameter geändert werden, aber andere ggf. häufigere Anzeigen auch auf andere Arten entstehen können. Eine häufige Form sind Bestätigungsfehler, bei denen die Daten so erzeugt werden, dass sie vorherige Annahmen bestätigen und dabei Informationen ignorieren, die dem widersprechen. Dies geht teilweise so weit, dass ein Modell so erzeugt wird, dass es genau die Ergebnisse erzeugt, die man haben möchte.

Ebenso kritisch ist es, darauf zu achten, dass keine zufälligen Korrelationen in den Daten enthalten sind. Das KI-Modell könnte diese lernen, statt die Eigentliche Aufgabe zu lösen. Werden bspw. alle Bilder an Krebs erkrankter Menschen hauptsächlich in einer Klink aufgenommen, da diese z. B. auf eine bestimmte Krebsart spezialisiert ist und die Bilder gesunder Menschen kommen aus anderen Klinken, kann es passieren, dass das KI-Modell die Unterscheidung zwischen krank und gesund auf Basis der Klinik trifft. Dies kann auch in der ZfP passieren: In zwei Werken werden die Werkstücke mit leicht unterschiedlichen Parametereinstellungen des Prüfsystems geprüft. Aus einem Werk kommen hauptsächlich gute Werkstücke, während es im anderen Werk vermehrt zu Anzeigen kommt. Werden die Datensätze aus beiden Werken verwendet, kann es sein, dass sich das KI-Modell auf die Unterschiede der Prozessparameter fokussiert, anstatt auf das eigentliche Anzeigenbild zu erlernen.

DER NEGATIVE EINFLUSS VON VORVERARBEITUNGSSCHRITTEN

Die Qualität der aufgenommenen Daten ist dabei oft nicht so ausschlaggebend und Daten die für die Interpretation menschlicher Prüfer aufbereitet wurden, müssen sich nicht notwendigerweise auch für die automatisierte Verarbeitung durch ein KI-System eignen. Insbesondere durch Vorverarbeitungsschritte, wie einem Weichzeichnen zur Artefaktreduktion oder dem Komprimieren der Daten zur Archivierung (siehe Abbildung 7), gehen wertvolle Informationen verloren, die das KI-System für seine Vorhersagen nutzen könnte.

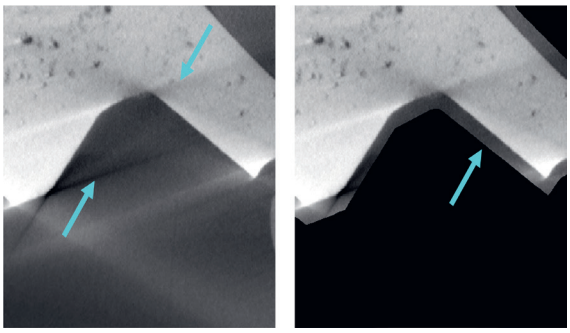

Abb. 7: Ein Beispiel für eine schädliche Vorverarbeitung. Die vermeintlich nicht benötigten Grauwerte in der Luft hart auf null zu setzen, erlaubt eine erhebliche Kompression der Daten. Dabei gehen aber zum einen wertvolle Informationen, z. B. über die Streifenartefakte in der Mitte des Bildes verloren, die an anderer Stelle zu falsch positiven Anzeigen führen können und zum anderen wird eine harte Kante mit einem großen Grauwertsprung in die Daten eingebracht, die die Stabilität des KI-Modells gefährden.

Auf der anderen Seite sind Vorverarbeitungsschritte auch beim Deep Learning nötig: Das Weißen (engl. „whitening“) des Datenhintergrundes sorgt für die nötige numerische Stabilität während des Trainings und muss daher auch zur Vorhersage im produktiven Einsatz mit denselben Parametern angewendet werden.

QUALITÄT UND KONSISTENZ DER ANNOTATIONEN

Wichtiger als die Qualität der Daten ist die Qualität der Referenzdaten. Die Label stellen die sogenannte Ground-Truth dar, gegen die die Ergebnisse des KI-Modells während des Trainings verglichen werden. Diese müssen in erster Linie konsistent sein, in dem Sinne, dass Gleiches immer gleich gelabelt ist, was bei einer manuellen Auswertung, gerade auf Pixelbasis, nie der Fall ist. Zudem sollten sie möglichst fehlerfrei sein. Zieht sich ein Fehler durch den Datensatz, wird das Modell diesen Fehler während des Trainings ebenfalls lernen und später im produktiven Einsatz wiedergeben. Ein KI-Modell kann immer nur so gut sein, wie die Daten, mit denen es trainiert wurde (mehr dazu in Abschnitt 3.3).

DATENERWEITERUNG UND DER EINSATZ SIMULIERTER DATEN

Nicht immer reichen die gelabelten Daten aus, um den Datenbedarf beim Trainieren tiefer neuronaler Netze zu decken. Einige vielversprechende Lösungsansätze sind zum einen die Datensatzerweiterung (engl. „data augmentation“) und zum anderen das Verwenden simulierter Daten und das Übertragen der Lernerfolge (engl. „transfer learning“) auf reale Daten. Bei der Datenerweiterung werden die vorhandenen Datenpunkte durch Transformationen vervielfältigt. In der Bildverarbeitung reichen diese Transformationen von einfachen Operationen wie dem Drehen oder Spiegeln der Datenpunkte bis hin zu beliebigen Verzerrungen in Form und Farbe. Wichtig ist dabei darauf zu achten, dass die transformierten Datenpunkte nicht im Widerspruch zur Aufgabe stehen. Komplexer ist die Verwendung simulierter Daten – entweder im Rahmen eines Vortrainings oder direkt zum Training der tatsächlichen Aufgabe. Wichtig ist hier auf die Realitätsnähe der Simulationen zu achten, vor allem wenn nicht auf Echtdaten nachtrainiert werden kann, z. B. weil bei der Annotation eine Präzision gefordert wird, die manuell nicht erreichbar ist.

2.3 Der Trainingsprozess

Im Werkzeugkasten des maschinellen Lernens finden sich viele Ansätze für unterschiedliche Fragestellungen, aus denen je nach Zielsetzung ein passender ausgewählt wird. In diesem Abschnitt werden einige dieser Werkzeuge im Detail beschrieben.

ÜBERWACHTES, NICHTÜBERWACHTES UND BESTÄRKENDES LERNEN

Die typische Aufgabenstellung im industriellen Umfeld, das Auffinden von Unregelmäßigkeiten, wird durch überwachtes Lernen trainiert. Dabei wird das KI-Modell durch das wiederholte „Zeigen“ von Daten und den entsprechenden Labels so eingestellt, dass es das richtige Label vorhersagt. Die Idee ist, dass sobald das KI-Modell genügend Beispiele mit entsprechendem Label „gesehen“ hat, es auch auf neuen Daten das richtige Label vorhersagt. Im Gegensatz dazu steht das nichtüberwachte Lernen, bei dem das KI-Modell so eingestellt wird, dass es die vorhandenen Daten, zu denen es keine Labels gibt, in Gruppen einteilt. Im besten Fall entsprechen diese Gruppen semantischen Kategorien, was die Gruppen jedoch genau darstellen, ist nicht vorgegeben. So lassen sich beispielsweise KI-Modelle auf Gut-Daten trainieren, die Abweichungen aller Art finden, diese aber nicht näher beschreiben oder kategorisieren können (Anomalieerkennung). Eine weitere Art des Trainings ist das bestärkende Lernen, das darauf beruht dem KI-Modell durch positive Verstärkung oder negative Bestrafung dazu zu bringen, gewünschte Verhaltensweisen zu wiederholen. Dieser Ansatz findet vor allem in der Robotik Verwendung, wenn es bspw. darum geht, dass ein Roboterarm etwas greifen soll. Ein Faktor für die Belohnung des KI-Modells könnte dabei die

benötigte Zeit sein. Ein weiteres prominentes Beispiel ist AlphaGo von DeepMind, das mittels bestärkendem Lernen und Millionen von Go-Partien trainiert wurde [4].

DER UNTERSCHIED ZWISCHEN HYPERPARAMETERN UND FREIEN PARAMETERN

Eine Gemeinsamkeit aller Deep-Learning-Modelle ist der Aufbau aus zwei verschiedenen Arten von Parametern: Den sog. Hyperparametern, welche die genaue Ausprägung der gewählten Methode beschreiben und vor dem Trainingsprozess festgelegt werden und den sog. Freien Parametern, die während des Trainings anhand der bereitgestellten Trainingsdaten durch den Trainingsansatz algorithmisch bestimmt werden. Je mehr freie Parameter eine Methode bietet, desto mehr Trainingsdaten werden benötigt. Insbesondere im Bereich der neuronalen Netze ist die Zahl der freien Parameter sehr groß und liegt teilweise im mehrstelligen Millionenbereich. Die Hyperparameter wären in diesem Fall beispielsweise die Zahl der verwendeten Schichten im neuronalen Netz oder die Zahl der Neuronen pro Schicht. In diesem Abschnitt soll das Vorgehen zur Bestimmung dieser freien Parameter aus den gelabelten Daten grob beschrieben, die wichtigsten Fachbegriffe erläutert und mögliche Fallstricke aufgezeigt werden.

DAS TRAINING – EIN OPTIMIERUNGSPROBLEM

Wie bereits angedeutet werden die freien Parameter – im Gegensatz zu herkömmlichen, regelbasierten Ansätzen – nicht von Hand bestimmt, sondern als Optimierungsproblem auf Basis der bereitgestellten Trainingsdaten und einer mathematischen Zielfunktion (sog. Kosten- oder Loss-Funktion) formuliert. Dies ist mitunter der Grund, warum die trainierten Modelle nicht besser sein können als die Daten, mit denen sie trainiert werden und warum eine Abweichung von den Trainingsdaten während des Betriebs einen negativen Einfluss auf das Ergebnis hat. Um die bestmöglichen Werte für die freien Parameter zu finden, wird z. B. für neuronale Netze ein Gradientenabstiegsverfahren verwendet. Dabei wird das neuronale Netz mit zufällig gewählten Werten initialisiert. Anschließend werden die Ergebnisse der Trainingsdaten berechnet und mithilfe der Zielfunktion mit den Annotationen verglichen. Aus den Fehlern wird eine Aktualisierung der freien Parameter errechnet und so das neuronale Netz iterativ verbessert. Aufgrund der großen Menge an Daten können nicht alle Trainingsdaten auf einmal zum Berechnen der Wertaktualisierung herangezogen werden, sodass die Aktualisierung der freien Parameter immer auf Basis weniger Beispieldaten erfolgt, der sog. (Mini-)Batch (siehe Abbildung 8). Damit der Trainingsprozess bzw. das Berechnen der Wertaktualisierungen möglichst optimal verläuft, ist es wichtig auf konsistente Annotationen zu achten [5].

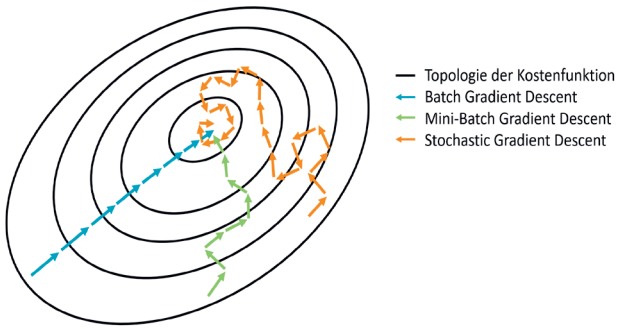

Abb. 8: Vereinfachte Darstellung des Gradientenabstiegsverfahren. Wird für jede Aktualisierung der Parameter der volle Trainingsdatensatz verwendet (blauer Pfad), führen die Aktualisierungen geradlinig zum nächsten Minimum. Bei der Größe der verwendeten Datensätze ist dies jedoch nicht praktikabel. Wird für jede Aktualisierung der Parameter nur ein Beispiel aus dem Trainingsdatensatz verwendet (orangener Pfad), werden wesentlich mehr Schritte benötigt, die sich aber deutlich schneller berechnen lassen. Das Mini-Batch-Verfahren, bei dem mehrere Beispiele auf einmal verwendet werden (grüner Pfad), bildet den Mittelweg. Ein optimales Ergebnis erfordert konsistent gelabelte Daten.

Sind alle Daten des Trainingsdatensatzes einmal zur Aktualisierung der freien Parameter betrachtet worden, ist eine sogenannte Epoche im Trainingsprozess abgeschlossen. Je nach Größe des Datensatzes werden die neuronalen Netze über mehrere Epochen trainiert, zwischen denen die Reihenfolge, in der die Beispieldaten gezeigt werden, zufällig neu gewählt wird. Dies ist dem (Mini-)Batch-Verfahren geschuldet. Je nach Modellgröße und Umfang des Datensatzes kann ein Training einige Stunden bis mehrere Tage in Anspruch nehmen. Um den Trainingsprozess abzukürzen, erlauben einige Methoden – insbesondere neuronale Netze – eine Initialisierung mit Parametern, die bereits für eine ähnliche Aufgabe trainiert wurden, das sogenannte Transfer Learning.

VON ÜBERANPASSUNG UND KREUZVALIDIERUNG

Es werden immer gelabelte Daten zur Validierung benötig, um bestimmen zu können, wie gut die Vorhersagen des trainierten KI-Modells sind. Die Validierungsdaten dürfen nicht in den Trainingsdatensatz aufgenommen werden. Sowohl der Trainings- als auch der Validierungsdatensatz müssen alle Anzeigenbilder abdecken – im Validierungsdatensatz sind einige wenige Beispiele pro Anzeige jedoch ausreichend. Sind nicht genügend gelabelte Daten vorhanden, um sowohl Trainings- als auch Validierungsdatensatz zu erstellen, werden mehr gelabelte Daten benötigt, damit eine KI-Entwicklung erfolgreich durchgeführt werden kann. Durch die große Zahl freier Parameter ist es möglich, dass diese auf den Trainingsdatensatz überangepasst werden (sog. Overfitting) – das Modell merkt sich die Trainingsdaten auswendig. Um eine Überanpassung zu erkennen, wird der gelabelte Datensatz zunächst in einen Trainings- und einen Validierungsdatensatz unterteilt. Der Validierungsdatensatz wird dabei zufällig aus dem gesamten Datensatz gezogen, wobei zu beachten ist, dass alle Klassen (oder Repräsentationen von Unregelmäßigkeiten) im Validierungsdatensatz vorkommen. Diese Daten werden nicht zum Training verwendet. Stattdessen werden diese Daten während des Trainings verarbeitet, um den Trainingsfortschritt aufzuzeigen und das Training zu überwachen. Verhält sich der Fehler auf dem Validierungsdatensatz ähnlich dem Fehler auf dem Trainingsdatensatz, ist alles in Ordnung; ist der Fehler auf dem Validierungsdatensatz jedoch signifikant größer, ist das Training in eine Überanpassung gelaufen.

Abb. 9: k-fache Kreuzvalidierung. Der gelabelte Datensatz wird zunächst in einen Trainingsdatensatz und einen Testdatensatz unterteilt. Letzterer wird für die finale Evaluierung des trainierten Modells beiseitegelegt. Für die Kreuzvalidierung wird der Trainingsdatensatz in k gleiche Teile unterteilt und sukzessive ein Teil als Validierungsdatensatz zur Auswertung verwendet und auf den verbleibenden k – 1 Teilen das Modell trainiert.

Für eine bestmögliche Evaluierung des Trainingsprozesses kann der Datensatz in beispielsweise vier gleiche Teile geteilt werden (siehe Abbildung 9). Im ersten Durchgang wird der letzte Teil als Validierungsdatensatz herangezogen und das Modell auf den restlichen drei Teilen trainiert. Im zweiten Durchgang dient der vorletzte Teil als Validierungsdatensatz usw., bis alle Teile einmal als Validierungsdatensatz verwendet wurden. Ist der Fehler in jedem Durchgang in etwa derselbe, ist das Training gelungen. Dieser Vorgang heißt Kreuzvalidierung (engl. „cross-validation“).

DAS VERZERRUNG-VARIANZ-DILEMMA ODER WARUM NIE 0 % FEHLERRATEN ERREICHT WERDEN

Um gegen eine Überanpassung vorzugehen, gibt es mehrere Möglichkeiten zur Regularisierung des Trainingsprozesses: Zum einen kann die Zahl der freien Parameter durch eine Anpassung der Hyperparameter reduziert werden, zum anderen kann der Wertebereich der freien Parameter durch zusätzliche Zielfunktionen eingeschränkt werden. Außerdem kann durch den Einsatz von mehr Trainingsdaten einer Überanpassung entgegengesteuert werden.

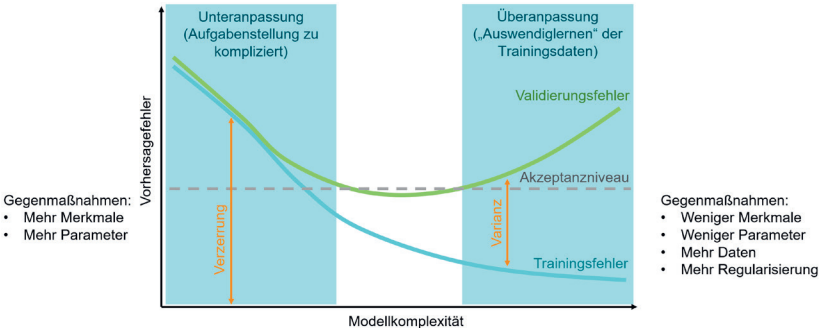

Abb. 10: Verzerrung-Varianz-Dilemma. Während des Trainings wird der Fehler, den das Modell auf den Trainings- und den Validierungsdatensatz macht, beobachtet. Weichen die Fehlerwerte deutlich voneinander ab und beginnt der Fehler auf dem Validierungs-datensatz gar wieder zu steigen, liegt eine Überanpassung des Modells auf den Trainingsdatensatz vor. Sind die Fehler in etwa gleich, aber signifikant größer als das vereinbarte, tolerierbare Fehlermaß, handelt es sich um eine Unteranpassung – das Modell kann die Komplexität der Fragestellung nicht abbilden. Dazwischen gibt es den optimalen Bereich, in dem das Modell die gewünschten Ergebnisse liefert.

Auf der anderen Seite kann es vorkommen, dass der Fehler auf Trainings- und Validierungsdatensatz in etwa gleich ist, aber deutlich über dem angestrebten tolerierten Fehlermaß liegt. Das Modell weist dann eine zu große Verzerrung auf (engl. „bias"). In diesem Fall kann es genügen, das Modell länger zu trainieren oder durch eine Erhöhung der freien Parameter die Kapazität des Modells zu erhöhen. Abbildung 10 zeigt die Abhängigkeit von Modellkomplexität, Trainingsfehler und Validierungsfehler.

2.4 Evaluierung und Auswertung

Am Ende eines Trainingsprozesses steht die Auswertung des trainierten Modells nach den zuvor vereinbarten Metriken, um die Erreichung der Zielsetzung sicherzustellen (siehe Abschnitt 2.1). Diese Auswertung findet auf dem zuvor vereinbarten Testdatensatz statt. Dieser Test kann nur abdecken, was in den Testdaten enthalten ist, daher ist es wichtig, dass – wie in den Trainingsdaten – alle zu erwartenden Repräsentationen von Unregelmäßigkeiten in den Testdaten enthalten sind.

ERWARTUNGEN AN DAS KI-MODELL

Oft wird KI-Modellen aufgrund ihres „Black-Box"-Charakters weniger Vertrauen zugesprochen als dem manuellen Prüfer. Dabei unterliegen menschliche Auswertungen oft zahlreichen, subjektiven Einflüssen, angefangen bei der Tagesform des Prüfers [6]. Ein KI-System unterliegt nur solchen Umwelteinflüssen, die z. B. einen Einfluss auf das bildgebende Verfahren haben – sofern die Datengrundlage stimmt (siehe Bias).

Entsprechen die erreichten Ergebnisse nicht den Erwartungen kann dies mehrere Ursachen haben. Die offensichtlichste ist, dass das gewählte KI-Modell schlicht nicht für die Komplexität der Aufgabenstellung geeignet ist oder nicht gut genug trainiert wurde. Manchmal ist der Fehler jedoch auch in den Labels des Testdatensatzes zu suchen, da diese den gleichen Schwankungen unterliegen wie die manuelle Prüfung im Allgemeinen. In diesen Fällen sind die Abweichungen der Vorhersagen des KI-Systems und der Labels des Testdatensatzes genauer zu prüfen und nach Rücksprache mit dem Auftraggeber zu klären.

UNTERSCHIEDE ZWISCHEN DER WELT DER KI UND DER WELT DER ZFP

Die Auswertung des KI-Modells findet aus der Sicht des KI-Entwicklers statt. Für den Einsatz in der zerstörungsfreien Prüfung ist eine zusätzliche Bewertung aus der Sicht der ZfP wichtig, das heißt unter Einbezug physikalischer Größen. Dadurch wird jedoch nicht mehr nur das KI-Modell betrachtet, sondern das gesamte KI-System bewertet, das schließt z. B. bildgebende Verfahren mit ein.

Entsprechend unterscheidet sich der Fokus der Auswertung und Evaluierung des KI-Modells aus Sicht des KI-Experten und der des KI-Systems aus Sicht des ZfP-Experten. Für die Evaluierung aus der ZfP-Sicht ist die Abdeckung der Grenzfälle wichtig, werden z. B. alle Poren, die größer als 0,5 mm sind mit einer gegebenen Wahrscheinlichkeit gefunden? Für die Evaluierung aus der KI-Sicht ist es wichtig alle Vorhersagen des KI-Modells zu prüfen wozu bspw. ein pixelbasiertes Maß wie die IoU herangezogen wird. Für den Erfolg eines KI-Projekts ist es wichtig in der Auswertung Metriken zu verwenden, die beide Seiten verstehen.

Je nach Anwendungsfall kann die Gewichtung falsch positiver und falsch negativer Ergebnisse variieren. Bei der Zwischenprüfung von Bauteilen vor der Nachverarbeitung sind falsch positive bspw. schlimmer als falsch negative Ergebnisse, da sie den Ausschuss unnötig erhöhen. Auf der anderen Seite sind – insbesondere bei sicherheitsrelevanten Bauteilen – falsch negative Ergebnisse bei der finalen Prüfung kritischer zu bewerten als falsch positive Ergebnisse. Die Abwägung zwischen der Bewertung falsch negativer und falsch positiver Anzeigen muss für jeden Anwendungsfall neu entschieden werden.

Wichtig ist, bereits bei der Zielsetzung (siehe Abschnitt 2.1) alle Erwartungen klar zu formulieren und auf eventuell nötige Zertifizierungsschritte einzugehen. Dies umfasst auch die Sicherstellung der Auswertung des KI-Systems als Ganzes und die Verknüpfung von Metriken aus der KI-Welt mit den Metriken aus der ZfP-Welt (siehe auch Abschnitt 3.5).

2.5 Verteilung, Anwendung und Monitoring

Nach dem Training des KI-Modells und der erfolgreichen Auswertung des gesamten KI-Systems steht die Bereitstellung und Verteilung beim Kunden. Je nach Anforderungen an das KI-System kann dies auf verschiedene Arten geschehen. Eine wichtige Rolle dabei spielen die Datenmenge, die bei der Prüfung anfällt, die Wichtigkeit des Datenschutzes sowie die Notwendigkeit für Aktualisierungen des KI-Systems (siehe Abbildung 11).

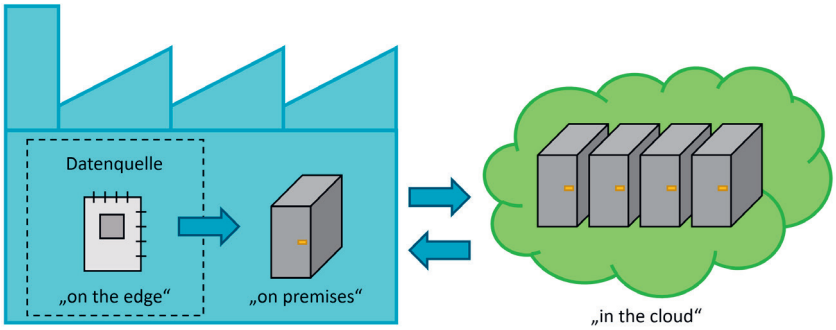

Abb. 11: Je nach Anforderungen an das KI-System, ist es sinnvoll das KI-System unterschiedlich „nahe“ an der Anwendung zu installieren. „On the edge“, z. B. im bildgebenden System, liefert es die schnellsten Ergebnisse, da die Daten nicht kopiert werden müssen. Eine Cloud-Lösung dagegen bietet den Vorteil, dass das zugrundeliegende KI-Modell schnell aktualisiert und ausgetauscht werden kann und keine zusätzliche Hardware vor Ort nötig ist – dafür müssen die Daten über das Internet zum Service-Anbieter kopiert werden.

„ON THE EDGE”, „ON-PREMISES” ODER IN DER CLOUD

Bei einer Cloud-basierten Lösung werden sämtliche Daten über das Internet an einen Server übertragen, dort analysiert und die Ergebnisse zurückgeschickt. Der große Vorteil dieser Variante liegt in der einfachen Skalierbarkeit der Hardware und den damit verbundenen Kosten – es muss nur die Rechenzeit bezahlt werden, die tatsächlich benötigt wird und die Hardware und Software wird vom Serverbetreiber aktuell gehalten, weshalb keine teure Hardware-Anschaffungen inhouse nötig sind. Üblicherweise bieten solche Lösungen auch einen hohen Standard in Bezug auf Cyber-Security, den typischerweise kleine und mittelständige Unternehmen nur schwer selbst erreichen können. Die Übertragung der Daten über das Internet auf Drittanbietersysteme erschwert jedoch die Nachvollziehbarkeit des Datenflusses und wo einzelne Daten überall gespeichert sind. Außerdem spielt bei großen Datenmengen auch die Datenübertragungsrate eine wichtige Rolle. Der Anbieter des KI-Systems kann bei einer solchen Lösung das KI-Modell (nach Rücksprache) einfach austauschen.

Alternativ dazu kann das KI-System „on-premises“ auf einer Maschine vor Ort installiert werden. Dies hat für den Anwender den Vorteil, dass seine Daten das Unternehmensnetz nicht verlassen müssen. Dafür muss er sich um die Bereitstellung und Wartung der entsprechenden Hardware, Software und der notwendigen Cyber-Security kümmern, was nicht nur Material- sondern auch Personalkosten nach sich zieht. Auch eventuelle Aktualisierungen des enthaltenen KI-Modells muss der Kunde dann in der Regel selbst vornehmen oder durch Drittanbieter vornehmen lassen.

Weiterhin ist es möglich, dass ein KI-Modell als Teil der benötigten Hardware ausgeliefert wird. Das KI-Modell befindet sich dann auf nicht mehr veränderbaren eingebetteten System der Hardware des KI-Systems. Hier liegt der Vorteil in der deutlich erhöhten Ausführungsgeschwindigkeit. Die Daten werden dort verarbeitet, wo sie entstehen und lediglich die extrahierten Ergebnisse werden weitergeleitet. Eine Aktualisierung des KI-Modells geht dann mit der Aktualisierung der Hardware des KI-Systems oder einem Softwareupdate einher.

ANWENDUNG UND WARTUNG

Die Aktualisierung des KI-Modells kann insbesondere dann notwendig werden, wenn sich die Anforderungen an das KI-System ändern. Dies kann zum Beispiel durch das Auftreten neuer Anzeigenbilder, durch Änderungen am Verfahren zur Datenerzeugung oder durch das Altern von dessen Komponenten (siehe Abbildung 12) begründet sein. Dies hat zur Folge, dass ein KI-System kontinuierlich weiterentwickelt werden kann und im Sinne der Optimierung auch sollte. Ein Monitoring ist hier der entscheidende Faktor. Dies kann z. B. dadurch geschehen, dass regelmäßig Bauteile mit bekannten Anzeigen geprüft und die Ergebnisse über-

wacht werden. Wichtig ist, dass jede Anpassung am KI-System einer erneuten Freigabe Bedarf. Dazu muss der alte oder ein angepasster Testdatensatz erneut durchlaufen werden.

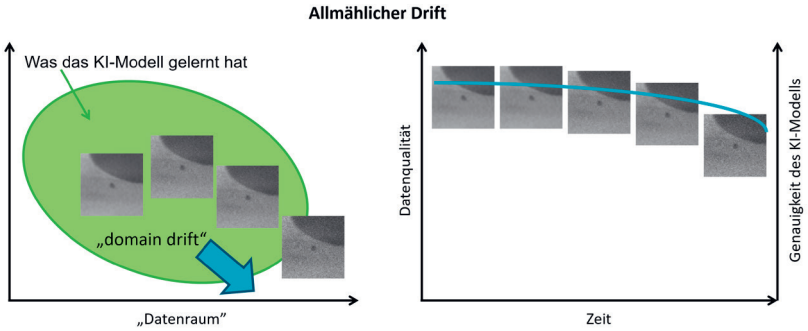

Abb. 12: Allmähliche Änderungen der Sensorik, z. B. das Altern bestimmter Komponenten im bildgebenden Verfahren, haben einen Einfluss auf die Genauigkeit des KI-Modells. Diese Änderungen passieren schleichend und sind zunächst kaum spürbar. Verlässt die Datenqualität jedoch die „Komfortzone“ des KI-Modells werden die Ergebnisse unzuverlässig. Eine kontinuierliche Überwachung des KI-Systems ist daher unerlässlich.

Jede Aktualisierung des KI-Modells stellt in der Regel einen neuen Durchgang durch den Projekt-Lebenszyklus dar und beginnt wieder mit der Zielsetzung. Wie in Abschnitt 2.1 erwähnt, wird ein KI-System immer auf den spezifischen Anwendungsfall hin entwickelt. Das bedeutet auch, dass ein KI-System, das in einer Produktionsstraße gute Ergebnisse liefert, nicht zwangsweise in anderen Produktionsstraßen die gleichen Ergebnisse liefern wird (siehe Abschnitt 3.5).

Bei vielen KI-Systemen, insbesondere in sicherheitskritischen Bereichen, ist es aufgrund von normativen oder sogar gesetzlichen Regelungen notwendig, dass eine menschliche Instanz die finale Entscheidung trifft (siehe dazu auch Abschnitt 3.2). Hier wäre es naheliegend, dass das KI-Modell kontinuierlich mit den Entscheidungen des Menschen im laufenden Betrieb mitlernt. Dieses aktive Lernen (engl. „active learning“) muss jedoch zum einen von Anfang an in der Entwicklung des KI-Systems berücksichtigt werden, da tiefe neuronale Netze per se nicht dafür ausgelegt sind, zum anderen birgt dieses Vorgehen erhebliche Risiken: Je nach Tagesform der prüfenden Person können sich so systematische Fehler in das KI-Modell einschleichen.

3 Voraussetzung für eine KI-Entwicklung

Bevor die Frage nach Voraussetzungen relevant wird, sollte man sich zunächst überlegen, ob eine aufwendige KI-Entwicklung überhaupt notwendig ist. Oft lassen sich selbst schwierige Fragstellungen mit klassischen Analysemethoden hinreichend genau, schnell und kosteneffizient lösen. Eine Fragestellung zunächst mit klassischen Ansätzen z. B. aus der Bildverarbeitung zu lösen, bietet zudem den Vorteil, eine Referenz zu haben, gegen die sich andere Ansätze vergleichen lassen. Auf Methoden des maschinellen Lernens zurückzugreifen ist vor allem dann lohnend, wenn klassische Ansätze unzureichende Ergebnisse liefern. Dies ist zum Beispiel der Fall, wenn die Taktzeit sehr kurz ist und die Datenqualität der aufgenommenen Daten darunter leidet. Wichtig ist auch zu berücksichtigen, dass nicht jede Zielsetzung einer KI-Entwicklung mittels Deep Learning gelöst werden muss: der Werkzeugkasten des maschinellen Lernens bietet eine große Zahl verschiedenster Ansätze, die je nach Fragestellung schneller und ressourcenschonender sein können.

Für jedes Projekt ist eine klare Zielvorgabe ausschlaggebend. Im Falle einer KI-Entwicklung sind die Zielvorgaben mit der Erfüllung gewisser Metriken verbunden. Diese hängen stark von den Daten ab, auf denen sie berechnet werden. Daher ist es wichtig für eine KI-Entwicklung den Validierungsdatensatz, auf dem die Metriken berechnet werden, mit in die Zielvorgabe aufzunehmen. Zudem müssen die zu erreichenden Werte klar festgehalten werden.

Eine weitere wichtige Voraussetzung ist, sich über den Anwendungsfall im Klaren zu sein. Wo wird das KI-System eingesetzt werden: auf einem eingebetteten System ohne Verbindungen zur Außenwelt (und damit ohne externe Wartungsmöglichkeit), auf einer lokalen Maschine mit beschränktem Zugang zum Internet oder serverseitig als Teil einer Cloud-Lösung auf einem leistungsstarken Server?

Neben diesen allgemeinen Voraussetzungen, die sich aus dem Projektzyklus der Projekte ergeben, gibt es noch einige Voraussetzung, auf die in diesem Abschnitt näher eingegangen werden soll. Wichtig ist zunächst zu wissen, was ein KI-System leisten kann und was nicht (Abschnitt 3.1) und was es darf. Daher geht Abschnitt 3.2 auf die Befugnisse ein, die dem KI-System eingeräumt werden können und einen Einfluss auf dessen Zulassung haben. Danach geht es um Abwägungen im Bereich der Daten: In Abschnitt 3.3 geht es um die Qualität der Labels, die insbesondere bei kleinen Datensätzen eine entscheidende Rolle spielt und in Abschnitt 3.4 geht es um Dateneigentum und die möglichen Vorteile, wenn eigene Daten für umfassendere Entwicklungen freigegeben werden. Abschließend beschäftigt sich Abschnitt 3.5 mit den Möglichkeiten zur Qualifizierung von KI-Systemen.

3.1 Erwartungen und Skepsis

Mit dem Begriff „Künstliche Intelligenz“ sind viele Erwartungen aber auch große Skepsis verbunden. Wichtig ist, dass sich hinter dem hochtrabenden Begriff künstliche „Intelligenz“ ein breites Spektrum an digitalen, datengetriebenen Werkzeugen verbirgt und keine mit der menschlichen Intelligenz vergleichbare kognitive Fähigkeit – auch wenn dies in den Medien oft anders suggeriert wird. Nichtsdestotrotz gibt es viele Fragen zu klären: Wie kann ich dem Ergebnis vertrauen? Wer übernimmt die Verantwortung für die Ergebnisse? Wieso macht die Maschine trotzdem noch Fehler? Und, und, und… auf die im Folgenden kurz eingegangen wird.

SIND REGELBASIERTE ALGORITHMEN WIRKLICH NACHVOLLZIEHBAR?

Der aus den (Trainings-) Daten abgeleitete Weg zur Entscheidungsfindung des KI-Modells ist weniger nachvollziehbar als es scheinbar bei herkömmlichen Methoden der Fall ist, bei denen der Mensch überlegt, wie aus den Rohdaten durch geschickte Verarbeitung das gewünschte Resultat erreicht wird. In vielen Fällen sind diese herkömmlichen Algorithmen für den Endanwender genauso undurchsichtig, wie ein KI-Modell – zumal die wichtigen Details der Algorithmen oft als Geschäftsgeheimnis geschützt und verborgen sind. So ist beispielsweise im Fall der Porendetektion alles, was über ein einfaches Schwellwertverfahren hinausgeht, nur schwer zu greifen. Daher ist es wichtig, KI-Modelle durch geeignete (Validierungs-) Daten zu qualifizieren. Welche Eigenschaften ein geeigneter Datensatz aufweisen sollte wird in Abschnitt 3.3 beschrieben und auf Möglichkeiten der Qualifizierung wird in Abschnitt 3.5 eingegangen.

AUCH EINE KÜNSTLICHE INTELLIGENZ MACHT FEHLER

Auf Grund der Tatsache, dass ein KI-Modell mit einer Menge an Daten trainiert wird und nicht darauf ausgelegt ist, aus diesen Daten zu extrapolieren, ist es nicht verwunderlich, dass auch eine künstliche „Intelligenz“ Fehler macht – vor allem wenn die Datengrundlage fehlt. Eine wichtige Frage, die sich hier aufdrängt, ist jedoch: Wie oft macht der Mensch Fehler, unbewusst oder bewusst? Ein KI-System bietet hier die Möglichkeit einen Prozess konstanter Ergebnisqualität aufzusetzen. Um die Wahrscheinlichkeit von Fehlern zu reduzieren, ist es ratsam, das KI-Modell auf die Extraktion der nötigen Informationen aus den Rohdaten zu reduzieren und den Gut-Schlecht-Entscheid auf Basis dieser Informationen regelbasiert nachzulagern (siehe Abschnitt 3.3).

DIE FRAGE NACH DER VERANTWORTUNG

Wenn nun auch ein KI-Modell nicht unfehlbar ist, wer trägt dann die Verantwortung für das (voll-) automatisierte System? Hier ändert sich nichts zu klassischen Methoden. Nicht der Prüfer, sondern eine übergeordnete Instanz, wie ein Vorgesetzter oder ein (noch nicht vorhandener) Standard sind hier in die Pflicht zu nehmen. Zusätzlich muss für den Algorithmus wie das KI-System nachgewiesen werden, dass es richtig arbeitet und es darf danach nicht mehr geändert werden. Die Kriterien der Qualifizierung sind unter anderem nach dem Automatisierungsgrad und der Kritizität der Anwendung auszurichten.

3.2 Assistenzsystem vs. vollautomatisierter Entscheider

Der Einsatz eines KI-Systems kann in verschiedenen Graden der Automatisierung erfolgen. Bei der Entscheidung, wie das KI-System eingesetzt werden soll, kann ein Vergleich mit dem menschlichen Prüfer hilfreich sein. KI-Systeme arbeiten unabhängig von Tagesform und Laune. Allerdings machen sie auch keine Ausnahmen. Wichtig beim Einsatz eines KI-Systems ist auch die Tatsache, dass ein KI-System keine Verantwortung übernehmen kann – diese verbleibt beim Menschen.

AUTOMATISIERUNGSGRAD DER KI

Je nach Betrachtungsweise werden KI-Systeme in der Literatur in drei Stufen – zum Teil mit Unterkategorien – eingeteilt. In der ersten Stufe dient die KI lediglich als Hilfestellung für den menschlichen Entscheider und hebt beispielsweise relevante Merkmale in den betrachteten Daten hervor oder bietet Entscheidungsmöglichkeiten an. Das Risko bei diesem Einsatz eines KI-Systems ist gering, da der menschliche Entscheider die finale Entscheidung trifft. Das KI-System kann allerdings die Entscheidungen des menschlichen Entscheiders beeinflussen, es besteht die Gefahr dem KI-System zu stark zu vertrauen, wodurch sich eine gewisse Nachlässigkeit einschleichen kann. In der zweiten Stufe läuft das KI-System noch immer als Assistenzsystem, trifft jedoch eine Vorauswahl dessen, was der menschliche Entscheider bei einer Prüfung berücksichtigen sollte. Das Risiko ist hier klar, dass alles, was das KI-System herausfiltert, nicht mehr in den Entscheidungsprozess mit einbezogen wird. Das KI-System sollte daher sehr gründlich bewertet werden und ggf. besser zu viele falsch positive Anzeigen liefern. In der dritten und finalen Stufe trifft das KI-System selbstständig Entscheidungen. Der menschliche Entscheider dient hier lediglich als letzte Kontrollinstanz, die den Prozess überwacht und Entscheidungen gegebenenfalls nachträglich korrigieren kann. Das Risiko ist, dass das KI-System eine falsche Entscheidung trifft und diese zu spät oder gar nicht vom menschlichen Entscheider gefunden wird. Näheres lässt sich zum Beispiel in der Richtlinie der EASA für Stufe 1 KI-Systeme [7], den Empfehlungen der ENIQ [8], oder dem Leitfaden des Fraunhofer Instituts [9] nachlesen.

3.3 Daten, Datenverteilung und Qualität der Annotationen

Erfahrungsgemäß ist eine der ersten Fragen, die im Rahmen einer KI-Entwicklung fällt, „wie viele gelabelte Trainingsdaten benötigen wir?“ Leider ist diese Frage nicht allgemeingültig zu beantworten. Selbst wenn wir uns auf Deep Learning beschränken und die Komplexität des KI-Modells als gegeben annehmen, hängt die Antwort von vielen weiteren Faktoren ab, z. B. (i) der Aufgabenstellung, auf welche Zielstellung trainiert werden soll, (ii) dem Umfang der Aufgabenstellung (eine einfache Detektion oder das Vorhersagen abgeleiteter Maße), (iii) der Qualität der Labels und Daten oder (iv) der erforderlichen Zielgenauigkeit.

DIE WICHTIGKEIT DER DATEN

Zunächst einmal ist es wichtig, dass die Trainingsdaten die gesamte Breite der Aufgabenstellung repräsentieren. KI-Modelle sind in der Lage selbst komplexe Zusammenhänge aus einer genügend großen Menge an Trainingsdaten zu extrahieren, haben jedoch erhebliche Schwierigkeiten, wenn das „bekannte Terrain“ verlassen wird. Um ein abstraktes Beispiel zu zeigen, wird in Abbildung 13 ein KI-Modell darauf trainiert, ein periodisches Signal nachzubilden. Für den weiß hinterlegten Bereich liegen Trainingsdaten vor. Hier hat selbst ein einfaches KI-Modell keine Schwierigkeiten, das Signal korrekt zu erfassen. Anders sieht es in den grau hinterlegten Bereichen aus, für die keine Trainingsdaten vorliegen: Hier weichen die Vorhersagen sehr schnell stark vom eigentlichen Signal ab. Ein KI-Modell lernt nur das, was es in den Trainingsdaten sieht.

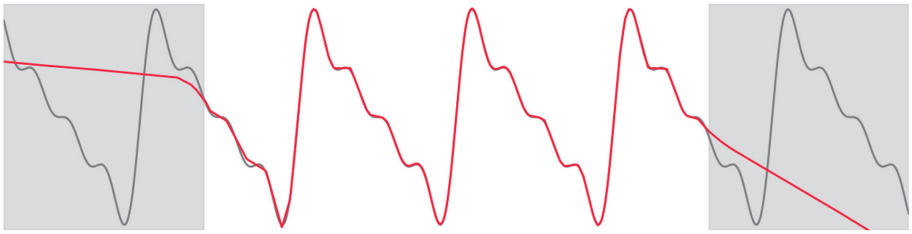

Abb. 13: Für den Bereich, für den Trainingsdaten vorliegen (weiß hinterlegt), liefert selbst ein einfaches neuronales Netz zuverlässige Ergebnisse (rote Linie). Sobald diese „Komfortzone“ verlassen wird und zur Trainingszeit keine Daten mehr vorlagen (grau hinterlegter Bereich), werden die Ergebnisse rapide schlechter, zumal das neuronale Netz keine periodischen Signale kennt. Die tatsächlichen Daten sind als dunkelgraue Linie dargestellt.

Übertragen auf die Erkennung von Anzeigen in CT-Aufnahmen (siehe auch Abbildung 6) bedeutet dies beispielsweise, dass nur die Anzeigen zuverlässig erkannt werden können, die auch in den Trainingsdaten vorhanden waren. Das bedeutet auch, dass mindestens so viele Daten benötigt werden, sodass alle Repräsentationen von Unregelmäßigkeiten sowohl in den Trainings- als auch in den Testdaten vorhanden sind.

Die Daten müssen dabei notwendigerweise mit den Systemen aufgenommen werden, die später Teil des KI-Systems sind. In der Qualitätssicherung mittels CT gibt es beispielsweise Unterschiede zwischen Stichprobenkontrollen und der 100 %-Serienprüfung. Für die Prüfung vereinzelter Stichproben aus der Produktion ist mehr Zeit verfügbar, die CT-Scans weisen daher nur wenig Artefakte auf. Im Gegensatz dazu stehen die CT-Scans der Inline-Prüfung: Der Zeitmangel schlägt sich in hohem Rauschen und starker Artefakt-Belastung nieder. Werden im Rahmen einer Machbarkeitsstudie CT-Scans aus der Stichprobenkontrolle genommen, kann das damit trainierte KI-Modell nicht in einem KI-System in der Produktionslinienprüfung eingesetzt werden.

QUALITÄT DER LABEL

Ähnlich wichtig, wie die Menge und der Umfang der Daten, ist die Qualität der Label. Sind die vorhandenen Label ungenau oder gar widersprüchlich, steigt der Bedarf an Daten stark an. Eine hohe Labelqualität zu erreichen ist aufwendig und selbst große, frei verfügbare Datensätze weisen diese nicht auf. Im ursprünglichen, oft verwendeten ImageNet [10] Datensatz befinden sich etwa 5 % fehlerhaft gelabelte Daten [11]. Diese sind lange Zeit nicht aufgefallen, da sie von der großen Menge an Daten kompensiert wurden. Das heißt, möchte man z. B. ein KI-Modell an die wenigen stark verrauschten Datenpunkte (entspricht ungenauen Labeln) in Abbildung 14a anpassen, hat man es sehr schwer. Jede der gezeigten Lösungen könnte richtig sein. Werden deutlich mehr Datenpunkte hinzugefügt, wird es leichter auf die zugrundeliegende Verteilung zu schließen (siehe Abbildung 14b). Ist die zur Verfügung stehende Menge an Daten jedoch limitiert, ist es umso wichtiger, dass diese korrekt gelabelt sind (siehe Abbildung 14c). Im Bereich des angewandten maschinellen Lernens gibt es daher auch eine Bewegung weg von „big data“ hin zu „good data“ [12]. Je nach Aufgabenstellung können sich die Fehler in den Labeln sehr subtil im trainierten Modell äußern oder bereits beim Training zu schwerwiegenden Problemen führen.

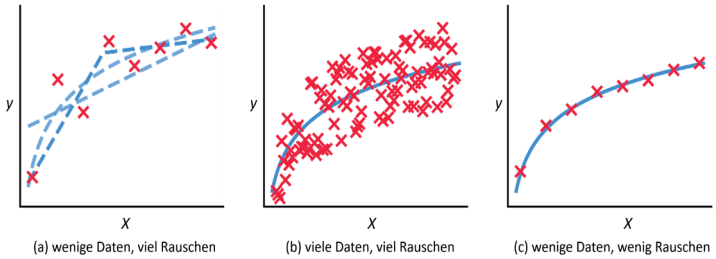

Abb. 14: Die Möglichkeiten des KI-Modells hängen immer von den verfügbaren Trainingsdaten und der Qualität der Label ab. Sind nur wenige Datenpunkte verfügbar und die Label sehr verrauscht, hat es das KI-Modell schwer, die richtige Zielfunktion zu treffen (a) Dies lässt sich durch das Hinzufügen vieler weiterer Trainingspunkte bedingt beheben (b) Besser ist es jedoch auf die Qualität der Label zu achten, um die bestmöglichen Ergebnisse zu erzielen.

Weniger subtil äußern sich die Fehler beispielsweise bei Segmentieraufgaben, wenn ähnliche Datenpunkte manchmal als gewünschte Anzeige und manchmal als Hintergrund markiert werden: Dies führt dazu, dass Ergebnisse des Modells ungenau werden. Sie verlieren ihre klaren Kanten und damit ihre Aussagekraft. Ein wichtiger Punkt beim Labeling von Daten aus Sicht der zerstörungsfreien Prüfung ist daher die Behandlung von nicht relevanten Anzeigen, die nicht zu einem Ausschuss führen würden. Ist es das Ziel ein KI-Modell zu trainieren, das bspw. Anzeigen in Durchstrahlungsbildern von Objekten findet, ist es wichtig, dass in den Trainingsdaten alle möglichen Anzeigen gelabelt wurden, ungeachtet dessen, ob sie relevant sind, oder nicht. Werden diese nicht gelabelt und damit implizit dem nicht relevanten Hintergrund zugeordnet, hat dies einen entscheidenden negativen Einfluss auf das Gesamtergebnis des KI-Systems. Widersprüchliche Labels führen zu ungenauen Vorhersagen.

WIE VIELE DATEN SIND NUN NÖTIG?

Die Frage nach der Menge der benötigten Trainingsdaten ist also immer eine Einzelfallentscheidung, ebenso wie die Abwägung der Datenqualität, die das jeweilige bildgebende Verfahren liefert. Letztere muss immer zur Aufgabenstellung passen. Die Erkennbarkeit von Anzeigen einer geforderten Größe muss gegeben sein. Es kann beispielsweise vorkommen, dass ein Qualitätsmaß bei einer gegebenen Datenqualität gar nicht erfüllbar ist. Es wird daher empfohlen, stets eine Machbarkeitsstudie im Vorfeld einer KI-Entwicklung durchzuführen.

MEHR QUALITÄT DURCH AUFGABENBESCHRÄNKUNG

Um für mehr Konsistenz in den Labels zu sorgen, kann es sinnvoll sein, das Modell nur nach Auffälligkeiten suchen zu lassen und eine IO/NIO-Entscheidung nachgelagert zu treffen, v.a. wenn es für diese Entscheidung bereits klare Regeln und strikte Normen gibt. Ein Beispiel aus der Bildverarbeitung: Im Aluminiumdruckguss ist es nahezu unmöglich, ein Bauteil ohne Unregelmäßigkeiten zu erstellen. Es ist deshalb wichtig sicherzustellen, dass diese nur in unkritischen Regionen eines Bauteils auftreten. Um konsistente Label zu gewährleisten, empfiehlt es sich, alle Unregelmäßigkeiten zu markieren, statt nur die zu labeln, die tatsächliche Defekte darstellen und die Entscheidung über tatsächlich Defekte anschließend über einen regelbasierten Ansatz zu bestimmen.

3.4 Datenschutz und Dateneigentum

Jedes KI-Modell lernt anhand von Trainingsdaten, bei bildgebenden Modellen beispielsweise mit Trainingsbildern. Da der Umfang und die Qualität dieser Trainingsdaten von entscheidender Bedeutung für die Leistungsfähigkeit der KI sind, ist die Frage nach dem Schutz dieser Trainingsdaten von besonderer Relevanz.

Die Europäische Kommission hat dabei erkannt, dass der eigentliche Wert von Daten in ihrer Nutzung und Weiterverwendung liegt, dass aber gleichzeitig für die innovative Weiterverwendung von Daten zur Entwicklung künstlicher Intelligenz nicht genügend Daten zur Verfügung stehen. Insbesondere hat sich nach Einschätzung der Kommission die gemeinsame Nutzung von Daten zwischen Unternehmen noch nicht ausreichend durchgesetzt, was vor allem an den fehlenden wirtschaftlichen Anreizen und letztlich auch der Furcht, etwaige Wettbewerbsvorteile einzubüßen, liegt (vgl. Europäische Kommission, Eine europäische Datenstrategie, COM (2020), 66 final, 19. Februar 2020, S. 7 f.).

DOCH WIE KANN DER SCHUTZ VON TRAININGSDATEN RECHTLICH BEGRÜNDET WERDEN?

Anders als sonst im bürgerlichen Recht üblich, lässt sich an Daten jedenfalls kein Schutz des Eigentums konstruieren. Das Bürgerliche Gesetzbuch (BGB) sieht Eigentum lediglich an "Sachen" vor (§ 903 BGB), d.h. nur an körperlichen Gegenständen – ein Charakteristikum, der Datenpunkt offenkundig fehlt.

Wegen des technischen Charakters könnte sich daher ein patentrechtlicher Schutz der Trainingsdaten aufdrängen. Patentierbar sind nach deutschem Recht allerdings nur neue Erfindungen auf dem Gebiet der Technik, wobei als Erfindung in diesem Sinne die Lösung einer Aufgabe durch technische Mittel verstanden wird. Da aber Trainingsdaten (ebenso wie auch Computerprogramme) selbst keine Lösung einer technischen Aufgabe darstellen, kann an ihnen kein Patentschutz entstehen.

Ob ein einzelnes Trainingsdatum dem urheberrechtlichen Schutz unterfällt, hängt entscheidend davon ab, ob das einzelne Trainingsdatum als "Werk" im Sinne von § 2 UrhG zu qualifizieren ist. Hierunter sind nur persönliche geistige Schöpfungen zu verstehen, also solche, die von einem Menschen geschaffen wurden und ein Mindestmaß an Individualität und Schöpfungshöhe aufweisen.

Jedwede von Maschinen autonom und ohne schöpferischen Beitrag eines Menschen geschaffenen Arbeitsergebnisse fallen damit von vornherein bereits aus dem Schutzbereich des Urheberrechts heraus.

URHEBERSCHUTZ FÜR DEN DATENSATZ

Aber auch ansonsten ist das Erreichen des für einen Urheberschutz erforderlichen Mindestmaßes an kreativer Leistung, der sog. Schöpfungshöhe, in Bezug auf das einzelne Trainingsdatum unwahrscheinlich. Möglich ist dagegen der urheberrechtliche Schutz der Datenbank selbst (§ 87a UrhG), also der Sammlung von Daten, die systematisch oder methodisch angeordnet und einzeln mit Hilfe elektronischer Mittel (oder auf andere Weise) zugänglich sind, und deren Beschaffung, Überprüfung oder Darstellung eine nach Art oder Umfang wesentliche Investition erfordert.

Hierüber ist der Datenbankhersteller als Investierender zumindest vor unberechtigter Vervielfältigung, Verbreitung oder öffentlicher Wiedergabe der Datenbank selbst (oder eines nach Art und Umfang wesentlichen Teils der Datenbank) durch einen Dritten geschützt. Allerdings ist nach der höchstrichterlichen Rechtsprechung ein rechtswidriger Eingriff in das Recht des Datenbankherstellers nur gegeben, wenn die Entnahmehandlungen hierauf gerichtet sind und im Fall ihrer Fortsetzung dazu führen würden, dass die Datenbank insgesamt oder ein nach Art oder Umfang wesentlicher Teil davon vervielfältigt, verbreitet oder öffentlich wiedergegeben wird. Ob dies bei einer Datenübernahme zum Training einer KI der Fall ist, hängt vom individuellen Einzelfall ab.

Exkurs: Soweit auch ein einzelnes Trainingsdatum ein Werk darstellt und damit dem urheberrechtlichen Schutz unterfällt und der Trainierende nicht zugleich auch Urheber (oder zumindest: Inhaber der entsprechenden Nutzungsrechte) hieran ist, stellt sich die Frage, ob und in welchem Umfang das Werk zum Trainieren einer KI ohne Erwerb einer entsprechenden Lizenz genutzt werden darf.

Technisch stellt das Training nämlich eine (zumindest vorübergehende) Vervielfältigung des Werkes dar. Hier eröffnet § 44b Abs. 1 UrhG dem Trainierenden allerdings die Möglichkeit zum Text und Data Mining, also zur automatisierten Analyse von einzelnen oder mehreren digitalen oder digitalisierten Werken, um daraus Informationen insbesondere über Muster, Trends und Korrelationen zu gewinnen.

(Vorübergehende) Vervielfältigungen von rechtmäßig zugänglichen digitalen oder digitalisierten Werken sind danach auch ohne Einwilligung des Urhebers zulässig. Die Vervielfältigungen sind aber zu löschen, wenn sie für das Text und Data Mining nicht mehr erforderlich sind – also jedenfalls nach Abschluss des Trainings.

DATEN ALS GESCHÄFTSGEHEIMNIS

Allerdings gibt das Gesetz dem Rechteinhaber die Möglichkeit einen Vorbehalt gegen das Text und Data Mining bezüglich seiner Werke zu erklären (was bei online verfügbaren Inhalten in maschinenlesbarer Form zu erfolgen hat). Macht der Rechteinhaber hiervon Gebrauch, ist die Nutzung der betreffenden Werke zu KI-Trainingszwecken nicht mehr zulässig.

Da Trainingsdaten aber letztlich auch "nur" Informationen sind, kann sich zumindest (bzw. zusätzlich) ein Schutz nach dem Gesetz zum Schutz von Geschäftsgeheimnissen (GeschGehG) ergeben. Als Geschäftsgeheimnis geschützt sind dabei Informationen, die

(i) weder insgesamt noch in der genauen Anordnung und Zusammensetzung ihrer Bestandteile den Personen in den Kreisen, die üblicherweise mit dieser Art von Informationen umgehen, allgemein bekannt oder ohne Weiteres zugänglich sind und daher von wirtschaftlichem Wert sind,

(ii) Gegenstand von den Umständen nach angemessen Geheimhaltungsmaßnahmen durch ihren rechtmäßigen Inhaber sind und

(iii) bei denen auch ein berechtigtes Interesse an der Geheimhaltung besteht.

Sämtliche Merkmale können auf Trainingsdaten grundsätzlich zutreffen. Gleichwohl hat dies zur Konsequenz, dass einer Lizenzierung von Trainingsdaten, d.h. deren Überlassung an einen Dritten, immer auch das Risiko des Verlusts des Geschäftsgeheimnisses innewohnt.

Angesichts der oben beschriebenen Schutzlücken ist es bei der Überlassung von Trainingsdaten daher unerlässlich, den Umfang und die Grenzen der zulässigen Nutzung solcher Daten durch einen Lizenzenznehmer, sowie die zu deren Schutz zu implementierenden Sicherungsmaßnahmen, vertraglich rechtssicher zu regeln und Sanktionen für den Fall eines Verstoßes (z. B. Vertragsstrafe) vorzusehen.

KI-VERORDNUNG DER EUROPÄISCHEN UNION

Losgelöst vom Schutz der Trainingsdaten gibt es auf europäischer Ebene weitere gesetzgeberische Vorhaben mit KI-Bezug. So soll mit der Verordnung zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz ("KI-Verordnung") der Einsatz von KI-Systemen europaweit einheitlich reguliert und damit Rechtssicherheit geschaffen werden.

Am 14. Juni 2023 verabschiedete das Europäische Parlament seine diesbezügliche Verhandlungsposition, es werden sich nun noch die Verhandlungen mit der Europäischen Kommission und dem Rat der Europäischen Union anschließen.

Inhaltlich soll die Verordnung gewährleisten, dass die auf dem Unionsmarkt in Verkehr gebrachten und verwendeten KI-Systeme sicher sind und dass bei deren Nutzung die Grundrechte und Grundfreiheiten der Bürgerinnen und Bürger gewahrt werden.

Dazu sollen verschiedene Anwendungsbereiche und Risikoklassen unterschieden werden. Die Anbieter von Hochrisiko-Systemen (gemeint sich solche Systeme, die potenziell signifikante Auswirkungen auf die Rechte und Freiheiten von Personen haben können, wie etwa medizinische Diagnosesysteme oder Systeme im Bereich der öffentlichen Sicherheit) müssen – nach dem Willen des europäischen Gesetzgebers – zunächst eine umfassende Risikobewertung im Hinblick auf die Sicherheit, Gesundheit, Privatsphäre und Grundrechte der betroffenen Personen durchführen.

Aber selbst Anbieter von Basismodellen (d.h. solche Systeme, die auf einer breiten Datenbasis trainiert wurden, auf eine allgemeine Ausgabe angelegt sind und an eine breite Palette unterschiedlicher Aufgaben angepasst werden können) müssen u.a. bestimmte Leistungsstandards, Energieeffizienz, die technische Dokumentation und ein angemessenes Risikomanagement sicherstellen.

Die KI-Verordnung wird voraussichtlich erst im Jahr 2026 in Kraft treten. Bereits jetzt können Anbieter oder Hersteller von KI-Systemen aber mit technischen Sicherheitsvorkehrungen und Überwachungsmechanismen beginnen, um die einem KI-System innewohnenden Fehler zu beheben und damit Risken oder Schäden zu verhindern.

TECHNISCHE BETRACHTUNG UND ENTWICKLUNGEN

Es gibt immer wieder Missverständnisse darüber, ob auf einzelne Daten zugegriffen werden kann, wenn diese in ein Modelltraining eingeflossen sind. Dies ist grundsätzlich nicht der Fall. So werden Daten genutzt, um die freien Parameter eines KI-Modells zu optimieren, nicht aber reproduzierbar eingespeichert. Damit werden die Daten einer Partei, die zum Training eines KI-Modells verwendet wurden, nicht für eine zweite Partei zugänglich, die ihrerseits Daten für ein Nachtraining einbringt. Der Datentransfer zwischen zwei z. B. im Wettbewerb befindlichen Parteien ist somit ausgeschlossen. Eine “Vermischung” verschiedener Datensätze aus unterschiedlichen Quellen zum gleichen Anwendungsfall ist förderlich für die allgemeine Leistungsfähigkeit des KI-Modells, vor allem hinsichtlich dessen Robustheit. Dies belegen zahlreiche Studien aus anderen Anwendungsgebieten, allen voran [13].

Die Entwicklung technischer Möglichkeiten, KI -Modelle auf vertraulichen Daten zu trainieren, ist ein aktives Forschungsthema. In der klinischen Bildgebung wird bspw. der Ansatz des "föderalen Lernens" (engl. "federated learning") erprobt, bei dem das KI-Modell nacheinander auf den vertraulichen Datensätzen verschiedener Kliniken trainiert wird, ohne dass diese die Daten an eine zentrale Stelle schicken müssen. Stattdessen reist das KI-Modell von einem Datensatz zum nächsten. Ansätze dieser Art sollen in Zukunft Daten-Marktplätze erlauben, über die die Besitzer von Datensätzen vergütet werden können im Verhältnis zu dem Wert, welches das KI-Modell durch das Training hinzugewonnen hat.

3.5 Qualifizierung von KI-Systemen

Insbesondere in reglementierten und sicherheitskritischen Bereichen wie der Luft- und Raumfahrt oder dem Nuklearenergiesektor ist für viele die Qualifizierung, Zertifizierung und Standardisierung von KI-Systemen ein entscheidendes Thema. Bei klassischen ZfP-Verfahren wird dazu jeder Arbeitsschritt mit allen seinen Parametern eindeutig festgeschrieben. Das mitunter sehr umfassende Standarddokument enthält dann die zu verwendende Software und die Werte der einzustellenden Parameter. Doch wie qualifiziert man ein KI-System mit Millionen oder Milliarden von Parametern, die zufällig initialisiert, aus einer breiten Datenbasis abgeleitet werden?

ÜBERLEGUNGEN FÜR EINEN ÜBERGREIFENDEN STANDARD

Zu welchem Grad muss tatsächlich verstanden werden, wie das KI-Modell zu einem Ergebnis kommt, um einen Standard zu etablieren? Inwieweit reicht es aus die Vorhersagegenauigkeit zu validieren? Anstatt den gesamten Prozess der Modellerstellung zu qualifizieren und zu standardisieren, wäre es denkbar, lediglich ein trainiertes KI-Modell oder eher das gesamte KI-System zu qualifizieren. Dazu wäre ein Datensatz an Referenzbauteilen notwendig, die es zu prüfen gilt. Die sorgfältig ausgewählten Referenzbauteile müssen dann entsprechend eines festgelegten Referenzmaßes, z. B. der POD, ausgewertet werden und ein bestimmtes Mindestmaß erfüllen. Wie Trainings- und Validierungsdatensatz muss auch ein solcher Referenzdatensatz alle zu erwartenden Repräsentationen von Unregelmäßigkeiten und andere Auffälligkeiten enthalten, damit er sich für eine Qualifizierung eignet. Die Referenzbauteile stellen einen physikalischen Anhang zum Standard dar.

Die Referenzbauteile sollten im Anschluss aufbewahrt und in regelmäßigen Abständen erneut geprüft werden, um einen allmählichen Drift (siehe Abschnitt 2.5) rechtzeitig erkennen zu können. Die Vorschrift eines solchen Monitorings sollte daher ebenfalls ein wesentlicher Bestandteil des Standards sein. Dabei muss nicht immer der komplette Datensatz an Referenzbauteilen geprüft werden, zunehmende Schwankungen lassen sich evtl. auch an einzelnen Referenzstücken erkennen.

Im Allgemeinen ist es wichtig, dass sich bei der Entwicklung für ein KI-Modell an geltende Entwicklungsstandards gehalten wird, die in einem Standard gefordert werden können. Darunter fällt z. B. die Archivierung und Versionierung der verwendeten Trainingsdaten und des verwendeten Quellcodes.

SKALIERBARKEIT UND ROBUSTHEIT

Definierte Referenzbauteile sind auch dann von Vorteil, wenn es darum geht, eine KI-basierte Inspektionslösung zu skalieren. Ein KI-System wurde z. B. testweise zur Überwachung einer Produktionslinie trainiert und soll nun auf weiteren Produktionslinien an verschiedenen Standorten ausgerollt werden. Dort herrschen andere Bedingungen, die sich unter anderem auf das bildgebende Verfahren auswirken können. Um sicherzustellen, dass das KI-System dort genauso zuverlässige Ergebnisse liefert, wie bei der initialen Produktionslinie, sollten dort die Referenzbauteile wieder geprüft werden. So können die Grenzen der Skalierbarkeit des KI-Systems bestimmt werden. Je mehr Daten verschiedener Produktionslinien zum Training des KI-Systems verwendet werden desto robuster ist das KI-System gegenüber leichter Veränderungen in den Bilddaten und desto wahrscheinlicher ist dessen gute Skalierbarkeit.

Robustheit ist aber nicht nur im Rahmen der Skalierbarkeit ein wichtiges Thema, es betrifft auch Änderungen auf ein und derselben Produktionslinie. Den beispielsweise durch Verschleiß an den Hardwarekomponenten des KI-Systems entstehenden Artefakten, wie zunehmendes Rauschen, kann durch eine entsprechende Berücksichtigung im Trainingsprozess bis zu einem gewissen Grad vorgebeugt werden. Problematisch ist es, wenn einzelne Bestandteile des KI-Systems ausgetauscht werden. Im Fall einer Durchstrahlungsprüfung kann es notwendig werden, den Detektor auszutauschen. Wird dann – aufgrund aktueller Entwicklungen – ein modernerer Detektor mit mehr Pixeln und einer besseren Kontrastauflösung gewählt, kann dies zu einem rapiden Abfall der Vorhersagequalität des KI-Systems führen, obwohl die entstehenden Bilder für den menschlichen Betrachter deutlich besser aussehen. Grund hierfür ist die Tatsache, dass solche Bilder nicht in den ursprünglichen Trainingsdaten enthalten waren und das KI-Modell diese daher nicht richtig verarbeiten kann. Ein solcher plötzlicher Drift (vgl. Abbildung 15) erfordert eine Anpassung des KI-Modells.

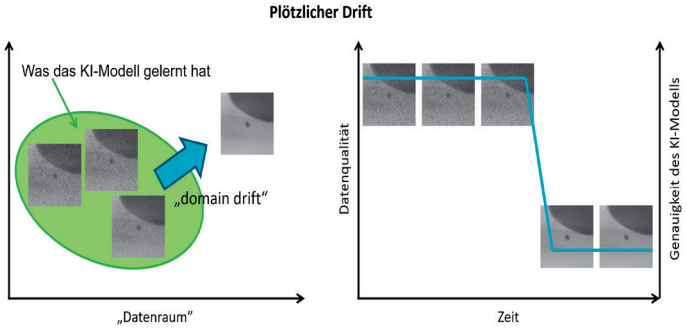

Abb. 15: Änderungen der Sensorik, z. B. durch den Austausch bestimmter Komponenten im bildgebenden Verfahren, haben einen Einfluss auf die Genauigkeit des KI-Modells. Diese Änderungen sind plötzlich und sofort spürbar, die Datenqualität verlässt die „Komfortzone“des KI-Modells und die Ergebnisse werden unzuverlässig. Die Notwendigkeit zum Nachtrainieren kann durch einen Test auf einem zuvor festgelegten Referenzdatensatz geprüft werden.

Eine solche Änderung kann auch eine Anpassung des Bauteils sein. Wird bei einem Gussstück beispielsweise der Anguss neu platziert oder generelle Formanpassungen vorgenommen, führt das zu Änderungen im Bild und hat so einen Einfluss auf die Vorhersagequalität des KI-Systems. Auch hier kann es sein, dass bei kleinen Änderungen das KI-Modell robust genug ist, mit den Änderungen weiterarbeiten zu können. Um sicher zu gehen ist es in diesem Fall notwendig, einen neuen Referenzdatensatz mit der neuen Bauteilform zu definieren.

4 Zusammenfassung und Ausblick

Dieses Merkblatt der Arbeitsgruppe für künstliche Intelligenz (KI) der DGZfP bietet einen Überblick über die Grundlagen der KI und des maschinellen Lernens im Kontext der zerstörungsfreien Prüfung mit dem Schwerpunkt Deep Learning. Im Fokus steht dabei ein strukturiertes Vorgehen bei der KI-Entwicklung, das die Schritte der Zieldefinition, der Datenaufbereitung, des Trainingsprozesses, der Evaluierung und Auswertung sowie der Verteilung, Anwendung und Wartung von KI-Systemen umfasst. Dabei wird insbesondere auf die verschiedenen technischen Aspekte eingegangen, die es bei der KI-Entwicklung zu berücksichtigen gilt.

Ebenso werden nicht-technische Voraussetzungen für eine erfolgreiche KI-Entwicklung detailliert behandelt, einschließlich der Erwartungen und Skepsis in der Zerstörungsfreien Prüfung, der Unterscheidung zwischen Assistenzsystemen und vollautomatisierten KI-basierten Entscheidungen, der Bedeutung von Datenqualität und -verteilung, Datenschutz sowie Dateneigentum und der Qualifizierung von KI-Systemen.

In der Zusammenfassung möchten wir betonen, dass dieses Merkblatt als Grundlagenwerk für den ZfP-Anwender dienen soll und für die Berücksichtigung der genannten Aspekte sensibilisieren möchte, um die Basis für erfolgreiche Entwicklungen zuverlässiger KI-Systeme in der zerstörungsfreien Prüfung zu legen. Ein bewusster Umgang mit Erwartungen, Datenschutz und Qualifizierung stellt die Grundlage für die Akzeptanz und Integration von KI dar.

Leider haben (noch) nicht alle aktuellen Entwicklungen ihren Weg in unser Merkblatt gefunden. Uns ist es wichtig, dass wir zunächst die längerfristig gültigen Grundlagen erläutern, bevor wir auf aktuelle Trends eingehen. Als Ausblick sehen wir vor, in weiteren Merkblättern z. B. über die aktuellen Fortschritte Large Language Models – oder die sog. Foundation Models im Allgemeinen – zu berichten. Diese zeichnen sich dadurch aus, dass sie mit natürlicher Sprache umgehen und so bspw. komplizierte Prüfabläufe für Einsteiger zugänglicher machen können. Zudem planen wir zusätzlich eine Reihe konkreter Beispiele von KI-Anwendungen in der zerstörungsfreien Prüfung zu präsentieren. Dadurch wollen wir die Anwendbarkeit und Vorteile von KI in diesem Bereich verdeutlichen und einen praktischen Einblick in die vielfältigen Einsatzmöglichkeiten bieten.

Schließlich möchten wir die Gelegenheit nutzen und Sie motivieren sich an der Weiterentwicklung des Merkblattes zu beteiligen. Wenn Sie interessante Anwendungsbeispiele haben, die künstliche Intelligenz, maschinelles Lernen und insbesondere Deep Learning verwenden, melden Sie sich gerne bei uns. Außerdem sind wir an Ihrer Rückmeldung interessiert, wie wir dieses Merkblatt verbessern können, welche Punkte nicht verständlich genug formuliert sind und welche Aspekte Sie interessieren würden.

5 Literaturverzeichnis

[1] A. Burkov, Machine Learning Engineering, True Positive Incorporated, 2020.

[2] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn und A. Zissermann, „The PASCAL Visual Object Classes Challenge: A Retrospective," International Journal of Computer Vision, pp. 98 - 136, 2015.

[3] C. Müller, M. Scharmach und F. Fücsök, „Measuring of the Reliability of NDE," in 8th International Conference of the Slovenian Society For NDT, Portoroz, Slovenia, 2005.

[4] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel und D. Hassabis, „Mastering the game of Go with deep neural networks and tree search," Nature, Bd. 529, pp. 484-489, 2016.

[5] I. J. Goodfellow, Y. Bengio und A. Courville, Deep Learning, Cambridge: MIT Press, 2016.

[6] D. Kahneman, O. Sibony und C. R. Sunstein, Noise: A Flaw in Human Judgment, William Collins, 2021.

[7] European Union Aviation Safety Agency, EASA Concept Paper: First usable guidance for Level 1 machine learning applications, Cologne, 2021.

[8] European Network for Inspection & Qualification, Qualification of Non-Destructive Testing Systems that Make Use of Machine Learning, Bd. ENIQ Report No. 65, Brussels: SNETP Association, 2021.

[9] M. Poretschkin, A. Schmitz, M. Akila, L. Adilova, D. Becker, A. B. Cremers, D. Hecker, S. Houben, M. Mock, J. Rosenzweig, J. Sicking, E. Schulz, A. Voss und S. Wrobel, Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz, Fraunhofer IAIS, 2021.

[10] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li und L. Fei-Fei, „Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009.

[11] C. G. Northcutt , A. Athalye und J. Mueller, „Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks," in 35th Conference on Neural Information Processing Systems, 2021.

[12] A. Ng, Interviewee, A Chat with Andrew on MLOps: From Model-centric to Data-centric AI. [Interview]. 2021.

[13] J. Yosinski, J. Clune, Y. Bengio und H. Lipson, „How transferable are features in deep neural networks?," in Neural Information Processing Systems, 2014.

6 Glossar

Annotation	Siehe "Label"
Anomalieerkennung	Das automatisierte Erkennen von ungewöhnlichen Mustern, Ausreißern oder potenziellen Anzeigen, ohne diese näher zu spezifizieren oder kategorisieren wird als Anomalieerkennung (engl. „anomaly detection“) bezeichnet. Dies ist besonders nützlich, wenn (noch) nicht alle potenziellen Anzeigen bekannt sind oder deren Variation zu groß ist. Dazu wird bei der Anomalieerkennung auf Abweichungen zum etablierten Normalzustand geachtet.
Bestärkendes Lernen	Das bestärkende Lernen (engl. „reinforcement learning“) basiert auf dem Prinzip von Belohnung und Bestrafung, die in regelmäßigen Abständen (nicht notwendigerweise nach jedem Schritt) verteilt werden. Das Ziel besteht darin eine Strategie zu erlernen, am Ende eine möglichst hohe akkumulierte Belohnung zu erhalten.
Bewertungsmetrik	Sammelbegriff verschiedener Maße zur Bewertung der Güte der Ergebnisse eines KI-Modells oder eines KI-Systems. Die Bewertungsmetriken des KI-Modells entstammen meist dem Bereich des maschinellen Lernens, während die Bewertungsmetriken des KI-Systems einen starken Anwendungsbezug haben.
Big Data	Der Begriff umfasst alle Anwendungen bei denen Daten in so großen Mengen anfallen, dass sie nicht mehr manuell verarbeitet werden können. Daher tritt er häufig in Verbindung mit maschinellem Lernen zur automatisierten Auswertung auf.
Datensatz	Ein Datensatz bezeichnet die Gesamtheit der Daten, die in einer strukturierten Form verfügbar sind. Dies schließt sowohl die gemessenen Daten als auch die zugehörigen Label mit ein.
Datensatzerweiterung	Die Datenerweiterung (engl. „data augmentation“) ist eine gezielte Modifikation der vorhandenen Trainingsdaten, um den Trainingsdatensatz zu erweitern und die Robustheit des KI-Modells zu steigern. Im Bereich der Bildverarbeitung werden Bilder z. B. rotiert, gespiegelt oder weichgezeichnet. Je nach Anwendungsfall sind nicht immer alle Modifikationen geeignet.
Datenwissenschaften	Die Datenwissenschaften (engl. „data science“) beschäftigen sich mit der automatisierten Informationsgewinnung aus großen Datenmengen mithilfe mathematisch statistischer Mittel und Methoden des maschinellen Lernens.
Deep Learning	Das tiefe oder tiefgreifende Lernen oder englisch Deep Learning, bezeichnet eine Unterkategorie des maschinellen Lernens, die sich mit der Verwendung tiefer neuronaler Netze befasst.
Detektionswahrscheinlichkeit	Unter der Detektionswahrscheinlichkeit wird die Wahrscheinlichkeit verstanden, dass ein Merkmal einer gegebenen Größe in den gegebenen Daten gefunden wird.
F-Maß	Das F-Maß (engl. „F-score“) liefert eine Aussage über die Genauigkeit eines Models und berechnet sich als gewichtetes harmonisches Mittel aus Precision und Recall.
Falsch negativ	Sagt ein KI-Modell vorher, dass alles in Ordnung sei, in Wahrheit aber eine Auffälligkeit vorliegt, ist dies ein falsch negatives (engl. „false negative“) Ergebnis. Gelegentlich werden auch Begriffe wie „Escape“ oder „Slip“ verwendet.
Falsch positiv	Sagt ein KI-Modell vorher, dass es sich um eine Auffälligkeit handelt, aber eigentlich alles in Ordnung ist, ist dies ein falsch positives (engl. „false positive“) Ergebnis. Gelegentlich wird auch der Begriff „Fehlalarm“ oder „Pseudo“ verwendet.
Falsch-Positiv-Rate	Die Falsch-Positiv-Rate oder Falsch-Alarm-Rate (engl. „false alarm rate“) bezeichnet eine Bewertungsmetrik, die angibt, welcher Anteil der tatsächlich als in Ordnung zu markierenden Elemente fälschlicherweise als Auffälligkeit erkannt wurden.
Freier Parameter	Die freien Parameter des KI-Modell werden während des Trainingsprozesses an die Daten des Trainingsdatensatzes angepasst.

Genauigkeit	Die Genauigkeit (engl. „precision“) beschreibt den Anteil relevanter Treffer in der Menge der Vorhersagen. Je näher der Wert an 1 ist, desto weniger falsch positive (Fehlalarme) liefert die Methode. Da dieses Maß keine Aussage über die Menge der gefundenen relevanten Elemente liefert sollte dieses Maß stets in Verbindung mit der Trefferquote berücksichtigt werden.
Grenzwertoptimierungskurve	Die Grenzwertoptimierungskurve, auch ROC-Kurve (engl. „receiver-operating-characteristic“) stellt die Leistungsfähigkeit eines binären KI-Modells unter Variation eines bestimmten Parameters dar. Sie zeigt den Kompromiss zwischen Sensitivität und Spezifität.
Ground Truth	Die Ground Truth (im Deutschen oft als „Grundgesamtheit“ bezeichnet) beschreibt das zu erwartende Ergebnis der KI-Modelle für einen gegebenen Datensatz.
Hyperparameter	Die Hyperparameter beschreiben den Aufbau des KI-Modells. Sie werden vom Entwickler (z. B. Data Scientist oder Machine Learning Engineer) vorgegeben und werden nicht durch den Trainingsprozess beeinflusst.
Intersection over Union	Die Intersection over Union (oder Jaccard-Index) beschreibt ein Ähnlichkeitsmaß von Mengen und ist definiert als Quotient aus Schnittmenge und Vereinigungsmenge. Je näher der Wert an 1 liegt, desto größer die Ähnlichkeit. Im Gegensatz zur Genauigkeit berücksichtigt dieses Maß keine True Negatives und eignet sich daher auch als Maß für unausgeglichene Datensätze.
KI-Entwicklung	Siehe „KI-Projekt“
KI-Modell	Das KI-Modell stellt den eigentlichen Kern des KI-Systems dar, das auf Basis der bereitgestellten Eingaben eine Aussage trifft.
KI-Projekt	Ein KI-Projekt beschreibt einen Zyklus in der Entwicklung eines KI-Systems von der Analyse der Fragestellung, über den Aufbau eines Datensatzes und dem Trainieren eines KI-Modells bis hin zur Auswertung und dem (ersten) Deployment in der produktiven Umgebung. Für Nachbesserungen, z. B. weil neue Anzeigenbilder auftreten, wird in der Regel ein neues KI-Projekt gestartet.
KI-System	Das KI-System beschreibt die Gesamtheit aus KI-Modell, bildgebenden Verfahren und Auswertung.
Konfusionsmatrix	Die Konfusionsmatrix (engl. „confusion matrix“) ist ein definiertes Tabellenlayout, das die vorhergesagten Ergebnisse in die Kategorien richtig positiv, richtig negativ, falsch positiv und falsch negativ unterteilt. Gibt es mehrere Klassen, werden die Ergebnisse den einzelnen Klassen zugeordnet. So lassen sich typische Verwechslungen zwischen den Klassen einfach erkennen.
Kreuzvalidierung	Bei der k-fachen Kreuzvalidierung (engl. „crossvalidation“) werden die verfügbaren Daten in k gleiche Teile zerlegt. Anschließend werden KI-Modelle trainiert, wobei je einer der k Teile als Validierungsdatensatz außenvor gelassen wird. Verhalten sich alle k KI-Modelle etwa gleich ist das ein gutes Zeichen für das Training.
Künstliche Intelligenz	Künstliche Intelligenz (engl. „artificial intelligence“) ist ein sehr weit gefasster Begriff und bezieht sich auf alle Computersysteme und Algorithmen, die in der Lage sind, Aufgaben auszuführen, die norma-lerweise ein menschliches Eingreifen erfordern.
Label	Ein Label ist eine eindeutige Kennzeichnung oder Kategorie, die einem Datenpunkt zugeordnet wird, um diesen zu klassifizieren oder zu kategorisieren.
Maschinelles Lernen	Maschinelles Lernen (engl. „machine learning“) ist ein Teilbereich der künstlichen Intelligenz, der sich auf die Entwicklung von Algorithmen und Modellen konzentriert, die es Computern ermöglichen, aus Daten zu lernen und automatisch Muster und Zusammenhänge zu erkennen, ohne explizite Programmierung. Diese Modelle werden in verschiedenen Anwendungen eingesetzt, um Vorhersagen, Mustererkennung und Entscheidungsfindung zu unterstützen.

Merkmal	In der ZfP werden gesuchte Eigenschaften und Strukturen im zu verarbeitenden Signal als Merkmal bezeichnet. Dies können z. B. Poren in Bilddaten oder signifikante Spitzen in einem kontinuierlichen Signal sein. Im Bereich des Maschinellen Lernens bezeichnet ein Merkmal eine beliebige messbare Eigenschaft in den Daten, dies können z. B. alle Kanten oder Ecken in Bilddaten sein. Ein ML-Merkmal kann dabei auch nur die Beschreibung einer Eigenschaft eines ZfP-Merkmals sein.
Metrik	Siehe „Qualitätskriterium“
Minimal umgebendes Rechteck	Ein Rechteck, das an den Bildachsen ausgerichtet ist, ein gesuchtes Merkmal enthält und möglichst wenig der direkten Umgebung im Bild einschließt, wird minimal umgebendes Rechteck genannt (engl. „bounding box“).
Modellübertragung	Die Modellübertragung (engl. „transfer learning“) ist ein Ansatz im maschinellen Lernen, bei dem ein KI-Modell, das auf eine bestimmte Aufgabe trainiert wurde, auf eine andere, verwandte Aufgabe angewendet wird. Anstatt das KI-Modell von Grund auf neu zu trainieren, nutzt die Modellübertragung die bereits erworbenen Fähigkeiten und Kenntnisse des bestehenden KI-Modells, um die Leistung auf der neuen Aufgabe zu verbessern.
Nichtüberwachtes Lernen	Sind für einen Datensatz keine Label vorhanden, lassen sich dennoch Muster in den Daten erkennen, die es bspw. erlauben die Daten zu gruppieren. Da kein Soll-Ergebnis bekannt ist mit dem das Ergebnis des KI-Modells verglichen werden kann, wird von unüberwachtem Lernen (engl. „unsupervised learning“) gesprochen.
Qualitätskriterium	Kriterium zur Bestimmung der Qualität (oder Fähigkeit) eines KI-Modells, z. B. Intersection over Union, Genauigkeit oder Trefferquote.
Rauschen	Unter Rauschen (engl. „noise“) wird eine ungewollte Variation in Daten (oder Labeln) verstanden.
Richtig negativ	Sagt ein KI-Modell vorher, dass alles in Ordnung sei und dies entspricht der Wahrheit, ist dies ein richtig negatives (engl. „true negative“) Ergebnis.
Richtig positiv	Sagt ein KI-Modell vorher, dass eine Anzeige vorliegt und dies entspricht der Wahrheit, ist dies ein richtig positives (engl. „true positive“) Ergebnis.
Richtigkeit	Die Richtigkeit (engl. „accuracy“) ist der Quotient aus allen richtigen Vorhersagen (Anzeigen und Gutbereiche!) und allen Vorhersagen. Da alle richtigen Vorhersagen berücksichtigt werden, kann es sein, dass sich der relevante Unterschied in nur einem kleinen Wertebereich abspielt: Bei 1 % Fehlerwahrscheinlichkeit erhält ein Klassifikator, der alles als gut markiert eine Richtigkeit von 99 %.
Sensitivität	Siehe „Trefferquote“
Systematischer Fehler	Eine systematische Abweichung oder Verzerrung (engl. „bias“) im Datensatz im Vergleich zur realen Welt wird als systematischer Fehler bezeichnet. Dieser kann sich in unterschiedlichsten Aspekten äußern, wie unterrepräsentierte Anzeigen oder immer größer gelabelten Merkmalen und kann daher zu Fehlern in den Vorhersagen des trainierten KI-Modells führen.
Testdaten	Der Anteil der Daten im Datensatz, der nicht für das Training (oder der Validierung während der Entwicklung des Modells) sondern ausschließlich für den abschließenden Test verwendet wird, heißt Testdatensatz (engl. „test data“).
Trainingsdaten	Der Anteil der Daten im Datensatz, der für das Training des KI-Modells verwendet werden soll, heißt Trainingsdatensatz (engl. „training data“).
Trainingspaket	Die Trainingsdaten können nicht auf einmal verarbeitet werden und werden daher in Pakete unterteilt. Ein Trainingspaket (engl. „batch“) ist daher die Menge an Daten, die gleichzeitig vom KI-Modell verarbeitet werden kann während des Trainings.
Trainingszyklus	Hat das KI-Modell während des Trainings alle verfügbaren Trainingsdaten einmal gesehen ist ein Trainingszyklus (engl. „epoch“) beendet.

Trefferquote	Die Trefferquote (engl. „recall“) beschreibt den Anteil der gefundenen Treffer in der Menge der möglichen Treffer. Je näher der Wert an 1 ist, desto weniger falsch negative (Escapes, Slips) liefert die Methode. Da dieses Maß keine Aussage über den irrelevanten Anteil in der Vorhersage (falsch positive Element) liefert sollte dieses Maß stets in Verbindung mit der Genauigkeit berücksichtigt werden.
Überanpassung	Umgangssprachlich wird oft vom Auswendiglernen des Trainingdatensatzes gesprochen. Die Überanpassung (engl. „overfitting“) beschreibt den Zustand, dass der Fehler auf den Trainingsdaten deutlich niedriger ist als auf den Validierungsdaten ist.
Überwachtes Lernen	Wird ein KI-Modell mit gelabelten Daten trainiert, für die das Soll-Ergebnis bereits bekannt ist, spricht man vom überwachten Lernen (engl. „supervised learning“). Dabei soll das KI-Modell die im Trainingsset vorhandenen Beispiele so gut wie möglich reproduzieren können und dies auf ungesehene Daten anwenden können. Es gilt beim überwachten Lernen eine Überanpassung auf den Trainingsdatensatz zu vermeiden, was durch einen geeigneten Validierungsdatensatz sichergestellt wird.
Unteranpassung	Das KI-Modell hat nicht die Möglichkeit alle Merkmale eines Trainingsdatensatzes zu erfassen, es ist daher unterangepasst (engl. „underfitting“).
Validierungsdaten	Der Anteil an Daten, der nicht in das Training mit einfließt, sondern ausschließlich zur Beurteilung des trainierten Modells verwendet wird, wird als Validierungsdatensatz (engl. „validation data“) bezeichnet. Siehe auch Kreuzvalidierung.
Verlustfunktion	Eine Verlustfunktion (engl. „loss function“) ordnet auf Basis der Ground Truth jeder Vorhersage des KI-Modells einen Verlust zu, um den die Vorhersage von der Ground Truth abweicht. Dieser Verlust wird während des Trainingsprozesses minimiert.
Verzerrung	Siehe „Systematischer Fehler“
Verzerrung-Varianz-Dilemma	Das Verzerrung-Varianz-Dilemma (engl. „bias-variance-tradoff“) beschreibt die gleichzeitige Minimierung zweier konträrer Einflüsse: Zum einen muss der systematische Fehler oder die Verzerrung berücksichtigt werden, dem KI-Modell muss es möglich sein von den Daten die Beziehung zur Ground Truth zu modellieren; zum anderen muss die Varianz berücksichtigt werden, sodass sich das KI-Modell beispielsweise nicht an das Rauschen in den Daten anpasst und dadurch zu sensibel gegenüber kleinerer Schwankungen in den Daten wird.
Weißen	Beim Weißen (engl. „whitening“) der Daten wird die Statistik von Daten so verändert, dass sie bestimmte Voraussetzungen erfüllt, um die Leistung von Algorithmen und KI-Modellen zu verbessern, indem beispielsweise Korrelationen zwischen den Daten reduziert werden.

7 Autoren-/Firmenverzeichnis

Dr. Carsten Brandt	**Airbus Operations GmbH, Bremen**
Dr. Nick Brierley	**diondo GmbH, Hattingen**
Christian Döpke	**Beiten Burkhardt Rechtsanwaltsgesellschaft mbH, Düsseldorf**
Christian Els*	**sentin GmbH, Bochum**
Dr. Patrick Fuchs*	**Volume Graphics GmbH, Heidelberg**
Dr. Sven Gondrom-Linke	**Volume Graphics GmbH, Heidelberg**
Dr. Frank Herold	**VisiConsult X-ray Systems & Solutions GmbH, Stockelsdorf**
Geo Jacob	**Deutsches Zentrum für Luft- und Raumfahrt e.V., Hamburg**
Prof. Dr. Ahmad Osman	**Fraunhofer IZFP, Saarbrücken**
Dr. Christiane Trela	**DB Systemtechnik GmbH, Standort Brandenburg-Kirchmöser**
Dr. Johannes Vrana	**Vrana GmbH, Rimsting**
Dr. Eric Wild	**DB Systemtechnik GmbH, Brandenburg-Kirchmöser**
Mathias Zimmer-Goertz	**Beiten Burkhardt Rechtsanwaltsgesellschaft mbH, Düsseldorf**
* HAUPTAUTOREN	

8 Bildquellennachweis

Bildnr.	Bildtitel	Autor-/in	DGZfP-VNB-Nr.
1	Oft synonym verwendet, sind die Begriffe "Künstliche Intelligenz", "Maschinelles Lernen" und "Deep Learning" eigentlich Untermengen des jeweils vorhergehenden Begriffs. Künstliche Intelligenz stellt daher einen Sammelbegriff aller Systeme dar, die automatisiert zu Entscheidungen kommen, während beim maschinellen Lernen stehts allgemeine Algorithmen auf Basis von Daten zu Entscheidungen kommen. Das Deep Learning fokussiert sich dabei auf eine spezielle Kategorie von Algorithmen, den tiefen neuronalen Netzen.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
2	Unterschiedliche Aufgabenstellungen, die ein KI-Modell in der Bildverarbeitung lösen kann. Bei der Anomalieerkennung werden Abweichungen vom Soll-Zustand gesucht, ohne diese näher zu beschreiben; bei der Klassifizierung wird das Bild als Ganzes kategorisiert; die Lokalisierung findet Bereiche im Bild und kategorisiert diese; und die semantische Segmentierung identifiziert gesuchte Objekte pixelgenau im Bild. Mit steigender Komplexität der Aufgabe, steigt auch der Aufwand für die Erstellung eines geeigneten Trainingsdatensatzes.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
3	Der Projektzyklus zur Entwicklung eines KI-Modells: Initiiert von einer Fragestellung werden zunächst ein realistisches Ziel formuliert und die Kriterien zu dessen Erreichung festgelegt. Anschließend werden die Trainingsdaten gesammelt und aufbereitet, bevor es daran geht das konkrete KI-Modell zur Lösung der definierten Aufgabenstellung zu trainieren. Das fertige KI-Modell wird dann mit Hilfe eines Testdatensatzes validiert. Erfüllen die Validierungsergebnisse die Kriterien der Zielsetzung kann das KI-Modell produktiv geschaltet und ins KI-System eingebettet werden. Treten während des Betriebs Änderungen am KI-System auf, beginnt der Zyklus von neuem.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
4	Die Tabelle auf der linken Seite zeigt eine Konfusionsmatrix. Die Formeln auf der rechten Seite zeigen, wie die Werte aus der Konfusionsmatrix zu aussagekräftigen Werten kombiniert werden. So bildet sich die Trefferquote beispielsweise aus dem Quotienten aus richtig als positiv erkannten Merkmalen und allen tatsächlich positiven Merkmalen zusammen. Deutlich ist auch der Unterschied zwischen Richtigkeit und IoU: Die IoU fokussiert sich auf die relevante Klasse und eignet sich daher auch für unausgeglichene Datensätze, bei denen es deutlich mehr negative als positive Merkmale gibt.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024

Bildnr.	Bildtitel	Autor-/in	DGZfP-VNB-Nr.
5	Eine alternative Betrachtung der Bewertungsmetriken ist die Berechnung aus den Inhalten der minimal umgebenden Rechtecke: Der richtig positive Anteil der Vorhersage ist beispielsweise die Schnittmenge des tatsächlichen und des vorhergesagten minimal umgebenden Rechtecks. Die Kombination aus richtig positiv und falsch positiv ist die Menge aller vorhergesagten Elemente und entspricht daher dem vorhergesagten minimal umgebenden Rechteck. Der Quotient ergibt dann die Genauigkeit.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
6	Bei der Erstellung eines Trainingsdatensatzes ist es wichtig, dass die Trainingsdaten die gesamte Spanne, der an tatsächlich im produktiven Einsatz auftretenden Zustände abbildet. Das heißt es müssen die gleichen Aufnahme-Modalitäten verwendet werden und die gleichen Unregelmäßigkeiten enthalten sein. Im obigen Beispiel wäre es kaum möglich, dass das trainierte KI-Modell Risse erkennt.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
7	Ein Beispiel für eine schädliche Vorverarbeitung. Die vermeintlich nicht benötigten Grauwerte in der Luft hart auf null zu setzen, erlaubt eine erhebliche Kompression der Daten. Dabei gehen aber zum einen wertvolle Informationen, z. B. über die Streifenartefakte in der Mitte des Bildes verloren, die an anderer Stelle zu falsch positiven Anzeigen führen können und zum anderen wird eine harte Kante mit einem großen Grauwertsprung in die Daten eingebracht, die die Stabilität des KI-Modells gefährden.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
8	Vereinfachte Darstellung des Gradientenabstiegsverfahren. Wird für jede Aktualisierung der Parameter der volle Trainingsdatensatz verwendet (blauer Pfad), führen die Aktualisierungen geradlinig zum nächsten Minimum. Bei der Größe der verwendeten Datensätze ist dies jedoch nicht praktikabel. Wird für jede Aktualisierung der Parameter nur ein Beispiel aus dem Trainingsdatensatz verwendet (orangener Pfad), werden wesentlich mehr Schritte benötigt, die sich aber deutlich schneller berechnen lassen. Das Mini-Batch-Verfahren, bei dem mehrere Beispiele auf einmal verwendet werden (grüner Pfad), bildet den Mittelweg. Ein optimales Ergebnis erfordert konsistent gelabelte Daten.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
9	k-fache Kreuzvalidierung. Der gelabelte Datensatz wird zunächst in einen Trainingsdatensatz und einen Testdatensatz unterteilt. Letzterer wird für die finale Evaluierung des trainierten Modells beiseitegelegt. Für die Kreuzvalidierung wird der Trainingsdatensatz in k gleiche Teile unterteilt und sukzessive ein Teil als Validierungsdatensatz zur Auswertung verwendet und auf den verbleibenden k – 1 Teilen das Modell trainiert.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
10	Verzerrung-Varianz-Dilemma. Während des Trainings wird der Fehler, den das Modell auf den Trainings- und den Validierungsdatensatz macht, beobachtet. Weichen die Fehlerwerte deutlich voneinander ab und beginnt der Fehler auf dem Validierungsdatensatz gar wieder zu steigen, liegt eine Überanpassung des Modells auf den Trainingsdatensatz vor. Sind die Fehler in etwa gleich, aber signifikant größer als das vereinbarte, tolerierbare Fehlermaß, handelt es sich um eine Unteranpassung – das Modell kann die Komplexität der Fragestellung nicht abbilden. Dazwischen gibt es den optimalen Bereich, in dem das Modell die gewünschten Ergebnisse liefert.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024
11	Je nach Anforderungen an das KI-System, ist es sinnvoll das KI-System unterschiedlich „nahe“ an der Anwendung zu installieren. „On the edge“, z. B. im bildgebenden System, liefert es die schnellsten Ergebnisse, da die Daten nicht kopiert werden müssen. Eine Cloud-Lösung dagegen bietet den Vorteil, dass das zugrundeliegende KI-Modell schnell aktualisiert und ausgetauscht werden kann und keine zusätzliche Hardware vor Ort nötig ist – dafür müssen die Daten über das Internet zum Service-Anbieter kopiert werden.	Volume Graphics GmbH, Heidelberg	VNB 284 vom 28.02.2024