

Vom Rohmaterial zum Endprodukt am Beispiel der Blechfertigung

# Optimierte Produktion durch KI und Quantencomputing

C. Nitsche, X. Wu, D. Brajovic, L. Lörcher, M. Roth, P. Wagner, A. Yaman, J. Schwab, H. Monke, C. Hennebold, M. F. Huber

**ZUSAMMENFASSUNG** Für optimierte Produktionsprozesse und weniger Materialverschwendung in Fertigungen werden Künstliche Intelligenz (KI) und Quantencomputing immer wichtiger. Diese Methoden kommen über die gesamte Prozesskette von der initialen Blechproduktion über das Blechschneiden und -schweißen (zum Beispiel mittels Laser) bis hin zum abschließenden Lackieren der gefertigten Teile zum Einsatz. Dazu gehören auch Ansätze wie erklärbare KI für transparente Modelle, Quantencomputing zur Materialoptimierung, physikbasierte KI-Modelle für präzise Schneide- und Schweißprozesse sowie generative KI als virtueller Assistent und zur Generierung synthetischer Datensätze.

## STICHWÖRTER

Fertigungstechnik, Industrie 4.0, Künstliche Intelligenz

## 1 Einleitung

Die Industrie steht vor der Herausforderung, Produktionsprozesse effizienter, flexibler und qualitativ hochwertiger zu gestalten. Künstliche Intelligenz (KI) bietet hierfür vielversprechende Lösungen, welche sich in der Praxis bereits bewähren. Laut einer Studie [1] des Fraunhofer-Zentrums für Internationales Management und Wissensökonomie IMW nutzen bereits 45 % der befragten Unternehmen KI-Anwendungen in ihren Dienstleistungen. Dabei zeigt sich, dass KI-basierte Systeme nicht nur die Effizienz und Qualität von Produktionsprozessen steigern können, sondern auch dabei helfen, die Produktionsqualität zu verbessern.

Im Rahmen des KI-Fortschrittszentrums [2] hat das Fraunhofer IPA im Verbund mit dem Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO schon mit über 300 Firmen KI-Projekte durchgeführt. Auch demonstrieren aktuelle Forschungsarbeiten des Fraunhofer IPA, wie verschiedene KI-Methoden, so etwa Maschinelles Lernen (ML), Quantencomputing, generative KI und physikbasierte Modelle, in der Blechverarbeitung eingesetzt werden können. Vor allem beim Laserschneiden von Metallblechen bietet der Einsatz von KI großes Potenzial, um Prozesse zu optimieren und die Produktqualität zu verbessern. Außerdem können Herausforderungen auf verschiedenen Ebenen, wie intransparente Black-Box-Modelle oder suboptimale Zuschnittsplanung in der Blechproduktion, adressiert werden. Die Integration von KI-basierten Lösungen in die Produktion ist ein

## Optimized manufacturing through AI and quantum computing

**ABSTRACT** To optimize production processes and reduce material waste in the manufacturing industry, artificial intelligence (AI) and quantum computing methods are getting more important. These methods are applied across the entire process chain from initial sheet metal production through cutting and welding (e.g., using lasers) to final painting of the fabricated parts. Amongst others, there are approaches such as explainable AI for transparent models, quantum computing for material optimization, physics-based AI models for precise cutting and welding processes, and generative AI as virtual assistant and for generating synthetic datasets.

wichtiger Schritt, um die Wettbewerbsfähigkeit der Industrie zu stärken.

Dieser Beitrag zeigt, wie KI und Quantencomputing entlang der gesamten Wertschöpfungskette der Blechproduktion integriert werden. Folgende Produktionsschritte werden betrachtet:

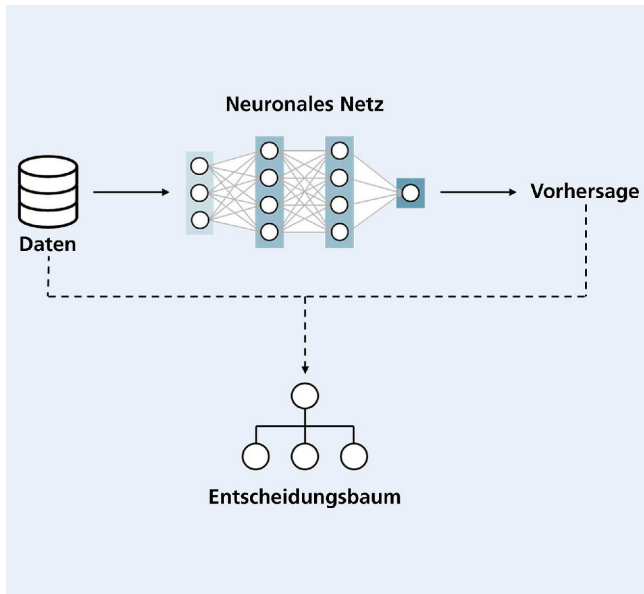
- Initiale Herstellung des Blechs
- Ausschneiden von Teilblechen
- Zusammenschweißen von Teilblechen zu einem Endprodukt
- Finaler Lackierprozess

Im ersten Kapitel wird beschrieben, wie KI die Fehler am frisch produzierten Blech erkennt, diese Fehler klassifiziert und das Klassifiziermodell in verständliche Beschreibungen umwandelt. Im zweiten Kapitel geht es um Quantencomputing zur Optimierung der Materialausnutzung beim Zuschnitt der Bleche. Die weiteren Kapitel behandeln den Einsatz von KI beim Schneiden, Schweißen und Lackieren der Teilbleche. Abschließend wird im sechsten Kapitel der Einsatz von generativer KI als virtueller Assistent über alle Produktionsschritte hinweg beschrieben.

## 2 Einsatz von KI und Quantencomputing in der Blechproduktion

### 2.1 Einsatz von KI und Methoden der Erklärbarkeit in der Produktion des Rohbleches

Die Herstellung von Blechteilen beginnt damit, die Metall- und Oberflächenqualität sicherzustellen. Durch den Einsatz von KI-gestützter optischer Inspektion können Hersteller Defekte wie



**Bild 1.** Schematisches Vorgehen zur Erstellung eines Stellvertretermodells (Entscheidungsbaum) eines künstlichen neuronalen Netzes (KNN).

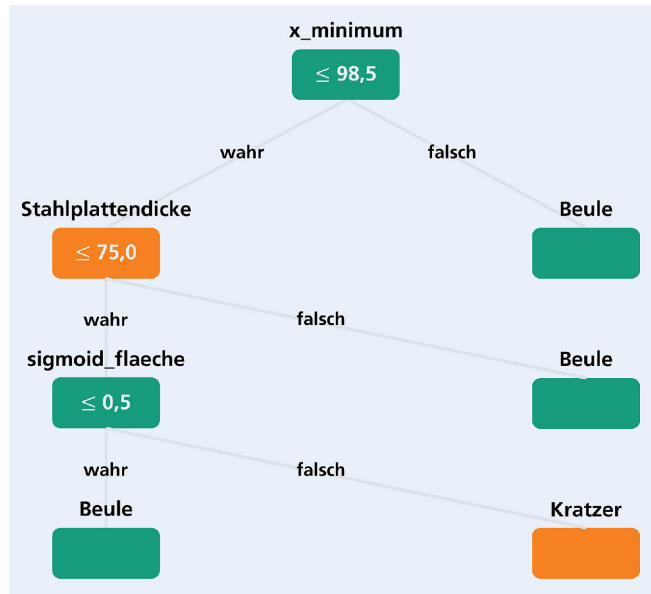
Grafik: Fraunhofer IPA

Beulen oder Kratzer früh im Produktionsprozess identifizieren und das Material gegebenenfalls aussortieren. Allerdings sind die Entscheidungen von KI-basierten Anwendungen häufig selbst für Experten nicht vollständig nachvollziehbar. Methoden der erklärbar KI (xAI: eXplainable Artificial Intelligence) machen undurchsichtige Vorhersagen des KI-Modells nachvollziehbar und steigern so das Vertrauen in dessen Entscheidungen. Zudem wird xAI auch im Zuge zunehmender Regulierungsmaßnahmen immer wichtiger. Es stehen verschiedene xAI-Methoden zur Verfügung, die je nach Anwendungskontext und Zielgruppe unterschiedlich gut passen. Ein entsprechendes Vorgehen verdeutlicht die im Folgenden beschriebene xAI-Lösung [3].

Ausgangspunkt ist ein künstliches neuronales Netz (KNN), das darauf trainiert wird, Defekte auf Stahlplatten zu klassifizieren. Dazu nehmen Sensoren verschiedene Merkmale der Fehler auf den Stahlplatten auf (wie Maße, Dicke oder verschiedene Helligkeitswerte der Platte), die als Eingangswerte des KNNs genutzt werden. Als Ausgabewert liefert das KNN den Fehlertyp.

Das resultierende Modell klassifiziert die Stahlplatten zwar sehr genau, es ist aber nicht klar, wie oder anhand welcher Merkmale das Modell seine Entscheidungen getroffen hat. Dies führt dazu, dass das Vertrauen in die KI-Lösung nicht besonders hoch ist. Um das Vertrauen zu steigern und die Entscheidungsfindung des Modells besser verstehen zu können, wird es durch ein Surrogatmodell, also ein Stellvertretermodell, das die Logik des KNNs abbildet, approximiert. In diesem Fall ist das Surrogatmodell ein einfacher Entscheidungsbaum, der anhand der Ausgabe des KNNs trainiert wird (**Bild 1**).

Die ursprünglichen Labels werden durch die Prädiktionen des neuronalen Netzes ersetzt. Außerdem werden beim Training zusätzliche Regularisierungsterme eingesetzt, mit denen Einfluss auf die Struktur des Netzes genommen wird [4]. Diese beiden Maßnahmen stellen sicher, dass der Entscheidungsbaum zum einen nicht zu groß wird und damit für den Menschen überblickbar ist. Zum anderen ist der Baum dem neuronalen Netz möglichst ähnlich und hat dadurch auch eine ähnlich hohe Genauigkeit bei der



**Bild 2.** Resultierendes Surrogatmodell (Entscheidungsbaum), das Entscheidungsregeln visualisiert. Der linke Ast eines Entscheidungsbaums stellt die Erfüllung der Bedingung im Knoten dar, der rechte Ast die Nichterfüllung. Jeder Knoten repräsentiert also eine Teilung der Daten und die jeweilige Farbe zeigt an, welche Klasse bei dieser Teilung dominiert (orange = Kratzer, grün = Beule). Grafik: Fraunhofer IPA

Klassifikation. Aus dem resultierenden Entscheidungsbaum (**Bild 2**) lassen sich dann verschiedene Erklärungen extrahieren.

Erstens wird eine globale Erklärung generiert, welche durch Entscheidungsregeln, Merkmalswichtigkeiten und Entscheidungsebenen darauf abzielt, das Modell als Ganzes zu erklären. Diese Entscheidungsregeln können mit bestehenden Vorgaben oder vorhandenem Expertenwissen abgeglichen werden, sodass eine gewisse Validierung des Systems möglich ist. Zweitens wird für jede konkrete Entscheidung eine lokale Erklärung ausgegeben, zum Beispiel durch Klassenverteilungen und Relevanzen in einzelnen Entscheidungsknoten im Baum.

Aus dem Entscheidungsbaum in Bild 2 lässt sich zum Beispiel schließen, dass das Merkmal „x\_minimum“, das minimale Maß des Defekts in x-Richtung, das wichtigste Merkmal für die Klasseneinteilung ist, da es für die Entscheidung im ersten Knoten verwendet wird. Die Merkmale „stahlplattendicke“ und „sigmoid\_flaeche“ sind die zweit- und dritt wichtigsten Merkmale für die Klassifikation. Nach Inspektion des gesamten Entscheidungsbaums ergibt sich beispielsweise folgende Entscheidungsregel: „Wenn x\_minimum 98,5 und stahlplattendicke > 75,0, dann wird der Defekt als Beule klassifiziert.“ Insgesamt steigert das Trainieren eines Surrogatmodells die Transparenz des Modells und das Vertrauen in das KI-System, was für die Akzeptanz und den erfolgreichen Einsatz von KI in der Industrie entscheidend ist.

Neben Surrogatmodellen gibt es zahlreiche weitere xAI-Ansätze, welche die Modellentscheidung des KNNs im vorliegenden Anwendungsfall erklären können. Populär sind etwa SHAP (Shapley Additive exPlanations) [5], ein lokales Verfahren, das auf spieltheoretischen Shapley-Werten basiert [6], oder Kontrafakte [7], ein Ansatz, der besonders menschenfreundliche Erklärungen generiert [8], indem er aufzeigt, welche minimalen Änderungen an den Eingabedaten notwendig wären, um ein anderes Vorhersageergebnis zu erzielen. Auf diese Weise helfen Kontrafakte, zu verstehen, unter welchen Bedingungen die Modellentscheidung

anders ausgefallen wäre. Insgesamt hängt der optimale xAI-Ansatz von vielen Faktoren wie der gewählten Modellarchitektur, dem betrachteten Anwendungsfall und dem Adressaten der Erklärung ab. Daher ist es wichtig, die spezifischen Anforderungen und Ziele der Erklärung in jedem Einzelfall zu analysieren. Verschiedene Ansätze können unterschiedliche Aspekte der Modellvorhersage beleuchten und so zu einem umfassenderen Verständnis beitragen.

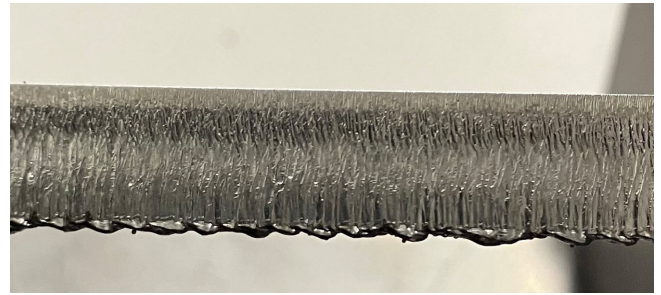
## 2.2 Optimierung des Zuschnittsplans für Teilbleche durch Quantencomputing

Sobald qualitativ hochwertiges Blech bereitsteht, ist die nächste Herausforderung im Fertigungsprozess, kleine, unterschiedlich geformte Stücke mit minimalem Abfall aus den großen Blechen zu schneiden. Dies ist eine Optimierungsaufgabe, deren Lösung mit Quantencomputing untersucht wird. Quantencomputing bezeichnet das Rechnen mit quantenmechanischen Zweizustandssystemen (Qubits). Unter Ausnutzung quantenmechanischer Eigenschaften wie Superposition und Verschränkung können Algorithmen entwickelt werden, die exponentielle Laufzeitvorteile gegenüber ihren klassischen Analoga aufweisen.

Obwohl bestehende Optimierungsverfahren und Heuristiken bereits praktikable Lösungen bieten, können diese in komplexeren Szenarien an ihre Grenzen stoßen, besonders wenn die Anzahl der Variablen und Einschränkungen zunimmt. Quantencomputing bietet das Potenzial, solche hochdimensionalen Probleme effizienter zu lösen. Der Mehrwert liegt in der Fähigkeit, größere und komplexere Problemstellungen in kürzerer Zeit und mit potenziell besseren Ergebnissen zu bewältigen. Ob und in welcher Art und Weise sich solche Vorteile tatsächlich realisieren lassen, wird sich erst in den kommenden Jahren mit zunehmender Reife der Technologie zeigen.

Auch wenn die auf derzeit verfügbaren Quantencomputern lauffähigen Algorithmen noch nicht diese herausragenden Vorteile aufweisen, lassen sie sich bereits auf reale Probleme anwenden. Für eine optimierte Materialausnutzung in der Blechfertigung wurde am Fraunhofer IPA ein heuristischer Algorithmus entwickelt, welcher das Problem in kleinere Teile zerlegt und diese mit quantenbasierter Optimierung löst [9]. Ziel ist es, bereits beim derzeitigen Technologiereifegrad Wege aufzuzeigen, wie Quantencomputing in einen produktiven Kontext integriert werden kann.

Der Algorithmus erstellt Cluster von Teilen, die räumlich so orientiert werden können, dass beim Schneiden möglichst wenig Verschnitt entsteht. Das Auffinden der passenden Kombinationen wird auf das „Traveling-Sales-Person“-Problem zurückgeführt, welches anschließend mit einem Quantencomputer gelöst werden kann. Die so entstehenden Cluster werden dann materialeffizient angeordnet. Der Gesamtalgorithmus basiert also auf einer Kombination klassischer und quantenmechanischer Methoden. Diese hybride Nutzung beider Systeme ist typisch für aktuelle Quantenalgorithmen, da zum einen derzeitige Quantencomputer noch fehlerbehaftet sind. Zum anderen sind auch sie zu klein, um große Probleme eigenständig zu lösen. Die zugrunde liegende Motivation ist, hierdurch die rechnerische Robustheit der bekannten Computersysteme zu nutzen und nur perspektivisch vielversprechende Teilaspekte auf den Quantencomputer auszulagern, um so einen guten Mittelweg zwischen potenziellem Vorteil durch Quantenressourcen und Stabilität zu erreichen [10].



**Bild 3.** Schnittkante eines Stahlblechs mit hohem Grat und hoher Rauigkeit. Foto: Trumpf Group

## 2.3 Optimierung des Blechschneideprozesses durch KI

Nach erfolgter Zuschnittsplanung müssen die Teilbleche aus dem Standardblech ausgeschnitten werden. Dabei kommt immer häufiger die Lasertechnik zum Einsatz. Die Präzision im Laserschneiden von Blechteilen ist ein entscheidender Faktor in der modernen Fertigungstechnik. Durch fortschrittliche, von KI unterstützte Laserschneidmaschinen erreicht dieser Prozess eine nie dagewesene Genauigkeit und Effizienz. KI-Systeme steuern die Maschinen so, dass sie nicht nur präzise Schnitte gewährleisten, sondern auch die Materialausnutzung optimieren, was zu enormen Einsparungen an Rohstoffen und Kosten führt. In **Bild 3** ist eine Schnittkante mit suboptimaler Qualität abgebildet. Der Grat der Schnittkante ist recht hoch und die Kante ist sehr rau.

Die Verwendung von physikbasierten Modellen ermöglicht den KI-Systemen, die Interaktionen zwischen dem Laserstrahl und dem Metall genau zu simulieren, wie eine Studie des Fraunhofer-Instituts für Lasertechnik ILT darstellt [11]. Diese Modelle ermöglichen, für jeden Schnitt die optimalen Betriebsbedingungen zu prognostizieren. So können Laserleistung, Schnittgeschwindigkeit und Fokus genau justiert werden, um die Schnittqualität zu maximieren und Materialverschwendung zu minimieren.

Zudem verbessern adaptive KI-Algorithmen die Maschinenparameter kontinuierlich, indem sie aus Produktionsdaten lernen und auf Änderungen in der Materialbeschaffenheit oder Designanforderungen reagieren. Diese Algorithmen optimieren dynamisch die Einstellungen wie Schnittgeschwindigkeit und Laserleistung, um stets die besten Schnittergebnisse zu erzielen [12]. Das ist besonders wichtig in einer Produktionsumgebung, in der Flexibilität und schnelle Anpassungsfähigkeit an neue Designs gefordert sind.

Eine weitere Studie des Fraunhofer IPA [13] zeigt, dass der Einsatz von Faltungsnetzen (englisch: convolutional neural networks) eine präzise Beurteilung der Qualität thermisch geschnittener Kanten aus Bildern ermöglicht. Dieses Vorgehen erlaubt eine schnelle und genaue Vorhersage der Produktionsqualität wie Kantenrauheit und Grathöhe, was die Effizienz des Fertigungsprozesses weiter verbessern und zu einer noch präziseren Materialbearbeitung führen kann.

## 2.4 Optimierung des Blechschweißprozesses durch KI

In Szenarien, in denen alleiniges Schneiden nicht ausreicht, wird das Laserschweißen eingesetzt, um die geschnittenen Teile zu einer Endstruktur zusammenzufügen, wie etwa Autoteile. Dieser Schritt umfasst oft komplexe Form- und Pressprozesse, die

weiter durch KI-Technologien optimiert werden, um Präzision und Festigkeit im Endprodukt zu gewährleisten.

Oft sind die genauen Wirkzusammenhänge zwischen Prozessparametern – wie Laserleistung, Vorschub, Brennweite, Lage des Fokus relativ zur Werkstückoberfläche – und Qualität der Schweißnaht unbekannt. So führen Prozessschwankungen und unbekannte Wechselwirkungen aufgrund der chaotischen Natur des Lasersystems zu unterschiedlichen Qualitätsstufen, etwa mit Fehlern wie Schweißdurchbrüchen [14]. Somit ist die Frage, welcher Prozesszustand zu einem Gut-Teilfenster – die Prozessparameterkombinationen, die ein OK-Teil hervorbringen – führt, um eine optimale und verlässliche Qualität auch bei schwankenden Prozessparametern zu erreichen, ohne dass es zu Schweißdurchbrüchen kommt.

Um diese komplexen Zusammenhänge im realen Prozess aufzudecken, sind systematische Experimente sehr aufwendig und auch oftmals unzulänglich. Denn sie verursachen eine enorme Datenmenge durch optische und thermische Hochgeschwindigkeitskameras, Röntgen-Computertomographie sowie akustische und optische Sensoren [15]. Außerdem sind Wechselwirkungen nur zum Teil als Domänenwissen bekannt und in weiten Teilen nicht oder unzureichend konkret beschrieben. Dies ist darauf zurückzuführen, dass Laserschweißen ein hoch komplexer Herstellungsprozess ist, der durch eine Reihe von Prozessvariablen und andere Faktoren beeinflusst wird, wie Defekte in der Mikrostruktur des Materials, Verunreinigungen auf der Oberfläche des Werkstücks und Veränderungen der Eigenschaften des Laserstrahls [16].

Ein neuer Ansatz, bei dem ein ML-Verfahren auf Basis von Prozessdaten in konventionellen Untersuchungsprozessen zum Einsatz kommt, bietet Abhilfe. Dabei werden aufgezeichnete Prozessdaten in Kombination mit Abschätzungen durch Finite-Elemente-Simulation von lokalen oder nicht direkt zugänglichen Prozessgrößen, wie etwa Temperaturverläufen im Inneren des Materials, herangezogen und maschinelle Lernverfahren mit diesen trainiert, um die Schweißnahtqualität zu präzisieren. Für das Training werden nur ausgewählte Features genutzt, die mit einer Feature-Importance-Analyse ausgewählt wurden. Danach werden verschiedene Modelle trainiert und verglichen und dasjenige ausgewählt, welches eine möglichst hohe Prädiktionsgüte unterschiedlicher Gütekriterien auf Basis der Prozessgrößen erlaubt. Dieses Modell bietet eine Möglichkeit, simulativ zu überprüfen, inwiefern eine Änderung der Prozessparameter Auswirkungen auf die Qualitätsmerkmale hat. Mithilfe eines gradientenbasierten Suchverfahrens wird dann diejenige Parameterkombination gesucht, die laut dem Modell einen möglichst guten Prozesszustand im Hinblick auf die Qualität ergibt. So kann durch Auslegung auf diesen Prozesszustand eine möglichst hohe und robuste Qualität erzielt werden.

## 2.5 Optimierung des Blechlackierprozesses durch KI

Die nahezu fertigen Blechteile durchlaufen eine letzte Transformation im Lackierprozess, bei dem KI die Lackanwendung optimiert, um ein perfektes Finish zu erreichen. Dies verbessert nicht nur die Ästhetik, sondern fügt den Metallteilen auch eine schützende Schicht hinzu.

Eine Möglichkeit, um den Lackierprozess gezielt zu optimieren, besteht in der Berechnung von Ursache-Wirkzusammenhängen zwischen relevanten Prozessvariablen wie Lackmenge, Luft-

menge sowie Drehzahl des Glockentellers und deren Auswirkung auf das Endergebnis. Durch die gezielte Bestimmung kausaler Effekte wird im Vorfeld gezielt mögliches Optimierungspotenzial offengelegt. So konnten beispielsweise *Tezuka* und *Kuroki* [17] anhand eines kausalen Modells zeigen, welche Prozessvariablen die Transferleistung von Lack auf Autokarosserien zu welchem Grad beeinflussen, was bei der Schaffung eines stabilen Lackierprozesses sehr wichtig ist.

Einen bekannten Standardisierungsansatz zur Berechnung kausaler Effekte stellt die „DoWhy“-Python-Bibliothek dar [18]. Der Ansatz ist in einer Pipeline-Struktur in vier Schritte unterteilt:

1. Modellierung einer Graphenstruktur, die sowohl vorhandene Daten als auch Domänenwissen kombiniert.
2. Identifizieren eines Ziel-Schätzwerts, der den Zusammenhang zwischen Prozessvariablen und der Zielvariable beschreibt.
3. Berechnen des kausalen Effekts anhand des in Schritt 2 identifizierten Ziel-Schätzwerts.
4. Durchführen statistischer Tests zur Bestimmung der Robustheit des berechneten Effekts.

Der berechnete kausale Effekt eignet sich zur weiteren Analyse gut, indem aufgezeigt wird, welche Stellschrauben im Prozess die größte Auswirkung auf das Ergebnis haben. Anhand der erlangten Erkenntnisse wird die Optimierung des Lackiervorgangs wesentlich zielgerichteter durchgeführt, was auf lange Sicht den Aufwand reduziert und somit ein wesentliche Faktor bei der Kostenreduktion ist.

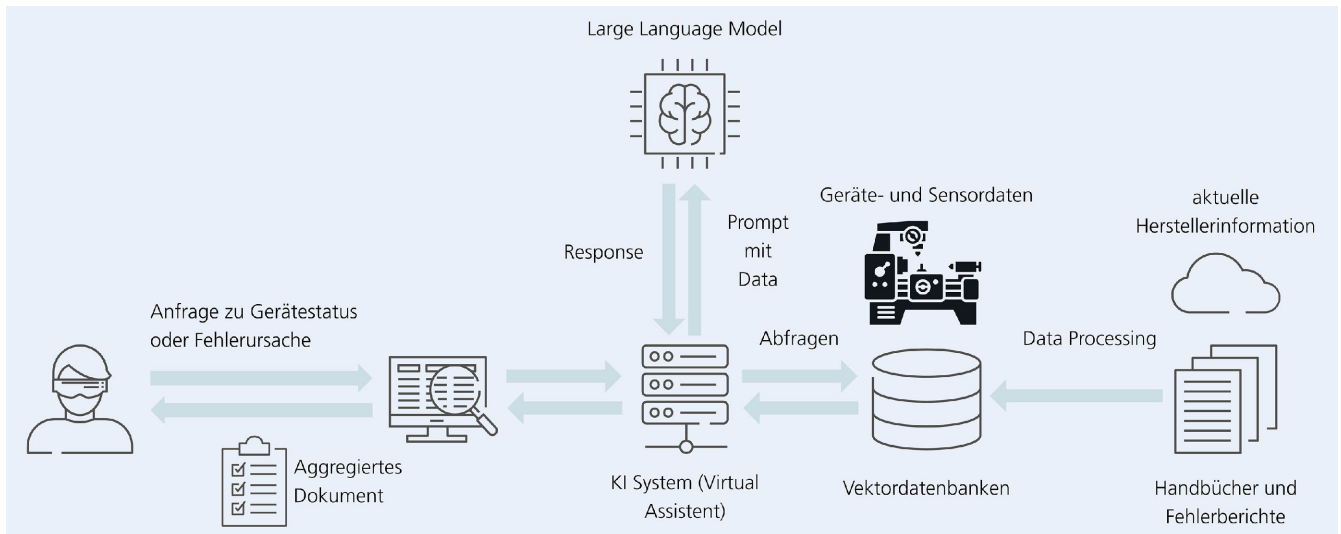
## 2.6 Einsatz von generativer KI im gesamten Blechfertigungsprozess

Generative KI ist eine Methode mit hoher Relevanz auch in der Fertigungsindustrie, bei der Modelle neue Inhalte wie realistische Bilder oder kohärente Texte generieren, während andere Arten der KI basierend auf vorhandenen Daten Vorhersagen, Klassifizierungen oder Regressionen vornehmen. Diese Technologie, die sich auf sogenannte Transformer-Modelle stützt, nutzt den Mechanismus der Selbstaufmerksamkeit [19], um die Bedeutung verschiedener „Tokens“ (das heißt Wörter oder Wortteile) in einem Text zu gewichten. Dadurch sind kontextbezogene Darstellungen möglich, die in verschiedenen Bereichen wie Sprachübersetzung, Texterzeugung und sogar Zeitreihenvorhersage Einsatz finden.

Die Entwicklung von „Generative Pre-trained Transformers“ [20] brachte erhebliche Fortschritte, welche die Generierung von Text aus Texteingaben sowie die Überbrückung von Modalitäten von Text zu Bildern ermöglichen. Ein prominentes Beispiel ist ChatGPT, das aktuell die Rolle von „Large Language Models“ (LLMs) in der industriellen Anwendung prägt. Diese Systeme, die mit Herausforderungen wie Halluzinationen und Verzerrungen, also falschen oder ungenauen Aussagen, zu kämpfen haben, werden durch Methoden wie „Retrieval-Augmented Generation“ (RAG) ergänzt. Diese verbessern die Genauigkeit und Relevanz der Ergebnisse durch die Nutzung interner und externer Wissensdatenbanken, bevor die Useringabe in das Modell eingebunden wird.

Der Umfang der LLM-Anwendungen verändert die Art und Weise, wie Unternehmen mit Informationen und Nutzern interagieren. LLMs, in Form von virtuellen Assistenten, werden zur Steigerung der Effizienz und zur Unterstützung in technischen





**Bild 4.** RAG (Retrieval-Augmented Generation)-Systemarchitektur für einen virtuellen Assistenten. Grafik: Fraunhofer IPA

Bereichen eingesetzt. Zum Beispiel sind Maschinenausfälle zeit- und produktivitätsintensiv. Virtuelle Assistenten helfen, diese Ausfallzeiten zu minimieren, indem sie sofortige Lösungen zur Fehlerbehebung anbieten. Durch den Zugriff auf umfangreiche Datenbanken mit Maschinenhandbüchern und historischen Leistungsdaten innerhalb eines RAG-Systems unterstützen sie technisches Personal mit präzisen, kontextbezogenen Reparaturstrategien, wie in **Bild 4** dargestellt. Diese Integration generativer KI in virtuelle Assistenten rationalisiert nicht nur die Wartungsroutinen, sondern verbessert auch die Fähigkeiten der Techniker, während sie lernen und ihr Fachwissen erweitern können.

Virtuelle Assistenten werden auch in der Blechfertigung immer relevanter und unterstützen bei verschiedenen Prozessen wie Schneiden, Biegen, Stanzen und Zusammenfügen der Bleche in die gewünschten Formen. Fehler in diesen Prozessen haben meist mehrere Ursachen, und es kann eine Herausforderung sein, das Problem zu identifizieren und eine Lösung zu finden. Bei Störungen in einer Blechproduktionsanlage können virtuelle Assistenten schnell alle relevanten Maschinenhandbücher und historischen Reparaturprotokolle durchforsten, um die effektivste Reparaturmethode zu empfehlen und sicherzustellen, dass die Maschine mit minimaler Verzögerung wieder in Betrieb genommen wird. Auch nutzen sie die Analyse von Daten aus Sensoren und Nutzungsprotokollen, um Ausfälle oder Wartungsbedarf vorherzusagen. Dieser Ansatz trägt dazu bei, Ausfallzeiten zu reduzieren, indem Wartungsarbeiten nur bei Bedarf geplant werden [21].

Bei der Auswahl eines bestimmten LLM für die Erstellung eines virtuellen Assistenzsystems gibt es verschiedene Optionen wie „Mistral“ [22], „Llama“ [23] oder deren für einen bestimmten Zweck angepassten Derivate. Sie unterscheiden sich in Größe, Hardwareanforderungen und Leistung. Daher muss eine sorgfältige, bedarfsorientierte Auswahl erfolgen. Für RAG-Systeme gibt es Open-Source-Lösungen, die als Basis für kundenspezifische Umsetzungen verwendet werden können, wobei sowohl Offline- als auch Cloud-Einsätze in Betracht kommen. Auch eine zunehmende Anzahl kommerzieller Angebote ist verfügbar [24, 25].

Während die zugrunde liegende Idee von RAG leicht zu verstehen ist, ist die robuste Implementierung des Gesamtsystems unter Berücksichtigung aller Teilsysteme und ihrer konfigurierbaren Parameter sowie der Datenvorverarbeitung eine Herausforderung,

auch aufgrund der hohen Varianz der Datenquellen und des endgültigen Systemzwecks. Neben der RAG ist oft auch eine Feinabstimmung eines Open-Source-Modells mit einem domänenspezifischen Datensatz nötig, um einen robusten virtuellen Assistenten zu erstellen. Dies ist mit noch mehr Herausforderungen verbunden, da die Feinabstimmung das zugrunde liegende LLM verändert und daher zusätzliche Sorgfalt erfordert. Diese Aspekte führen zu einer steigenden Nachfrage nach Experten für generative KI und Kenntnissen in allen Branchen, um solche Systeme zu erstellen, zu bewerten oder zu überwachen, selbst wenn ihr Kerngeschäft nicht im Bereich der generativen KI liegt.

Neben dem Einsatz generativer KI als virtuellem Assistenten ist auch die Fähigkeit dieser Technologie, synthetische Datensätze zu erzeugen (sowohl tabellarische Daten als auch Bilddaten), von großer Bedeutung für die Fertigungsindustrie. Dadurch können etwa für die automatische optische Inspektion von gefertigten Teilen synthetische Defektbilder erstellt werden. Diese erweitern den Trainingsdatensatz und verbessern somit die KI-gestützte Inspektion [26].

### 3 Diskussion und Ausblick

Die Implementierung von KI und Quantencomputing in der Fertigungsindustrie ist vielversprechend, jedoch gehen diese mit Herausforderungen einher. Eine zentrale Schwierigkeit liegt in der Notwendigkeit großer und qualitativ hochwertiger Datensätze. Oft sind die verfügbaren Daten unvollständig, verrauscht oder nicht für alle möglichen Produktionsszenarien repräsentativ. Zukünftige Arbeiten sollten sich darauf konzentrieren, Methoden zur Generierung synthetischer Daten weiterzuentwickeln und zu optimieren, um diese Lücken zu schließen.

Die Integration von KI und Quantencomputing in bestehende Produktionssysteme ist eine weitere Herausforderung. Die meisten aktuellen Produktionssysteme sind nicht darauf ausgelegt, die hohen Anforderungen dieser Technologien zu erfüllen, was zu erheblichen Anpassungskosten führen kann. Außerdem muss die Skalierbarkeit der Lösungen gewährleistet sein, um in großem Maßstab implementiert werden zu können. Zukünftige Arbeiten sollten daher auch die Entwicklung von standardisierten Schnitt-

stellen und Protokollen untersuchen, um die Integration dieser Technologien in bestehende Systeme zu erleichtern.

Dieser Beitrag hat die vielfältigen Anwendungsmöglichkeiten von KI und Quantencomputing in der modernen Fertigungsindustrie betrachtet. Veranschaulicht wurde die Rolle dieser Technologien am Beispiel der Blechverarbeitung. Der Einsatz von KI und Quantencomputing hat das Potenzial, die Produktionsprozesse erheblich zu optimieren und die Materialverschwendung zu reduzieren. Von der initialen Blechproduktion über das Schneiden und Schweißen bis hin zum finalen Lackieren der Teile können diese Technologien signifikante Effizienzsteigerungen und Qualitätsverbesserungen bewirken. Besonders hervorzuheben sind Ansätze wie erklärbare KI für transparente Modelle, Quantencomputing zur Materialoptimierung und physikbasierte KI-Modelle für präzise Prozesse. Die Nutzung generativer KI als virtuellen Assistenten und zur Erstellung synthetischer Datensätze trägt zusätzlich zur Verbesserung der Produktionsprozesse bei.

Die Integration dieser Technologien in die Fertigungsprozesse ist ein klarer Beleg für ihre transformative Kraft. Sie ermöglicht Herstellern, den schnell wachsenden und sich ständig ändernden Anforderungen der Industrie nicht nur zu begegnen, sondern diese Erwartungen sogar zu übertreffen. Insgesamt zeigen diese Technologien großes Potenzial, die Fertigungsindustrie grundlegend zu transformieren und zukunftssicher zu gestalten. Um dieses Potenzial vollständig auszuschöpfen, sind jedoch weitere Forschungsarbeiten und technologische Entwicklungen erforderlich. Die kontinuierliche Zusammenarbeit zwischen Industrie und Wissenschaft wird dabei entscheidend sein, um nachhaltige und skalierbare Lösungen zu entwickeln.

## Literatur

- [1] Abdelkafi, N.; Döbel, I.; Drzewiecki, J. et al.: Künstliche Intelligenz (KI) im Unternehmenskontext. Leipzig: Fraunhofer IMW 2019, <https://doi.org/10.24406/publica-fhg-299826>
- [2] KI-Fortschrittszentrum: Homepage. Stand: 2024. Internet: [www.ki-fortschrittszentrum.de/](http://www.ki-fortschrittszentrum.de/). Zugriff am 06.08.2024
- [3] Fraunhofer IPA: xAI-Demonstrator. Stand: 2024. Internet: [730-xai-demo.ipa.fraunhofer.de/](http://730-xai-demo.ipa.fraunhofer.de/). Zugriff am 06.08.2024
- [4] Schaaf, N.; Huber, M.; Maucher, J.: Enhancing decision tree based interpretation of deep neural networks through l1-orthogonal regularization. 18th IEEE International Conference on Machine Learning and Applications (ICMLA), Boca Raton/USA, 2019, pp. 42–49, [doi.org/10.48550/arXiv.1904.05394](https://doi.org/10.48550/arXiv.1904.05394)
- [5] Lundberg, S. M.; Lee, S. I.: A unified approach to interpreting model predictions. Advances in neural information processing systems, NIPS 2017, Long Beach/USA, 2017, [doi.org/10.48550/arXiv.1705.07874](https://doi.org/10.48550/arXiv.1705.07874)
- [6] Shapley, L. S.: 17. A value for n-person games. In: Kuhn, H. W.; Tucker, A. W. (eds.): Contributions to the Theory of Games II. Princeton: Princeton University Press, 1953, pp. 307–318, [doi.org/10.1515/9781400881970-018](https://doi.org/10.1515/9781400881970-018)
- [7] Wachter, S.; Mittelstadt, B.; Russell, C.: Counterfactual explanations without opening the black box: Automated decisions and the GDPR. arXiv 2018, [doi.org/10.48550/arXiv.1711.00399](https://doi.org/10.48550/arXiv.1711.00399)
- [8] Molnar, C.: Interpretable machine learning. Stand: 2024. Internet: [christophm.github.io/interpretable-ml-book/](https://christophm.github.io/interpretable-ml-book/). Zugriff am 06.08.2024
- [9] Bharti, K. et al.: Noisy intermediate-scale quantum algorithms. arXiv 2021, [doi.org/10.48550/arXiv.2101.08448](https://doi.org/10.48550/arXiv.2101.08448)
- [10] Matt, P.-A.; Roth, M.: A heuristic for solving the irregular strip packing problem with quantum optimization. arXiv 2024, [doi.org/10.48550/arXiv.2402.17542](https://doi.org/10.48550/arXiv.2402.17542)
- [11] Schulz, W.; Nießen, M.; Eppelt, U. et al.: Simulation of laser cutting. In: Dowden, J. (ed.): The Theory of Laser Materials Processing: Heat and Mass Transfer in Modern Technology. Dordrecht: Springer Netherlands, 2009, pp. 21–69
- [12] Tatzel, L.: Verbesserungen beim Laserschneiden mit Methoden des maschinellen Lernens. Dissertation, Karlsruher Institut für Technologie (KIT), 2021
- [13] Stahl, J.; Frommknecht, A.; Huber, M.: Comprehensive Quantitative Quality Assessment of Thermal Cut Sheet Edges using Convolutional Neural Networks. Proceedings of the 34th British Machine Vision Conference (BMVC). Stand: 2023. Internet: [papers.bmvc2023.org/0480.pdf](https://papers.bmvc2023.org/0480.pdf). Zugriff am 06.08.2024
- [14] Stavridis, J. et al.: Quality assessment in laser welding: a critical review. International Journal of Advanced Manufacturing Technology 94 (2018), pp. 1825–1847
- [15] Maculotti, G. et al.: Optimisation of laser welding of deep drawing steel for automotive applications by Machine Learning: A comparison of different techniques. Quality and Reliability Engineering International 40.1 (2024), pp. 202–219
- [16] Chryssolouris, G.: Manufacturing systems: theory and practice. New York: Springer Science & Business Media 2013
- [17] Tezuka, T.; Kuroki, M.: An unbiased estimator of the causal effect on the variance based on the back-door criterion in Gaussian linear structural equation models. Journal of Multivariate Analysis 197 (2023) #105201, [doi.org/10.1016/j.jmva.2023.105201](https://doi.org/10.1016/j.jmva.2023.105201)
- [18] Sharma, A.; Kiciman, E.: DoWhy: An end-to-end library for causal inference. arXiv 2020, [doi.org/10.48550/arXiv.2011.04216](https://doi.org/10.48550/arXiv.2011.04216)
- [19] Vaswani, A.; Shazeer, N.; Parmar, N. et al.: Attention is all you need. Advances in Neural Information Processing Systems. arXiv 2017, [doi.org/10.48550/arXiv.1706.03762](https://doi.org/10.48550/arXiv.1706.03762)
- [20] Brown, T. et al.: Language models are few-shot learners. Advances in Neural Information Processing Systems. arXiv 2020, [doi.org/10.48550/arXiv.2005.14165](https://doi.org/10.48550/arXiv.2005.14165)
- [21] Xia, Y.; et al.: Towards Autonomous System: Flexible Modular Production System Enhanced with Large Language Model Agents. 28th International Conference on Emerging Technologies and Factory Automation ETFA Padova/Italy, 2023. ArXiv 2023, [doi.org/10.48550/arXiv.2304.14721](https://doi.org/10.48550/arXiv.2304.14721)
- [22] Jiang, Albert Q. et al.: Mistral 7B. arXiv 2023, [doi.org/10.48550/arXiv.2310.06825](https://doi.org/10.48550/arXiv.2310.06825)
- [23] Touvron, H.; Martin, L.; Stone, K. et al.: Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv 2023, [doi.org/10.48550/arXiv.2307.09288](https://doi.org/10.48550/arXiv.2307.09288)
- [24] Gao, Y. et al.: Retrieval-augmented generation for large language models: A survey. arXiv 2023, [doi.org/10.48550/arXiv.2312.10997](https://doi.org/10.48550/arXiv.2312.10997)
- [25] Chen, J. et al.: Benchmarking large language models in retrieval-augmented generation. arXiv 2023, [doi.org/10.48550/arXiv.2309.01431](https://doi.org/10.48550/arXiv.2309.01431)
- [26] De Mitri, O.; Frommknecht, A.; Huber, M. et al.: Synthetic data generation for improvement of machine learning-based optical quality control: a practical approach in the welding context. In: Multimodal Sensing and Artificial Intelligence: Technologies and Applications III. Proceedings SPIE 12621 (2023), pp. 237–245, [doi.org/10.1117/12.2682047](https://doi.org/10.1117/12.2682047)



**Dr. techn. Dipl.-Ing. Christof Nitsche**

*christof.nitsche@ipa.fraunhofer.de*

*Tel. +49 711 / 970-1212*

*Foto: Fraunhofer IPA*


**Xinyang Wu, M.Sc.** 

**Danilo Brajovic, M.Sc.**

**Lena Lörcher, M.Sc.** 

**Dr. rer. nat. Marco Roth** 

**Philipp Wagner, M.Sc.** 

**Alper Yaman, Ph.D** 


**Jannik Schwab, M.Sc.**

**Helena Monke, M.Sc.**

**Christoph Hennebold, M.Sc**

**Prof. Dr.-Ing. Marco F. Huber** 

Fraunhofer-Institut für Produktionstechnik

und Automatisierung IPA 

Nobelstr. 12, 70569 Stuttgart

[www.ipa.fraunhofer.de](http://www.ipa.fraunhofer.de)

## LIZENZ



Dieser Fachaufsatz steht unter der Lizenz Creative Commons  
Namensnennung 4.0 International (CC BY 4.0)