



# Das KI-Bereitschaftsmodell

## So evaluieren Sie, ob ein Unternehmen mit künstlicher Intelligenz geschäftlichen Nutzen generieren kann

### Inhaltsverzeichnis

Wo befinden Sie sich auf dem Weg zur KI .....	1
Grundsätzliche Bereitschaft .....	2
Einsatzbereitschaft .....	3
Transformationsbereitschaft.....	4
Fazit .....	4

Intel arbeitet mit zahlreichen Organisationen zusammen, die sich mit Anwendungslösungen für künstliche Intelligenz (KI) befassen. Bilderkennung (Image Recognition), natürliche Sprachverarbeitung (Natural Language Processing, NLP) und vorausschauende Instandhaltung (Predictive Maintenance) entwickeln sich zu besonderen Schwerpunkten. Einige Unternehmen beschäftigen sich zum ersten Mal mit diesen KI-Anwendungsfällen, andere prüfen, wie sie sie von einem erfolgreichen Ausgangspunkt aus weiterentwickeln können.

Um Unternehmen auf ihrem Weg zur erfolgreichen KI-Anwendung zu unterstützen, hat Intel ein Bereitschaftsmodell entwickelt. Dies hilft Entscheidungsträgern zu verstehen, welchen Aktivitäten sie priorisieren sollten. Wir haben dieses Modell auf Grundlage unserer Erfahrung in der Zusammenarbeit mit Kunden aus verschiedenen Industriezweigen und im Rahmen unterschiedlicher Szenarien entwickelt. Die Beispiele umfassen Hersteller, die ihre Qualitätskontrolle verbessern möchten, und Finanzdienstleister, die KI im algorithmischen Handel einsetzen wollen. Dieses Whitepaper bietet eine Orientierungshilfe bei der Evaluierung der Fähigkeit und Bereitschaft eines Unternehmens, KI zur Generierung von Geschäftswerten zu nutzen. Das Paper enthält zudem einen Fragenkatalog, den Sie als Leitfaden zur Selbsteinschätzung verwenden können.

Zunächst lassen sich Unternehmen in drei Kategorien einteilen, abhängig davon, wo sie sich aktuell auf ihrer KI-Reise befinden - ob sie KI also zum ersten Mal verwenden, sie vergrößern oder ausweiten wollen.

### Wo befinden Sie sich auf dem Weg zur KI

#### Für einige Unternehmen ist KI etwas Neues

Viele Unternehmen sind noch nicht vertraut mit dem Einsatz von KI-Anwendungen wie Bilderkennung, natürliche Sprachverarbeitung (NLP) und vorausschauende Instandhaltung (Predictive Maintenance). Wir sehen drei häufige Szenarien:

- **Ein Unternehmen mit vorhandenen Datenbeständen kann vom Einsatz einer KI profitieren.** So hat beispielsweise ein Solarpark, mit dem Intel regelmäßig zusammengearbeitet hat, Bilder von seinen Anlagen aufgenommen, um so potenzielle Schäden aufzuspüren. In diesem Szenario wollte das Unternehmen verstehen, ob Bilderkennungsalgorithmen verwendet werden können, um Fehler automatisch zu erkennen. Ähnliche Gegebenheiten finden sich im Gesundheitswesen bei der Auswertung von Röntgenbildern oder anderen medizinischen Daten.
- **Ein Unternehmen arbeitet in einer traditionellen Umgebung und möchte KI oder maschinelles Lernen einsetzen, um Optimierungsmöglichkeiten zu erkunden.** Dies können zum Beispiel Unternehmen sein, die bereits mit Analyseverfahren vertraut sind, oder Forschungseinrichtungen, die physikalische Algorithmen einsetzen. Ebenso kann maschinelles Lernen für Hersteller von Nutzen sein, die festgelegte Verfahren zur Fehlererkennung bei Maschinen, Materialien, Gütern oder Prozessen verwenden.
- **Einige Unternehmen haben das Potenzial von KI bereits erforscht, so beispielsweise den Einsatz von NLP im Kundendienst oder die Bilderkennung zur Optimierung von Verhaltensanalysen im Einzelhandel.** In einigen Fällen sind Lösungen, die KI bieten kann,

nicht auf den ersten Blick ersichtlich. Dazu gehören zum Beispiel interne Vorhersagemodelle für die Lagerverwaltung.

Für diese Unternehmen ist es schwierig, die Vorteile von KI vorherzusagen. Zudem sind Daten nicht immer vorhanden oder in einer für die KI nutzbaren Form verfügbar.

#### **Andere Unternehmen sind hingegen bereits so weit, den Einsatz von KI auszuweiten**

Diese Unternehmen haben die Möglichkeiten der KI schon erforscht, besitzen eine Vorstellung davon, wie sie sie einsetzen wollen, und haben bereits erfolgreich Testmodelle implementiert. Hier sind eine Reihe von Szenarien denkbar:

- **Ein Unternehmen hat eine Proof-of-Concept-KI-Lösung entwickelt und auf einer Workstation oder einem Einzelgerät installiert.** Die Herausforderung besteht nun darin, diese Lösung in das Rechenzentrum zu übertragen, damit sie für die Produktion genutzt werden kann. Ein Beispiel dafür wäre, in der Produktion mit Erkenntnissen aus der vorausschauenden Instandhaltung zu arbeiten.
- **Ein Unternehmen hat eine eigene Lösung entwickelt und will jetzt eine Infrastruktur und/oder Software nach Industriestandard nutzen.** Ob Hardware, Software oder beides: Die Herausforderung kann darin bestehen, auf eine Architektur zu migrieren, die (kurzfristig) weniger optimale Ergebnisse liefert.

Obwohl diese Unternehmen auf ihrem Weg zur KI weiter fortgeschritten sind, fehlen ihnen oft noch die Fähigkeiten, die Nutzung der KI über eine kleinere Anzahl von Ingenieuren und eine relativ einfache Hardware- und Softwarekonfiguration hinaus zu erweitern. Hinzu kommt: Die Umstellung auf eine Multi-Node-Lösung muss die Tatsache berücksichtigen, dass KI nicht linear skaliert. Während eine Single-Node-Konfiguration beispielsweise Hunderte von Bildern pro Sekunde verarbeiten kann, liefert ein Wechsel auf eine Multi-Node-Lösung mit 50 Prozessoren nicht die 50-fache Leistung. Datenquellen können auch zu Engpässen führen, wenn sie nicht mehr für einmalige Stichproben, sondern in großem Umfang verwendet werden müssen.

#### **Andere Unternehmen implementieren KI in großem Maßstab**

Eine dritte Kategorie von Unternehmen setzt bereits maschinelles Lernen oder KI ein und versucht nun, sie auf ein breiteres Spektrum von Anwendungsfällen auszuweiten. Wir sehen hier folgende Szenarien:

- **Ein Unternehmen setzt KI erfolgreich in einem bestimmten Geschäftszweig ein und möchte nun weitere Anwendungsbereiche erkunden.** Ein Unternehmen nutzt beispielsweise die Bilderkennung zur Qualitätskontrolle in der Produktion und möchte nun NLP in Callcentern einsetzen.
- **Ein Unternehmen nutzt KI erfolgreich, um aus Daten zu lernen und sie zu interpretieren, und möchte sie nun auf die Inferenzbasierte Instandhaltung und Aktualisierung von Modellen ausweiten.** Es könnte die Ergebnisse der KI auch nutzen wollen, um die Automatisierung voranzutreiben, z. B. durch die Verwendung von Bestandsdaten für die Verwaltung, Planung und Beschaffung von Ersatzteilen. Aus infrastruktureller Sicht könnte dasselbe Unternehmen auch danach streben, die Energieeffizienz oder

Leistung zu verbessern, um so die Gesamtbetriebskosten (Total Cost of Ownership, TCO) zu senken.

In diesen Szenarien besteht die Herausforderung darin, sicherzustellen, dass die Plattform für mehr als einen Anwendungsfall geeignet ist, und Ressourcen zu verwalten, da die Anforderungen und der Ressourcenbedarf zwischen den einzelnen Fällen schwanken.

## **Die drei Formen von KI-Bereitschaft**

Während sich Unternehmen in unterschiedlichen Stadien ihrer KI-Reise befinden können, hängt der benötigte Fortschritt für die nächste Entwicklungsstufe oder für den anhaltenden Erfolg davon ab, dass die richtigen Elemente in Bezug auf Fähigkeiten und Ressourcen, Infrastruktur und Technologie sowie Prozesse und Modelle vorhanden sind.

Wir können Bereitschaft auch anhand der folgenden Bereiche betrachten:

- **Grundsätzliche Bereitschaft** - Voraussetzung für KI sind geeignete Infrastrukturen und Schnittstellen
- **Einsatzbereitschaft** - Geeignete Management- und Governance-Mechanismen sind der Schlüssel zur Nachhaltigkeit von KI-Lösungen
- **Transformationsbereitschaft** - Die Fähigkeit eines Unternehmens, den Wert der KI zu maximieren

Grundsätzliche Bereitschaft ist der erste Schritt. Aber der Erfolg von KI hängt von der Einsatzbereitschaft und dann von der Empfänglichkeit des Unternehmens für KI ab - das bedeutet, von seiner Transformationsbereitschaft. Dies führt zu neuen und aktualisierten Anforderungen, die die Grundlage für KI-Implementierungen weiter vorantreiben.

## **Grundsätzliche Bereitschaft**

Voraussetzungen für KI sind geeignete Infrastrukturen und Schnittstellen. Dies liegt zwar auf der Hand, aber es ist nicht immer ersichtlich, welcher Bedarf besteht, bevor potenzielle Szenarien getestet und bewertet werden. Gleichzeitig kann ein Mangel an Fähigkeiten und Fachwissen bestehen. In Anbetracht dieser Faktoren ist Folgendes zu berücksichtigen.

### **Infrastruktur-Plattform**

Viele Unternehmen möchten wissen, ob bestehende Rechenzentrumseinrichtungen für KI-Workloads geeignet sind. Während die Antwort für ein Proof of Concept (PoC) positiv sein kann, sind einige Einrichtungen möglicherweise nicht für die massiv skalierbare Verarbeitung geeignet, die für maschinelles Lernen und KI erforderlich ist. Je nach Szenario kann der von der KI geforderte Datenfluss das Netzwerk stark belasten. Einige KI-Lösungen können außerhalb der Geschäftszeiten geplant werden, um den Durchsatz zu maximieren. In der Zwischenzeit können andere, zeitkritischere Szenarien das Netzwerk stärker belasten, z. B. wenn die Daten im Rahmen von vorausschauender Instandhaltung genutzt wurden.

Für weitere Informationen über Plattformanforderungen für KI, so etwa zur Verarbeitung, Speicherung und Vernetzung, lesen Sie bitte unser [Paper über die Erstellung eines PoCs](#).

### Cloud-Ressourcen

Betrachten Sie Cloud-basierte Dienste als Grundlage für KI, insbesondere in Bereichen wie Bild- und Sprachverarbeitung. Die Cloud bietet den Vorteil kostengünstiger Pay-per-Use-Dienste, was sie zu einer ausgezeichneten Wahl für Schulungen und Tests macht. Wenn Unternehmen die Nutzung von KI ausbauen wollen, müssen sie allerdings prüfen, ob ihre Cloud-basierten Ressourcen noch geeignet sind. Anbieter von Cloud-Diensten können Staffelpreise und Service-Level-Agreements haben, die möglicherweise durch die zugrundeliegende Computing-Infrastruktur definiert sind, auf der der Workload läuft. Entscheider müssen ein optimales Verhältnis zwischen Performance und Kosten finden - möglicherweise auf Basis von Workloads.

### Datenquellen

Wichtige Datenquellen müssen verfügbar und zugänglich sein. Entscheider und Ingenieure benötigen die Gewissheit, dass Daten in der richtigen Quantität und Qualität für Deep-Learning-Algorithmen zur Verfügung stehen - unabhängig davon, ob sie intern oder extern bezogen werden. Im letzteren Fall müssen möglicherweise vertragliche Vereinbarungen mit Dritten getroffen werden.

### Softwarepakete

Heute gibt es zahlreiche Softwarepakete für Inferenz, maschinelles Lernen und KI, jedes mit bestimmten Vorteilen und Einschränkungen. Neben Open-Source-Paketen wie TensorFlow\* und BigDL existieren kommerzielle Lösungen wie die [Intel® Saffron™ Software](#) und Cloud-basierte Services. Jedes dieser Pakete muss in Tools für Datenmanagement, Visualisierung usw. integriert werden. Und jeder Anwendungsfall erfordert eine Softwarearchitektur, die die besten Tools für den jeweiligen Auftrag auswählt und dabei möglicherweise nachgelagerte Systeme, Anpassungen, Optimierungen und andere Modifikationen berücksichtigt.

### Einsatzbereitschaft

Geeignete Management- und Governance-Mechanismen sind der Schlüssel zur Nachhaltigkeit von KI-Lösungen. IT-Entscheider können mehrere Bereiche untersuchen, um sicherzustellen, dass ihre KI-Implementierungen betriebsbereit sind. Wir empfehlen eine Überprüfung der folgenden Punkte.

#### Schnelle Ergebnisse

Bewährte Modelle wie DevOps und andere agile Methoden können Unternehmen in jedem Schritt auf ihrem Weg zur KI erheblich unterstützen. In den frühen Phasen reagiert die kontinuierliche Entwicklung und Bereitstellung auf schnell wechselnde Anforderungen, unklare Ergebnisse und die Notwendigkeit einer wiederholten Überprüfung, Bewertung und Prüfung. Und für den fortgeschrittenen Einsatz kann die kontinuierliche Verbesserung Änderungen sowohl in den Datenquellen als auch bei den Modellen bewirken.

#### Operatives Management

Die operative Seite von DevOps muss auch Kriterien für eine effektive Servicebereitstellung (über die als Grundlage für die KI gewählte Plattform hinweg) sowie ein effektives Management aller internen und

## KI-Software-Optimierung für Intel® Xeon® Prozessoren

Damit Data Scientists und Entwickler mit ihrem bevorzugten Framework arbeiten können, hat Intel eine Reihe von Deep-Learning-Bibliotheken für viele der beliebtesten KI-Frameworks optimiert, darunter TensorFlow\*, Theano\* und MXNet\*.

Unter diesen Frameworks arbeitet die [Intel® Math Kernel Library for Deep Neural Networks](#) (Intel® MKL-DNN) als neuer Beschleuniger mit spezifischer Mathematik für Deep Learning und optimiert für x86 mit Intel® Advanced Vector Extensions 2 (Intel® AVX-2) und Intel® Advanced Vector Extensions 512 (Intel® AVX-512). Als Open-Source-Projekt wird es auch künftig neue und sich entwickelnde Trends für alle wichtigen Frameworks verfolgen.

Alternativ [kann BigDL als Open-Source verteilte Deep Learning Library](#) für Spark\* direkt auf bestehenden Spark oder Apache Hadoop\* Clustern laufen. Von Intel entwickelt, ermöglicht sie das Laden vortrainierter Torch\*-Modelle in das Spark-Framework und kann effizient skaliert werden, um Datenanalysen bei großen Datenmengen durchzuführen.

Erfahren Sie mehr über diese beiden Lösungen in unserer Infografik über die [Anatomie eines Proof of Concept für KI](#).

externen Datenquellen einbeziehen. Zusätzlich sollten Kriterien zur Messung der Geschäftseffektivität der KI existieren - zum Beispiel, ob die generierten Erkenntnisse oder die Automatisierung den erwarteten Geschäftswert liefern.

#### Fähigkeiten und Fachwissen

Fehlende Fähigkeiten sind für Unternehmen ein häufiges Problem in den frühen Phasen ihrer KI-Reise. Am Anfang müssen Sie möglicherweise auf externe Kompetenzen setzen, insbesondere auf die von Lösungsarchitekten, die PoC-Lösungen durchführen können. Es ist jedoch sinnvoll, interne Kompetenzen sowohl in der IT als auch in den Geschäftsbereichen aufzubauen. Das Vorgehen über einen PoC hinaus ohne eigene Kompetenzen birgt unnötige Risiken für ein KI-Projekt, insbesondere in der Überprüfungs- und Evaluierungsphase.

#### Cybersicherheit

Angesichts der Tatsache, dass die Beziehung zwischen KI und Automatisierung menschliche Eingriffe und damit die Überwachung minimieren kann, sollte der Cybersicherheit von Daten, Infrastruktur und Algorithmen hohe Priorität eingeräumt werden. Mögliche Sicherheitsrisiken ergeben sich aus der Verfälschung der Eingangsdaten in die KI, der Manipulation der Modelle oder dem unbefugten Zugriff auf die daraus resultierenden Erkenntnisse.

#### Governance, Compliance und Risiko

Die Wechselbeziehungen zwischen KI und Governance sind komplex und vielschichtig. In den frühen Unternehmensphasen unterscheiden sich die Governance-Fragen nicht von anderen datenzentrischen IT-Projekten: Kann das Projekt Ergebnisse liefern? Ist der Schutz der Kundendaten gewährleistet? ... Mit zunehmender Nutzung der KI

## Von Daten und Erkenntnissen bis hin zu Innovationen und Umsätzen: skalierbare Intel® Xeon® Prozessoren

Intel® Xeon® Prozessoren sind das Herzstück des Rechenzentrums und führen die meisten der kritischsten und innovativsten Workloads aus. Während die bestehende Intel®-Architektur in Ihrem Rechenzentrum Ihnen den Einstieg in maschinelles Lernen und Deep Learning ermöglicht, bieten die neuen skalierbaren Intel® Xeon® Prozessoren die agilste Intel-Plattform, die Ihre KI weiter voranbringen kann. Hinsichtlich der Rechenleistung, die für Deep Learning und Inferenz erforderlich ist, bietet der skalierbare Intel® Xeon® Prozessor eine bis zu 2,2-fache Leistung im Vergleich zur vorherigen Generation für Deep-Learning-Training und Inferenzleistung. Und durch zusätzliche Software-Optimierungen - zum Beispiel TensorFlow\*, Theano\* und Torch\* - kann eine bis zu 113-fache Leistung im Vergleich zu nicht optimierten, drei Jahre alten Servern für Deep Learning erreicht werden, was eine solide Basisarchitektur für KI-Workloads bietet<sup>1</sup>. Sechs Speicherkanäle im Vergleich zu den vier der Vorgängergenerationen sorgen für eine deutlich höhere Speicherbandbreite und Kapazität für speicherintensive Workloads.

ergeben sich zusätzliche Auswirkungen: Wie viel Personaleinsatz ist beispielsweise bei der vorausschauenden Planung und Instandhaltung in der Ersatzteilbeschaffung (falls vorhanden) erforderlich?

### Transformationsbereitschaft

Der dritte (und möglicherweise wichtigste) Aspekt der Bereitschaft ist die Frage, ob ein Unternehmen den Wert maximieren kann, den es von der KI erhält. Unabhängig davon, ob die KI zur Mustererkennung, Erkenntnisgewinnung oder Prozessautomatisierung eingesetzt wird, hat sie in der Regel eine der folgenden Auswirkungen:

- Unterstützen Sie Entscheidungsfindungen auf Geschäftsführungs- und nachfolgenden Ebenen
- Automatisieren Sie einen Teil der Geschäftsprozesse oder fördern Sie automatisierte Rückmeldungen

In beiden Fällen kann KI einen erheblichen Einfluss auf das Tagesgeschäft haben - was bedeutet, dass das Unternehmen in der Lage sein muss, sich den daraus resultierenden Veränderungen anzupassen.

### Strategische Führung

Der Vorstand eines Unternehmens sieht digitale Technologien im Allgemeinen und KI im Besonderen nicht unbedingt als Motor für das Geschäftswachstum. Allerdings ist es hilfreich, wenn die richtigen Denkweisen vorhanden sind, um die Umsetzung von der Unternehmensspitze aus voranzutreiben. Wenn KI als Quelle eines Strategievorteils gesehen wird, werden Prioritäten und Budgets entsprechend festgelegt.

### Geschäftschancen

In diesem Sinne können auch die Erfolgsaussichten von KI verbessert werden, wenn sie neue Möglichkeiten für Geschäftswachstum, neue Wege der Kundenansprache oder neue Arten von Geschäftsprozessen eröffnet. Wenn dies der Fall ist, sollte das Unternehmen sich so strukturieren, dass es die möglicherweise vorhandenen, von der KI gesteuerten Geschäftsmöglichkeiten nutzen kann. Es ist wichtig, sich im Klaren darüber zu sein, wie bestehende und erwünschte Betriebsmodelle die Ergebnisse und Vorteile der KI durch Automatisierung oder Erweiterung integrieren können.

### Klarheit des Geschäftsmodells

Ein Wegbereiter für jede geschäftliche Veränderung ist das Vorhandensein einer klaren Vorstellung von den Vorteilen, die der Wandel mit sich bringt. Die Erfahrung von Intel zeigt, dass sich Unternehmen bei ihren ersten Schritten im KI-Bereich verstärkt auf die TCO fokussieren - sich als auf die Frage konzentrieren, ob die KI zu akzeptablen Kosten die erwarteten Ergebnisse liefert (und durch Automatisierung Geld spart). Weiter entwickelte Projekte zielen darauf ab, die KI-Leistung zu steigern - und die fortgeschrittensten wollen den ROI betriebswirtschaftlich erfassen, z. B. in Form von Zeit, die für andere Aufgaben zur Verfügung steht. Der Businessplan muss klare, kalkulierte Kriterien für den Erfolg aufzeigen.

### Akzeptanz im Unternehmen

Die Lösung sollte an die Bedürfnisse des Unternehmens bis hin zu den täglichen Aktivitäten der Mitarbeiter vor Ort angepasst werden. Akzeptanz zu erreichen, wird nicht immer einfach sein, insbesondere wenn sich die Rollen und Verantwortlichkeiten durch die Einführung der KI ändern.

### Fazit: Beginnen Sie jetzt mit der Planung Ihres KI-Erfolges

Unabhängig davon, ob Ihr Unternehmen am Anfang seiner KI-Reise steht oder bereits einige Fortschritte erzielt hat: Sie können sich eine Reihe von Fragen stellen, um zu beurteilen, an welchem Punkt Sie sich befinden, und um Ihre Bereitschaft zu steigern. Einige empfohlene Fragen, die Ihnen bei der Planung helfen könnten, sind im Folgenden aufgeführt:

#### Erstmalige Nutzung der KI

- Ist das Szenario, der Anwendungsfall oder das mit KI zu lösende Problem klar definiert?
- Sind die Prioritäten so gesetzt, dass KI den größten geschäftlichen Nutzen bringt?
- Ist die geplante Infrastrukturarchitektur klar und zweckmäßig?
- Sind alle notwendigen Datenquellen klar ersichtlich und zugänglich?
- Können die von Ihnen gewählten Softwarepakete die KI-Lösung durchgängig umsetzen?
- Sind ausreichende Fähigkeiten und Ressourcen vorhanden (intern oder extern)?
- Wurden Erwartungen an die Schulungs- und Lernzeiten gestellt?
- Ist die TCO der End-to-End-Lösung eindeutig und genehmigt?

## Optimale Nutzung der vorhandenen Infrastruktur mit Intel® Xeon® Prozessoren

Die Nutzung Ihrer bestehenden Rechenzentrumsinfrastruktur ist ideal, um den Wert der KI für Ihr Unternehmen auf einer flexiblen, universellen Grundlage nachzuweisen.

Intel® Xeon® Prozessoren nutzen ein konsistentes Infrastruktur- und Programmiermodell für bestehende Analyse-Pipelines und unterstützen den großen Arbeitsspeicherbedarf von KI-Modellen. Durch die bereits erwähnten Software-Optimierungen kann die auf Intel® Xeon® Prozessoren basierte Infrastruktur ein Unternehmen bei seinem Weg zur KI je nach Auslastung weiter unterstützen - vom Experiment über den PoC bis hin zur Produktion.

Erfahren Sie mehr über diese beiden Lösungen in unserer Infografik [über die Anatomie eines Proof of Concept für KI](#).

- Wird die Architektur der KI als Plattform und nicht als Einzellösung realisiert?
- Sind die jeweiligen Unternehmensbereiche umfassend über die Auswirkungen der KI auf ihre Prozesse informiert?
- Werden die Governance-Bedürfnisse der KI-Lösung klar verstanden?
- Wird KI als zentrale Säule einer IT-gestützten Geschäftsstrategie gesehen?

Durch die Beantwortung dieser Fragen erhöhen Sie die Erfolgswahrscheinlichkeit - und die Akzeptanz von KI-basierten Lösungen im Unternehmen. Es ist keine realistische Erwartung, auf Anhieb alles richtig zu machen: Vielmehr sollten Sie darauf achten, Fähigkeiten und Fachwissen aufzubauen, während Sie die Vorteile von KI entdecken.

Am wichtigsten ist ein klares Verständnis für das Problem, das Sie lösen wollen. Erstellen Sie eine Problembeschreibung und arbeiten Sie dann daran, die Lösung umzusetzen. Egal ob es darum geht, Prozesse zu automatisieren oder Einblicke zu gewinnen: Das eigentliche Ziel von KI ist es, Ihnen und Ihrem Unternehmen Freiraum für Innovationen und Wachstum zu geben.

### Ausbau der KI

- Kann die geplante Lösung über die ersten Tests und Evaluierungen hinaus skaliert werden?
- Wurde ein klar definiertes Geschäftsszenario mit dem Geschäftsbereich vereinbart?
- Sind genügend direkte Ressourcen verfügbar, mit zugeteilter und reservierter Zeit?
- Ist die Netzwerkbandbreite ausreichend, um eine zeitnahe Datenbereitstellung zu gewährleisten?
- Gibt es operative Managementprozesse, die die KI-Ergebnisse einbeziehen?
- Entspricht die Architektur den Industriestandards und bewährten Verfahren?
- Wurde eine Risikobewertung der Cybersicherheit vorgenommen und umgesetzt?
- Wurden realistische Umsetzungspläne formuliert und vermittelt?

### Ausweitung der KI-Nutzung

- Gibt es ein Team, das die kontinuierliche Verbesserung der KI überwacht?
- Wurden die erweiterten KI-Möglichkeiten für das Unternehmen untersucht und sind sie bekannt?
- Wurden KI-Lösungen nach bewährten Verfahren entwickelt und eingesetzt?
- Gibt es Maßnahmen zur Überwachung der Geschäftseffektivität der KI-Lösungen?

### Weitere Informationen

Weitere Informationen über das Intel® KI-Portfolio und wie es Sie auf dem Weg zur KI unterstützen kann, finden Sie hier: [ai.intel.com](https://ai.intel.com)

Die leistungsoptimierten Bibliotheken für maschinelles Lernen und Deep Learning sowie die Frameworks von Intel sind hier verfügbar: <https://software.intel.com/en-us/ai-academy>





<sup>1</sup> INFERENCE mit FP32 Batch Size Caffe GoogleNet v1 256 AlexNet 256.

Die Leistungsschätzungen wurden vor der Implementierung der neuesten Software-Patches und Firmware-Updates ermittelt, die als Gegenmaßnahmen für die als „Spectre“ und „Meltdown“ bezeichneten Exploits eingesetzt wurden. Die Implementierung dieser Updates kann dazu führen, dass diese Ergebnisse auf Ihr Gerät oder System nicht zutreffen. In Leistungstests verwendete Software und Workloads können speziell für die Leistungseigenschaften von Intel®-Mikroprozessoren optimiert worden sein. Leistungstests wie SYSmark\* und MobileMark\* werden mit spezifischen Computersystemen, Komponenten, Softwareprogrammen, Operationen und Funktionen durchgeführt. Jede Veränderung bei einem dieser Faktoren kann abweichende Ergebnisse zur Folge haben. Als Unterstützung für eine umfassende Bewertung Ihrer geplanten Anschaffung sollten Sie noch andere Informationen und Leistungstests heranziehen – auch im Hinblick auf die Leistung des betreffenden Produkts in Verbindung mit anderen Produkten. Weiterführende Informationen finden Sie unter <http://www.intel.de/performance>. Quelle: Intel Messergebnisse, Juni 2017. Anmerkung zur Optimierung: Unter Umständen können Intel®-Compiler bei Optimierungen, die nicht für Intel®-Mikroprozessoren spezifisch sind, auch bei Mikroprozessoren anderer Hersteller denselben Optimierungsgrad erzielen. Zu diesen Optimierungen gehören Befehlssätze für SSE2, SSE3 und SSSE3 sowie weitere Optimierungen. Intel übernimmt keine Garantie für die Verfügbarkeit, Funktionalität oder Wirksamkeit von Optimierungen für Mikroprozessoren, die nicht von Intel hergestellt wurden. Vom Mikroprozessor-abhängige Optimierungen in diesem Produkt sind für die Anwendung in Verbindung mit Intel®-Mikroprozessoren bestimmt. Bestimmte, nicht für die Intel®-Mikroarchitektur spezifische Optimierungen sind für Intel®-Mikroprozessoren reserviert. Entnehmen Sie weitere Informationen zu den spezifischen Befehlssatzerweiterungen, auf die dieser Hinweis zutrifft, bitte den entsprechenden Benutzer- und Referenzhandbüchern.

#### Konfigurationen für Inferenz-Durchsatz:

Prozessor: Intel® Xeon® Platinum 8180 (2 Sockel, 2,5 GHz, 28 Kerne, HT aktiviert, Turbo aktiviert). Arbeitsspeicher: 376,46 GB (12 Slots, 32 GB, 2.666 MHz). CentOS Linux-7.3.1611-Core. SSD sda RS3WC080 HDD 744,1 GB, sdb RS3WC080 HDD 1,5 TB, sdc RS3WC080 HDD 5,5 TB, Deep-Learning-Framework Caffe-Version: f6d01efbe93f70726ea3796a4b89c612365a6341. Topologie: googlenet\_v1. BIOS: SE5C620.86B.00.01.0004.071220170215. MKL-DNN-Version: ae00102be506ed0fe2099c6557df2aa88ad57ec1 NoDataLayer. Messwert: 1190 Bilder/s vs. Plattform: 25 Intel® Xeon® Prozessor E5-2699 v3 @ 2,30 GHz (18 Kerne), HT ON, Turbo OFF, Skalierungsregler über intel\_pstate Treiber auf "Performance" gesetzt, 256 GB DDR4-2133 ECC-RAM. CentOS Linux, Release 7.3.1611 (Kern), Linux-Kernel 3.10.0-514.el7.x86\_64. Betriebssystemlaufwerk: Seagate\* Enterprise ST2000NX0253, 2 TB, 2,5" interne Festplatte. Leistung gemessen mit: Umgebungsvariablen: KMP\_AFFINITY=granularity=fine, compact, 1,0, OMP\_NUM\_THREADS=36, CPUFreq festgelegt mit: cpupower frequency-set -d 2.2G -u 2.2G -g performance. Deep-Learning-Framework: Intel® Caffe: (<http://github.com/intel/caffe/>), Revision b0ef3236528a2c7d2988f249d347d5fdae831236. Inferenz gemessen mit „caffe time --forward\_only“-Befehl, Training gemessen mit „caffe time“-Befehl. Für die „ConvNet“-Topologien wurde ein Test-Datensatz verwendet. Für andere Topologien wurden Daten im lokalen Datenspeicher gespeichert und vor dem Training im System Speicher zwischengespeichert. Topologie-Spezifikation von [https://github.com/intel/caffe/tree/master/models/intel\\_optimized\\_models](https://github.com/intel/caffe/tree/master/models/intel_optimized_models) (GoogLeNet, AlexNet und ResNet-50), [https://github.com/intel/caffe/tree/master/models/default\\_vgg\\_19](https://github.com/intel/caffe/tree/master/models/default_vgg_19) (VGG-19) und [https://github.com/soumith/convnet-benchmarks/tree/master/caffe/imagenet\\_winners](https://github.com/soumith/convnet-benchmarks/tree/master/caffe/imagenet_winners) (ConvNet-Benchmarks; die Dateien wurden aktualisiert und verwenden das neuere prototxt-Format, sind aber funktional gesehen äquivalent). GCC 4.8.5, MKLML-Version 2017.0.2.20170110. BVLC-Caffe: <https://github.com/BVLC/caffe>, Inferenz & Training gemessen mit dem „caffe time“-Befehl. Für die „ConvNet“-Topologien wurde ein Test-Datensatz verwendet. Für andere Topologien wurden Daten im lokalen Datenspeicher gespeichert und vor dem Training im System Speicher zwischengespeichert BVLC Caffe (<http://github.com/BVLC/caffe/>), Revision 91b09280f5233cafc62954c98ce8bc4c204e7475 (Zusagedatum 14.5.2017). BLAS: atlas ver. 3.10.1.

#### Konfiguration für den Trainingsdurchsatz:

Prozessor: Intel® Xeon® Platinum 8180 (2 Sockel, 2,5 GHz, 28 Kerne, HT aktiviert, Turbo aktiviert). Arbeitsspeicher: 376,46 GB (12 Slots, 32 GB, 2.666 MHz). CentOS Linux-7.3.1611-Core. SSD sda RS3WC080 HDD 744,1 GB, sdb RS3WC080 HDD 1,5 TB, sdc RS3WC080 HDD 5,5 TB, Deep-Learning-Framework Caffe-Version: f6d01efbe93f70726ea3796a4b89c612365a6341. Topologie: alexnet. BIOS: SE5C620.86B.00.01.0009.101920170742. MKL-DNN-Version: ae00102be506ed0fe2099c6557df2aa88ad57ec1 NoDataLayer. Gemessen: 1023 Bilder/s im Vergleich zu Plattform: Intel® Xeon® Prozessor E5-2699 v3 (2 Sockel, 2,3 GHz, 18 Kerne, HT aktiviert, Turbo deaktiviert, CPU-Skalierung mittels intel\_pstate Treiber auf „Performance“ gesetzt), Arbeitsspeicher: 256 GB, DDR4-2133 ECC RAM. CentOS Linux, Release 7.3.1611 (Kern), Linux-Kernel 3.10.0-514.el7.x86\_64. Betriebssystemlaufwerk: Interne Festplatte Seagate\* Enterprise ST2000NX0253 (2 TB, 2,5") Leistung gemessen mit: Umgebungsvariable: KMP\_AFFINITY=granularity=fine, compact, 1,0; OMP\_NUM\_THREADS=36, CPUFreq festgelegt mit: cpupower frequency-set -d 2.3 G -u 2,3G -g performance. Deep Learning-Framework: Intel Caffe: (<http://github.com/intel/caffe/>), Revision b0ef3236528a2c7d2988f249d347d5fdae831236. Inferenz gemessen mit „caffe time --forward\_only“-Befehl, Training gemessen mit „caffe time“-Befehl. Für die „ConvNet“-Topologien wurde ein Test-Datensatz verwendet. Für andere Topologien wurden Daten im lokalen Datenspeicher gespeichert und vor dem Training im System Speicher zwischengespeichert. Topologie-Spezifikation von [https://github.com/intel/caffe/tree/master/models/intel\\_optimized\\_models](https://github.com/intel/caffe/tree/master/models/intel_optimized_models) (GoogLeNet, AlexNet, and ResNet-50), [https://github.com/intel/caffe/tree/master/models/default\\_vgg\\_19](https://github.com/intel/caffe/tree/master/models/default_vgg_19) (VGG-19) und [https://github.com/soumith/convnet-benchmarks/tree/master/caffe/imagenet\\_winners](https://github.com/soumith/convnet-benchmarks/tree/master/caffe/imagenet_winners) (ConvNet-Benchmarks, die Dateien wurden aktualisiert und verwenden das neuere Caffe-Prototxt-Format, sind aber funktional gesehen äquivalent). GCC 4.8.5, MKLML-Version 2017.0.2.20170110. BVLC-Caffe: <https://github.com/BVLC/caffe>, Inferenz & Training gemessen mit dem „caffe time“-Befehl. Für die „ConvNet“-Topologien wurde ein Test-Datensatz verwendet. Für andere Topologien wurden Daten im lokalen Datenspeicher gespeichert und vor dem Training im System Speicher zwischengespeichert BVLC Caffe (<http://github.com/BVLC/caffe/>), Revision 91b09280f5233cafc62954c98ce8bc4c204e7475 (Zusagedatum 14.5.2017). BLAS: atlas ver. 3.10.1.

Durch Technologien von Intel ermöglichte Funktionsmerkmale und Vorteile hängen von der Systemkonfiguration ab und können entsprechend geeignete Hardware, Software oder die Aktivierung von Diensten erfordern. Die Leistungsmerkmale variieren je nach Systemkonfiguration. Kein Computersystem bietet absolute Sicherheit. Informieren Sie sich beim Systemhersteller oder Fachhändler oder auf [intel.de](http://intel.de).

In Leistungstests verwendete Software und Workloads können speziell für die Leistungseigenschaften von Intel®-Mikroprozessoren optimiert worden sein. Leistungstests wie SYSmark\* und MobileMark\* werden mit spezifischen Computersystemen, Komponenten, Softwareprogrammen, Operationen und Funktionen durchgeführt. Jede Veränderung bei einem dieser Faktoren kann abweichende Ergebnisse zur Folge haben. Als Unterstützung für eine umfassende Bewertung Ihrer geplanten Anschaffung sollten Sie noch andere Informationen und Leistungstests heranziehen – auch im Hinblick auf die Leistung des betreffenden Produkts in Verbindung mit anderen Produkten. Ausführlichere Informationen finden Sie unter <http://www.intel.de/benchmarks>.

Die Benchmark-Ergebnisse wurden vor der Implementierung der bereitgestellten Software-Patches und Firmware-Updates ermittelt, die als Gegenmaßnahmen für die als „Spectre“ und „Meltdown“ bezeichneten Exploits bereitgestellt wurden. Die Implementierung dieser Updates kann dazu führen, dass diese Ergebnisse auf Ihr Gerät oder System nicht zutreffen.

Alle hierin gemachten Angaben können sich jederzeit ohne besondere Mitteilung ändern. Wenden Sie sich an Ihren Ansprechpartner bei Intel, um die neuesten Produktspezifikationen und Roadmaps zu erhalten.

Intel, Xeon, Saffron und das Intel Logo sind Marken der Intel Corporation in den USA und/oder anderen Ländern.

\*Andere Marken oder Produktnamen sind Eigentum der jeweiligen Inhaber.