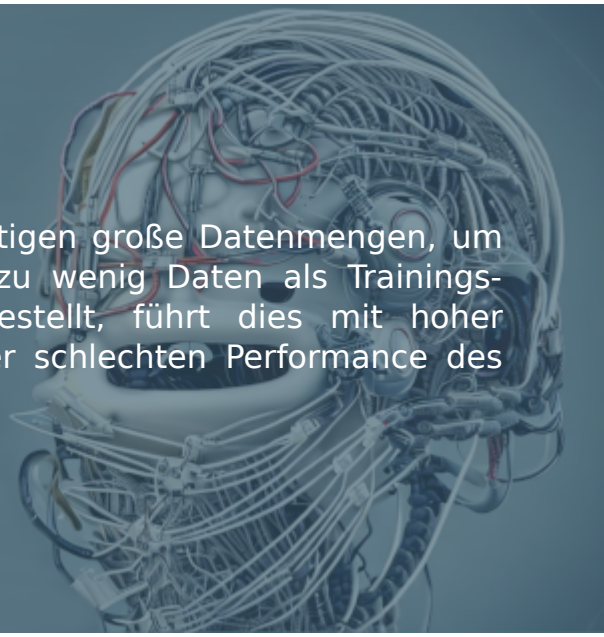
# SYNTHETIC DATASETS

ai in production

# AI FACTS

## DATA DRIVEN METHODS

"KI-basierte Systeme benötigen große Datenmengen, um effektiv zu sein. Werden zu wenig Daten als Trainings-objekte zur Verfügung gestellt, führt dies mit hoher Wahrscheinlichkeit zu einer schlechten Performance des KI-basierten Systems."



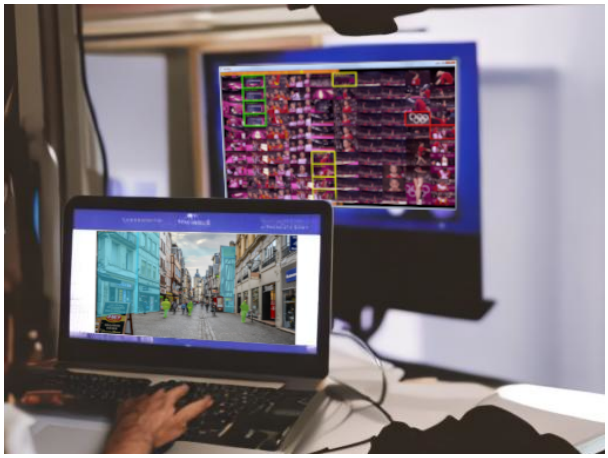
Künstliche Intelligenz (KI) basierte Fehlererkennung und -klassifizierung haben die Möglichkeiten für Unternehmen, Fehler zu identifizieren, zu klassifizieren und zu managen, revolutioniert. KI-Algorithmen sind in der Lage, Anomalien in Daten zu erkennen, die andernfalls von Menschen unbemerkt bleiben würden, sodass Unternehmen Probleme schneller ange-hen können, bevor diese sich zu kostspieligen Fehlern entwickeln.

KI-basierte Systeme - bzw. präziser eingegrenzt Deep-Learning-Systeme - benötigen jedoch große Datenmengen, um effektiv zu sein. Denn die Deep-Learning-Modelle bestehen aus einer Vielzahl von Parameterebenen, die zum Trainieren enorme Datenmengen benötigen. Werden diesen Deep-Learning-Modellen zu wenig Daten als Trainingsobjekte zur Verfügung gestellt, führt dies mit hoher Wahrscheinlichkeit zu einer schlechten Performance des KI-basierten Systems.

Große Mengen korrekt gelabelter Daten sind nicht immer verfügbar oder leicht zu beschaffen. Oftmals eignen sich frei verfügbare Datensätze nur für allgemeine Anwendungen, während die benötigte Anwendung eine zu große Nische darstellt. Entscheidet man sich dennoch für den Einsatz von KI in einem solchen speziellen Fall, benötigt man früher oder später eine große Menge an gelabelten Daten.

Die Generierung von gelabelten Daten kann über automatisierte Methoden erfolgen, z. B. mit Hilfe eines spezialisierten KI-Systems, oder durch manuelle Aufnahme von Bildern und Verwendung von Software zur Annotation. Beide Methoden zur Erzeugung von gelabelten Datenpunkten für einen ausreichend großen Datensatz erweisen sich jedoch als sehr zeitaufwändig, wenn es sich um den Einstieg in die KI handelt. Während die erste Methode ein hohes Maß an KI-Kenntnissen und das Verständnis für die Feinabstimmung eines Modells erfordert, um brauchbare Ergebnisse zu erzielen, erfordert die zweite Methode einen enormen Arbeitsaufwand für die Aufnahme von Bildern und deren manuelle Annotation.

Wenn der ursprüngliche Trainingsdatensatz, der für das Training der KI verwendet wird, von Menschen manuell gelabelt wurde, können außerdem noch Fehler vorhanden sein, die später zu falschen Ergebnissen des KI-basierten



Manuelle Datenannotation

Systems führen können. Zudem erfordern von Menschen gelabelte Datensätze spezielles Wissen über den jeweiligen Bereich, damit diese korrekt gelabelt werden können.

Das Abwägen der Aspekte manueller Datenerfassung und -annotation führt oftmals dazu, dass synthetische, automatisch generierte Datensätze zu einer attraktiven Option werden, wenn es um Kosteneffizienz und Genauig-keitssteigerung für verschiedenste Anwendungsfälle geht.

Unser automatisierter Datengenerierungsservice für KI-basierte Fehlererkennungssysteme bietet einen einzigartigen Vorteil gegenüber herkömmlichen Methoden der Datenerfassung und -generierung.

Unser System zur Generierung von synthetischen Datensätzen ist so konzipiert, dass die resultierenden Daten exakt auf die spezifischen Anforderungen Ihres Projekts zugeschnitten sind, um möglichst genaue und zuverlässige Ergebnisse zu gewährleisten.

Darüber hinaus können unsere synthetischen Datensätze schnell und effizient mit minimalen manuellen Eingaben von Ihnen oder Ihrem Team erstellt werden, was Ihnen wertvolle Zeit und

"Synthetisch generierte Datensätze bieten verschiedene Vorteile im Vergleich zu traditionell erstellten Datensätzen"

Die automatische Generierung von Trainingsdaten und das automatisierte Labelling bieten eine hohe Skalierbarkeit sowie eine verbesserte Effizienz (da keine zusätzlichen Arbeitskräfte eingestellt, bezahlt und an einem Arbeitsplatz eingesetzt werden müssen) und eine Geheimhaltung der Daten, die bei Labelling mittels menschlich basierten Dienstleistungen einfach nicht erzielt werden kann.

Stellen Sie sich vor, dass Ihre hochsensiblen Produktionsdaten von einem Labeldienst an Subunternehmer ausgelagert werden, die keinerlei Kontrolle über die Geheimhaltung Ihrer innovativen Produktlinie haben.

Auch aus statistischer Sicht können manuell Ressourcen spart und dennoch hochwertige Datensätze liefert.

Das bedeutet, dass nur ein Minimum an Eingaben erforderlich ist, wie z. B. einige Beispielbilder zum Abgleich von Kamera- und Beleuchtungsparametern.

Darüber hinaus gewährleistet unser automatisiertes Verfahren zur Generierung synthetischer Datensätze jedes Mal genaue und konsistente Ergebnisse, so dass unsere Kunden sicher sein können, zuverlässige Ergebnisse mit ihren KI Modellen zu erzielen.

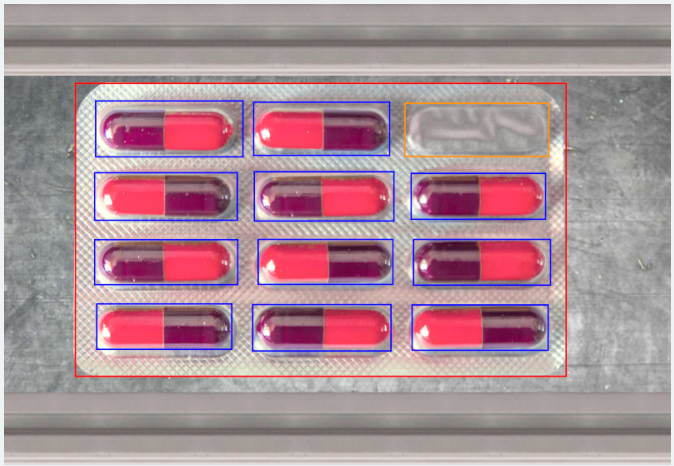
Im folgenden Abschnitt präsentieren wir Ihnen im Rahmen einer kurzen Fallstudie ein KI-Projekt eines Kundens aus der Pharmaindustrie.

# CASE STUDY

## ONLINE MONITORING

Kommen Sie mit auf die Reise eines Hidden Champions in der Pharmaindustrie, der seine Marktposition durch die Einführung von KI-basierten, in Echtzeit durchgeführten, Inline-Qualitätskontrollen stärkt.
Wir haben unseren Kunden von der ersten Idee und Problemstellung über Testläufe bis hin zur Einführung eines Live-Produktionssystems begleitet.

Falsch verpackte Artikel, die von der weiteren Verpackung und dem Versand ausgeschlossen werden sollten, waren das Hauptproblem, als unser Kunde das erste Mal an uns herangetreten ist.
In einem ersten Schritt wurde die aktuelle Produktions- und Qualitätssicherungslösung des Kundens evaluiert. Basierend auf unserer Expertise in der automatisierten Produktion sowie in KI-basierten Systemen erkannten wir die Fallstricke der derzeitigen mechanischen Qualitätskontrollmethode in der Produktionslinie unseres Kunden.
Nach Abwägung aller Gegebenheiten schlugen wir den Wechsel zu einer optischen Kontrollmethode und einem KI-basierten System zur Fehlererkennung und -klassifizierung vor.



KI-basierte Fehlererkennung und Klassifikation

Eine große Herausforderung bei der Zusammenarbeit mit Kunden in wettbewerbsintensiven und hoch-innovativen Branchen ist die Wahrung der Vertraulichkeit von Daten und Produktionsprozessen. Dies ist insbesondere relevant, wenn es um das Labeln von Daten, die von KI-Systemen verwendet werden sollen, in hochsensiblen Bereichen, wie in diesem Projekt, geht.
Um dieses Problem anzugehen, entschied unser Kunde frühzeitig, dass die Auslagerung der manuellen Kennzeichnung von Daten an Unternehmen mit Subunternehmerstrukturen keine praktikable Option ist, da menschliche Fehler möglich sind und Datenlecks durch schwer zu kontrollierende Subunternehmer, die Zugang zu sensiblen Daten haben, zu befürchten sind.
Zudem standen nicht genügend personelle Ressourcen für das manuelle Erfassen und Labeln der Trainingsdaten eines solchen KI-Systems zur Verfügung.
Um diese Probleme zu lösen, entschied sich der Kunde, uns, die Brengtech GmbH, mit der Datensatzerstellung und dem Training der KI-Modelle zu beauftragen. Die Brengtech GmbH hält sich an strenge Datenrichtlinien, wie z.B. die ausschließliche Übermittlung von Datensätzen über eigene, verschlüsselte und in Deutschland gehostete Cloud-Dienste,
Um die synthetischen Datensätze für unseren

Kunden automatisiert zu generieren, haben wir proprietäre Algorithmen entwickelt, die speziell für die Generierung und Annotation großer Datenmengen ausgelegt sind. Im Rahmen dieses Projekts haben wir unserem Kunden auch die Infrastruktur (Rechenleistung) zur Verfügung gestellt, um den Datensatz zu generieren und KI-Modelltests durchzuführen. Auf diese Weise hat der Kunde den Vorteil, alles aus einer Hand und mit einer Geheimhaltungsvereinbarung zu erhalten.

Um den Aufwand für die Erstellung des Datensatzes abzuschätzen, wurden lediglich ein Dutzend Beispielbilder inklusive einer kurzen Beschreibung, welche Fehler in den Produkten erkannt und klassifiziert werden sollten, vom Kunden benötigt. Daraufhin konnten wir einen ersten Satz synthetischer Daten generieren. Nach Rücksprache mit unserem Kunden und Optimierung der resultierenden Daten sind wir in die Produktion übergegangen.

Qualitätskontrollsystems eliminiert.

Nach der erfolgreichen Implementierung erweitert unser Kunde die KI-basierten QS-Systeme auf andere Produkte und Produktionsprozesse. Durch unseren skalierbaren Service können wir neue Datensätze leicht an geänderte Produkte anpassen, was zu deutlich geringeren Kosten für nachfolgende Datensätze führt. Während bei manueller Annotation eine geringfügige Änderung der Daten eine Wiederholung der gesamten Arbeit bedeuten würde, können wir mit unserer auf Algorithmen basierenden Datengenerierungslösung sofort und zu akzeptablen Kosten auf veränderte Kundenanforderungen reagieren.

Der Erfolg mit diesem KI-basierten Projekt erwies sich als Aufbruch in ein neues digitales Zeitalter für unseren Kunden. Jetzt werden weitere Projekte auf mögliche Integrationen KI-basierter Lösungen geprüft. Unzuverlässige, veraltete Bildverarbeitungssysteme, die auf manuellen Heuristiken beruhen, werden dabei durch KI-basierte Systeme ersetzt.

"Innerhalb kurzer Zeit erhielt unser Kunde einen Datensatz und ein KI-Modell, das in der Produktion eingesetzt werden konnte"

Innerhalb kurzer Zeit erhielt unser Kunde einen umfangreichen Datensatz sowie ein trainiertes KI-Modell, das dieser in der Produktion einsetzen konnte. Für unseren Kunden erwiesen sich der synthetisch generierte Datensatz und das KI-Modell als äußerst erfolgreich. Daher wurde das trainierte KI-Modell in das Inline-QS-System integriert. Durch die volle Erfolgsquote des KI-basierten System kann der entsprechende Verpackungsprozess die strengen Standards der Pharmaindustrie einhalten.

Darüber hinaus konnten die bisherigen Ausfallzeiten eliminiert werden. Durch das KI-basierte System werden die engen Service-Intervalle der Wartung des alten, mechanisch basierten

Wenn Sie erfahren möchten, wie Sie von KI in Ihren Produktionsprozessen profitieren können und wie wir Ihnen helfen können, KI schnell in Ihrer Produktion zu etablieren, kontaktieren Sie uns via info@brengtech.de

Copyright © 2023 Brengtech GmbH