ONLOGIC

KI-Crashkurs

Alles, was Sie schon immer über KI wissen wollten, aber sich nicht trauten, ChatGPT zu fragen



Inhaltsverzeichnis

Einleitung 03

Was ist KI? 03

Wie funktioniert KI? 04

- KI-Training 04
- KI-Inferenzierung 04

Die Bereiche der KI 05

- Machine Learning (ML) 05
- Deep Learning 05
- Natural Language Processing (NLP) 05
- Expertensystem 05
- Robotik 06
- Machine Vision 06

Die Vorteile von KI für die Fabrikautomation 06

- Verlängerte Lebensdauer der Geräte 06
- Weniger Ausschuss 06
- Sicherere Arbeitsumgebungen 06
- Gesteigerte Produktivität 07
- Verbesserte Bestandsverwaltung und Bedarfsprognose 07
- Rationalisierte Fabriklayouts 07

KI im Edge 07

- Ein Beispiel für Edge-KI 07

Technologien für Lösungen im Bereich der künstlichen Intelligenz 08

- GPU vs. CPU 09
- Wofür wird eine GPU verwendet? 09

Die verschiedenen Arten von KI-Beschleunigern 10

- Integrierte GPU (iGPU) 10
- Diskrete GPU 10
- TPU 10
- NPU 10

Industrie-Computer und KI 10

Überlegungen zu einer GPU in einem Industrie-Computer 11

- Eine iGPU für industrielle KI 11
- Diskrete GPU für industrielle KI 11

Einstieg in eine KI-Lösung – eine Schritt-für-Schritt-Anleitung 12

Hardware für Lösungen im Bereich der künstlichen Intelligenz 13

Die wichtigsten Erkenntnisse 15

Ein Glossar von Begriffen rund um KI 16

Ein Crashkurs zu künstlicher Intelligenz (KI)

Künstliche Intelligenz ist in kürzester Zeit zu einem der wichtigsten Schlagworte aufgestiegen, nicht zuletzt, weil sie das Potenzial hat, praktisch jede Branche zu revolutionieren. Wenn Sie noch nicht genau wissen, was KI ist, was sie kann oder wie sie sich auf Ihr Leben oder Ihr Unternehmen auswirkt, sind Sie hier genau richtig.

Lesen Sie weiter und erhalten Sie einen Crashkurs in KI sowie einen Einblick in die Elemente und Fähigkeiten, die sie so leistungsstark machen. Wir gehen auch auf die Vorteile und Möglichkeiten von KI ein und zeigen Ihnen, was Sie jetzt tun müssen, um die Implementierung von KI-Lösungen vorzubereiten. Dazu gehört auch eine Schritt-für-Schritt-Anleitung. Abschließend haben wir ein Glossar mit Begriffen rund um KI zusammengestellt. So finden Sie sich in der KI-Landschaft leichter zurecht.

Was ist KI?

KI ist die Abkürzung für künstliche Intelligenz und befasst sich mit Technologien, die versuchen, die Ergebnisse menschlicher Intelligenz zu replizieren. Um dies zu erreichen, muss KI die komplexe Verarbeitung von Informationen beherrschen. Dazu gehören Lernen, logisches Denken, Problemlösung, Sprachgebrauch und Wahrnehmung.



Wie funktioniert KI?

Bei KI handelt es sich nicht nur um ein einzelnes Computerprogramm oder eine Anwendung, sondern um einen ganzen Bereich der Informatik. Im Wesentlichen gibt es zwei Phasen der KI: Training und Inferenz.

KI-Training

Das Training für KI ist der Prozess, bei dem einem KI-System beigebracht wird, Daten zu erkennen, zu verarbeiten und daraus zu lernen. Ziel des Trainings ist es, dem System die Inferenz (Entscheidungsfindung) zu ermöglichen, sobald es eingesetzt wird. Während des Trainingsprozesses werden große Mengen von Eingabedaten und erwarteten Ausgaben analysiert, um eine Reihe strukturierter Muster zu formulieren. Diese strukturierten Muster werden als “Modell” bezeichnet.

Das Ziel ist, dass das KI-Modell fundierte Vorhersagen machen kann, wenn es neue Daten erhält. Effektives KI-Training erfordert riesige Datenmengen und viel Rechenleistung. Oft werden dabei Multicore-Prozessoren und GPUs verwendet (dazu gleich mehr).

Was ist eine GPU?

GPU steht für Graphics Processing Unit. Manchmal werden sie auch Grafikkarten oder Videokarten genannt. Wie der Name schon sagt, haben Ingenieur:innen GPUs ursprünglich zur Verarbeitung von Grafiken wie Bildern und Videos entwickelt, damit sich die CPU auf andere Aufgaben konzentrieren kann.

Grafikprozessoren sind für die Parallelverarbeitung ausgelegt. Dadurch können sie komplexe Probleme in einzelne Aufgabenströme aufteilen und diese parallel bearbeiten, sodass sie alle auf einmal abgeschlossen werden. Im Gegensatz zu einer CPU werden sie nicht einzeln bearbeitet.

Es hat sich gezeigt, dass dieses Design nicht nur für das Rendern von Bildern und Videos geeignet ist. Dank der Parallelverarbeitung ist es auch äußerst effektiv für künstliche Intelligenz (KI), Machine Learning, Computer Vision und mehr. GPUs beseitigen mögliche Verarbeitungsengpässe und sorgen so für schnellere Ergebnisse, mehr Möglichkeiten und ein besseres Benutzererlebnis.

KI-Inferenzierung

KI-Inferenzierung ist der Prozess, bei dem neue Daten in ein trainiertes Modell eingespeist und in verwertbare Erkenntnisse umgewandelt werden. Bei der KI-Inferenzierung wird der Wert der KI erst richtig deutlich, denn die erstellten Modelle sind nur so nützlich wie die aus der Inferenzierung abgeleiteten Ergebnisse.

Aus Sicht der Hardware ist die Durchführung von Inferenzvorgängen im Vergleich zum KI-Training im Allgemeinen weit weniger rechenintensiv. Dies eröffnet die Möglichkeit, KI-Inferenzen auf leistungsärmeren und kostengünstigeren Prozessoren wie NPUs und Allzweck-CPUs durchzuführen.

Die Bereiche der KI

Der Bereich der künstlichen Intelligenz ist groß und wird von Tag zu Tag größer. Die Anwendungen der künstlichen Intelligenz lassen sich in mehrere Bereiche unterteilen. Dazu gehören Machine Learning, Deep Learning, Natural Language Processing, Expertensysteme, Robotik und Machine Vision.

Machine Learning (ML)

Machine-Learning-Modelle verwenden Daten und Algorithmen, um bestimmte Aufgaben zu erfüllen, ohne explizit programmiert zu werden. Diese Machine-Learning-Algorithmen verbessern ihre Genauigkeit mit der Zeit.

- Plus One Robotics zum Beispiel entwickelt Machine-Learning-Lösungen für die Lagerhaltung und den Vertrieb.

Im Allgemeinen gibt es drei große Unterkategorien von Machine Learning

- Algorithmus des überwachten Lernens
    - Datenwissenschaftler:innen liefern gekennzeichnete Trainingsdaten. Maschinen verwenden die Daten, um Ergebnisse vorherzusagen.
- Algorithmus des unüberwachten Lernens
    - Anwendungen suchen nach Mustern in nicht gekennzeichneten Daten.
- Machine Learning durch Verstärkung
    - Der Algorithmus lernt, indem er mit seiner Umgebung interagiert und sein Verhalten auf der Grundlage positiver oder negativer Reaktionen auf Aktionen verbessert.

Deep Learning

Deep Learning ist eine Form des Machine Learning, bei der Algorithmen in Schichten strukturiert werden, um ein „künstliches neuronales Netzwerk“ zu schaffen. Deep Learning ist ein fortschrittlicherer Ansatz des Machine Learning und kann zur Lösung komplexerer Probleme verwendet werden. Autonomes Fahren ist eine Form des Deep Learning.

- Bear Flag Robotics zum Beispiel entwickelt selbstfahrende autonome Traktoren, die durch Deep Learning betrieben werden.

Natural Language Processing (NLP)

NLP-Algorithmen werden verwendet, um geschriebene oder gesprochene menschliche Sprache zu verarbeiten. Sie werden zur Übersetzung, Zusammenfassung oder zur Durchführung einer Aktion verwendet.

- Google Translate ist ein Beispiel für NLP. Es übersetzt Text von einer Sprache in eine andere.

Expertensystem

Ein Expertensystem ist eine Software, die KI einsetzt, um Probleme zu lösen und das Urteil menschlicher Expert:innen zu simulieren.

- MYCIN, ein in den 1970er Jahren in Stanford entwickeltes Programm, ist ein Beispiel für ein Expertensystem. Es unterstützt Ärzt:innen bei der Diagnose und Behandlung von Infektionen.



Robotik

Die Robotik ist ein Bereich der Technik, in dem künstliche Intelligenz eingesetzt wird, um Maschinen bei der Navigation und Steuerung ihrer Umgebung zu unterstützen.

- So stellt advanced.farm beispielsweise Roboter für die Obsternte her. Sie helfen beim Erkennen und Pflücken von reifem Obst und Gemüse.

Machine Vision

Machine Vision nutzt die neuesten KI-Technologien, um Industrieanlagen zur Visualisierung ihrer Umgebung zu befähigen und schnelle Entscheidungen auf der Grundlage dessen zu treffen, was sie „sehen“.

- So entwickelt Artemis Vision beispielsweise Bildverarbeitungslösungen für die Qualitätskontrolle in der Fertigung. Sie erfassen kleinste Details, die dem menschlichen Auge entgehen könnten.

Die Vorteile von KI für die Fabrikautomation

KI bietet zahlreiche Vorteile für die Fabrikautomation. Sie bringt Möglichkeiten zur Kostensenkung und zur Herstellung von qualitativ hochwertigeren Produkten mit sich. Hier sind unsere 6 wichtigsten Vorteile.

Verlängerte Lebensdauer der Geräte

- Die KI-gestützte vorausschauende Wartung nutzt Sensordaten von Maschinen, um Schlüsselattribute wie Temperatur, Vibration, Druck usw. zu erkennen und festzustellen, wann eine Wartung der Anlage erforderlich ist. Dies kann ungeplante Ausfallzeiten von Maschinen verringern, die Wartungskosten senken und sogar die Lebensdauer von Maschinen verlängern.

Weniger Ausschuss

- In einer Produktionslinie können mehrere Qualitätskontrollpunkte eingerichtet werden. Sie erfassen Qualitätsprobleme, die dem menschlichen Auge entgehen könnten. Dadurch wird die Prüfgenauigkeit drastisch erhöht. Wenn Fehler frühzeitig im Produktionsprozess erkannt werden, lässt sich der Ausschuss reduzieren.

Sicherere Arbeitsumgebungen

- KI-gestützte Lösungen, wie z. B. die Qualitätskontrolle, können an Orten platziert werden, die für Menschen nicht sicher sind, wie etwa direkt an der Produktionslinie.

- Roboterlösungen können für die schwierigsten und/oder sich wiederholende Aufgaben in einer Produktionsstätte entwickelt werden.

Gesteigerte Produktivität

- KI-gestützte Automatisierung, etwa durch den Einsatz von Robotern, macht einen Betrieb rund um die Uhr möglich. Dabei werden selbst sich wiederholende Aufgaben mit Spitzenleistung und einem hohen Maß an Genauigkeit ausgeführt.

Verbesserte Bestandsverwaltung und Bedarfsprognose

- KI kann Trends im Bestand und in der Nachfrage erkennen, um die Bestände zu optimieren, die Nachfrage vorherzusagen und sogar Sendungen in einer komplexen globalen Lieferkette zu verfolgen. Außerdem kann sie sicherstellen, dass Entscheidungen in der Fertigung auf den neuesten und aktuellsten Daten basieren.

Rationalisierte Fabriklayouts

- KI kann Ineffizienzen in den Produktionsabläufen erkennen, Engpässe beseitigen und die Produktion in einer Fabrikhalle steigern. Eine Möglichkeit dazu ist die Erstellung eines digitalen Zwillings der Fabrikhalle. An diesem Zwilling können Anpassungen vorgenommen werden, um Ergebnisse zu ermitteln.

KI im Edge

Um eine Entscheidungsfindung nahezu in Echtzeit zu ermöglichen, verlagern viele Unternehmen KI-Lösungen von der Cloud zum Edge und damit näher an die Quelle der Systeme, die die Daten erzeugen. So werden verwertbare Erkenntnisse in Echtzeit geliefert.

Die Erstellung von KI-Modellen und Schlussfolgerungen kann natürlich in einer zentralen Cloud oder in externen Rechenzentren erfolgen. Die Bereitstellung von KI-Funktionen am Rand (Edge) des Netzwerks trägt jedoch dazu bei, Entscheidungen und Maßnahmen auf der Grundlage der von den Sensoren vor Ort erfassten Informationen zu beschleunigen. Dieser Rand des Netzwerks kann sich in einem Lager, an einer Fertigungsstraße, auf einem Gabelstapler oder sogar an einem entfernten, externen Standort befinden.

Ein Beispiel für Edge-KI

Autonome Fahrzeuge wie die von Bear Flag Robotics entwickelten Traktoren sind ein Beispiel für KI im Edge. Informationen aus einer Fülle von Eingaben, darunter 3D-Kameras, LiDAR und Positionssensoren werden gesammelt und verarbeitet. So erhalten die Fahrzeuge Informationen darüber, wo sie abbiegen müssen, um Bodenunebenheiten zu vermeiden, wie sie ihre Arbeitsgeräte einstellen müssen und wann sie bremsen müssen, um Hindernissen auszuweichen.

All diese Entscheidungen müssen schnell getroffen werden – es bleibt keine Zeit, die Daten an die Cloud zu senden, sie zu verarbeiten und dann Anweisungen an den Traktor zurückzusenden. Wenn die Daten direkt auf dem Computer im Traktor und nicht in der Cloud erfasst und verarbeitet werden, wird diese Verzögerung, die sogenannte Latenz, drastisch reduziert.

Ganz zu schweigen davon, dass der Zugang zu einer Internetverbindung auf einem landwirtschaftlichen Feld herausfordernd sein kann. Mit KI am Edge kann die Datenverarbeitung einfach weitergehen, unabhängig davon, wo die Daten erzeugt werden.

Technologien für Lösungen im Bereich der künstlichen Intelligenz

Das explosive Wachstum von KI und der daraus resultierende Wert ist eng mit der explosionsartigen Zunahme der verfügbaren Daten und Modelle sowie der Weiterentwicklung der Technologie zur Verarbeitung und Auswertung der gesammelten Informationen verbunden. Zu den wichtigsten Fortschritten gehören unter anderem:

Verfügbarkeit, Organisation und Klassifizierung von Daten

- Das Internet of Things (IoT) bietet die Möglichkeit, Daten aus zahlreichen verbundenen Geräten und offenen Datensätzen zu erfassen.
- Die datenzentrierte KI ist eine aufstrebende Wissenschaft, die sich mit der Optimierung von Datensätzen befasst, um eine bessere Leistung bei praktischen ML-Anwendungen zu erzielen.

Was ist das Internet of Things (IoT)?

IoT bezieht sich auf die Sammlung von Geräten, die über das Internet miteinander kommunizieren. Zu den gängigen IoT-Hardwaregeräten gehören tragbare Geräte (Wearables), Smart-Home-Geräte und mit der Cloud verbundene industrielle Sensoren.

Was ist das AIoT?

Die AIoT kombiniert KI mit dem IoT, um vernetzte Geräte in die Lage zu versetzen, Informationen zu verarbeiten und zu erlernen. Dies wiederum kann dazu beitragen, künftige Muster oder Ereignisse vorherzusagen, die Entscheidungen darüber ermöglichen, wie zu reagieren ist.

Was bedeutet Open-Source?

Der Begriff “Open Source” bezieht sich auf etwas, das von allen verändert und weitergegeben werden kann, weil das Design öffentlich zugänglich ist. Open-Source-Software ist beispielsweise Software mit Quellcode, den jede:r einsehen, verändern und verbessern kann.

Zu den Vorteilen von Open-Source-KI-Tools gehören Transparenz, kollaborative Beteiligung, schnelles Prototyping und gemeinschaftsorientierte Entwicklung. Dies trägt zu einem schnelleren Fortschritt bei, da die Entwickler:innen voneinander lernen und auf der Arbeit anderer aufbauen können. Einige sehen darin auch einen Weg, mögliche Quellen von Verzerrungen in der Entwicklung von KI zu identifizieren und ein KI-Monopol von Unternehmen zu vermeiden.

Verfügbarkeit von Open-Source-KI-Tools

- Open-Source-KI-Frameworks, Toolkits und vorgefertigte Modelle haben eine schnelle Einführung von KI ermöglicht. Sie bieten die Bausteine für die Entwicklung, das Training, die Validierung und den Einsatz von KI-Lösungen. Durch den Wegfall der manuellen Kodierung und der traditionellen Softwareentwicklungsmethoden können diese Toolkits die Kosten erheblich senken und die Zeit bis zur Bereitstellung verkürzen.
- Einige Beispiele für Toolkits sind OpenVino, TensorFlow, Apache, Hugging Face und PyTorch.

Fortschritte bei den Prozessoren

- Multi-Core-CPUs mit integrierten GPUs und NPUs (Neural Processing Units) sowie moderne eigenständige Beschleuniger und GPUs mit neuen KI-spezifischen Erweiterungen sind die Grundlage für den steigenden Wert von KI. Sie liefern die Rechenleistung für die Verarbeitung und Interpretation von Datensätzen zur Erstellung eines KI-Algorithmus.

> ### Was ist eine CPU?
>
> CPU steht für Central Processing Unit (zentrale Verarbeitungseinheit) und wird allgemein als das Gehirn des Computers bezeichnet. Vereinfacht gesagt, ist die CPU ein elektronischer Schaltkreis, der Daten steuert – einschließlich Eingabe, Ausgabe und Speicherung. Die CPU sorgt dafür, dass das Betriebssystem und die Anwendungen des Computers funktionieren.

GPU vs. CPU

Wir haben die Definitionen von GPU und CPU erläutert, aber was ist der Unterschied? Sowohl die GPU als auch die CPU sind siliziumbasierte Mikroprozessoren, die Daten verarbeiten. Die Unterschiede liegen in der Art und Weise, wie sie Daten verarbeiten, und in ihren physischen Verarbeitungseinheiten – ihren Kernen.

Datenverarbeitung

- CPUs führen serielle Berechnungen durch, das heißt, sie verarbeiten eine Aufgabe nach der anderen.
- GPUs führen parallele Berechnungen durch, das bedeutet, sie verarbeiten viele Berechnungen auf einmal.

Die Cores

- CPUs haben lediglich einige wenige Cores.
  - So sind zum Beispiel Intel® Hybrid Core Prozessoren der 13. Generation mit bis zu 24 Cores erhältlich.
  - Beachten Sie, dass Intel mit der 12. Generation die Hybrid-Core-Technologie eingeführt hat. Mehr dazu erfahren Sie hier.
- GPUs können Hunderte bis Tausende von Cores haben, die den CPU-Cores ähnlich, aber kleiner sind..
  - So bietet beispielsweise die Nvidia RTX A4000 GPU über 6.300 Cores, darunter 48 RT-Cores der zweiten Generation, 192 Tensor-Cores der dritten Generation und 6.144 CUDA-Cores.

Wofür wird eine GPU verwendet?

Sowohl CPUs als auch GPUs sind wirklich gut in Mathe. Aber wie bereits erwähnt, verarbeiten CPUs die Gleichungen nacheinander, Aufgabe für Aufgabe. Ein Grafikprozessor hingegen kann die Gleichungen alle gleichzeitig verarbeiten. Aus diesem Grund bieten GPUs eine außergewöhnliche Berechnungsleistung für komplexe mathematische Berechnungen.

Das Rendern von Bildern und Grafiken läuft auf eine Reihe mathematischer Probleme hinaus – insbesondere die Operationen der Multiplikation und Addition in einer strukturierten Weise, die als Matrixmultiplikation bezeichnet wird. Dies unterstützt die verschiedenen Funktionsblöcke der KI. Und wie wir bereits erwähnt haben, sind GPUs darin hervorragend. Sie beschleunigen diese Aufgaben für eine überragende Grafikleistung, weshalb sie auch als „Beschleuniger“ bezeichnet werden.

Während GPUs ursprünglich zur Beschleunigung der Grafikverarbeitung entwickelt wurden, fanden Ingenieur:innen bald heraus, dass GPUs auch die fortgeschrittenen Herausforderungen der künstlichen Intelligenz bewältigen können. Dazu gehört auch die Fähigkeit, eine große Menge an Daten aufzunehmen, sie zu klassifizieren und diese Informationen zu nutzen, um Rückschlüsse und Vorhersagen für zukünftige Ergebnisse zu treffen.

Die verschiedenen Arten von KI-Beschleunigern

Integrierte GPU (iGPU)

Eine integrierte GPU, manchmal auch iGPU genannt, ist direkt in den Prozessor oder die CPU des Computers eingebaut. Sie teilt sich Strom und Systemspeicher mit der CPU. Infolgedessen sind Computer mit integrierten Grafikprozessoren stromsparend und im Allgemeinen leichter und kleiner. Da der Speicher jedoch gemeinsam genutzt wird, kann die Verarbeitung umfangreicher GPU-Aufgaben zu Latenzzeiten mit der allgemeinen CPU-Verarbeitung führen.

Diskrete GPU

Eine diskrete oder dedizierte GPU ist ein Grafikprozessor, der vollständig von der CPU getrennt ist. Diese GPU verfügt über einen eigenen Speicher, den Videospeicher oder VRAM. Mit ihrem eigenen Speicher ist diese GPU nicht auf den Arbeitsspeicher des Computers angewiesen (manche würden sagen, sie nimmt ihn in Beschlag). Diskrete Grafikkarten bieten eine bessere Leistung, verbrauchen aber mehr Strom und erzeugen daher viel Wärme.

Die Kühlung eines diskreten Grafikprozessors ist ein wichtiger Aspekt bei der Entwicklung, um eine maximale Leistung zu erzielen. Aus diesem Grund haben die Ingenieur:innen diese GPUs mit eigenen Lüftern ausgestattet. Diese können den Platzbedarf und den Energieverbrauch erhöhen.

Weitere Informationen finden Sie in unserem GPU-Leitfaden zum Einstieg.

TPU

Eine Tensor Processing Unit (TPU) ist eine anwendungsspezifische integrierte Schaltung (ASIC), die von Google entwickelt wurde. Die Hauptaufgabe von TPUs ist die mathematische Matrixverarbeitung, die Machine-Learning-Workloads beschleunigt. In den Worten von Google: “TPUs können keine Textverarbeitungsprogramme ausführen, Raketentriebwerke steuern oder Banktransaktionen ausführen, aber umfangreiche Matrixoperationen in neuronalen Netzwerken mit hoher Geschwindigkeit verarbeiten.”

NPU

Eine Neural Processing Unit (NPU) wurde speziell dafür entwickelt, die Ausführung neuronaler Netzwerke für Deep Learning zu beschleunigen. NPUs bieten hohe Leistung bei minimalem Stromverbrauch. Daher werden NPUs häufig in mobilen Geräten, Edge-Computing und anderen energiesensiblen Anwendungen eingesetzt.

Industrie-Computer und KI

Industrie-Computer können mit Beschleunigern wie GPUs für KI und andere Deep-Learning-Anwendungen ausgestattet werden. Ingenieur:innen haben diese Computer so konzipiert, dass sie extremen Umgebungen standhalten und es Unternehmen ermöglichen, KI fast überall dort einzusetzen, wo sie benötigt wird.

So werden sie zum Beispiel für den Betrieb von Robotern in Fertigungsanlagen, Warenlagern und sogar auf landwirtschaftlichen Feldern eingesetzt. Dort sind sie oft Vibrationen und extremen Temperaturen ausgesetzt. Sie werden auch für die Qualitätskontrolle verwendet, manchmal direkt an einer Produktionslinie, deren Umgebung für Menschen möglicherweise nicht sicher ist, ganz zu schweigen von einem herkömmlichen Desktop-Computer.



Überlegungen zu einer GPU in einem Industrie-Computer

Industrie-Computer sind so konstruiert, dass sie extremen Umgebungsbedingungen standhalten. Ein gängiges Konstruktionsmerkmal für Industrie-PCs ist ihr lüfterloses und belüftungsfreies Design. Dieses Konzept bietet zwar viele Vorteile, eignet sich aber nicht für alle GPUs, da einige von ihnen im Vergleich zu CPUs einen deutlich höheren Stromverbrauch haben und viel Wärme erzeugen können.

Eine iGPU für industrielle KI

Die Verwendung und Kühlung integrierter Grafikkarten, die eine einfache Bild- und Objekterkennungslösung versorgen können, ist mit einem lüfterlosen Design durchaus möglich. Dies gilt insbesondere für einige der neuesten und unglaublich leistungsstarken Prozessoren.

Diskrete GPUs werden nicht immer benötigt und können die Kosten eines Computers erheblich erhöhen. Die Ausführung von KI-Inferenzierung auf einer CPU oder einem integrierten Grafikprozessor (iGPU) trägt dazu bei, die Kosten für einen effektiven KI-Einsatz zu senken.

Wenn es um Industrie-Computer für KI geht, ermöglicht ein lüfterloses Design einen kleinen Formfaktor, der ein wenig mehr Flexibilität bei der Installation bietet. Der Computer benötigt keinen Lüfter, um die Luft in das Gehäuse zu leiten. Daher können diese lüfterlosen Computer auch in Bereichen installiert werden, in denen viel Staub oder Schmutz in der Luft vorhanden ist und/oder in denen Vibrationen auftreten können.

Diskrete GPU für industrielle KI

Eine komplexere KI- oder ML-Lösung erfordert möglicherweise eine diskrete oder dedizierte Grafikkarte. Wie bereits erwähnt, erzeugen GPUs eine Menge Wärme. Das bedeutet, die Kühlung eines Grafikprozessors ist ein wichtiger Aspekt. Die meisten diskreten Grafikprozessoren sind mit eigenen Lüftern und Temperaturbegrenzungen für die Umgebung konfiguriert.

Eine Option ist ein lüfterloser Hybrid-Computer, der einen separaten Erweiterungsschacht bietet. In diesem Schacht kann eine GPU installiert und aktiv mit einem Lüfter gekühlt werden. Gleichzeitig sind das Motherboard und andere empfindliche interne Komponenten im belüftungsfreien Teil des Gehäuses geschützt. Falls Sie die GPU austauschen müssen, ist sie leicht zugänglich.

Einstieg in eine KI-Lösung – eine Schritt-für-Schritt-Anleitung

Sie sind also zu dem Schluss gekommen, dass die Implementierung von KI für Ihr Unternehmen von Vorteil sein könnte. Nun sind Sie bereit, damit zu beginnen. Aber wie fangen Sie an? Es ist leicht, sich sofort auf die Suche nach Hardware und Software zu machen, aber es gibt ein paar wichtige Schritte, die Sie zuerst beachten sollten.

1. **Führen Sie eine Bedarfsanalyse durch**
    - Ermitteln Sie die Bereiche Ihres Unternehmens, die von KI-gestützter Automatisierung profitieren können. Einige gängige Beispiele umfassen:
        - sich wiederholende Aufgaben
        - Qualitätskontrolle
        - vorausschauende Wartung
        - Lieferketten-Management.
    - Anhand Ihrer Evaluierung können Sie dann feststellen, ob eine KI-gestützte Lösung für Ihr Unternehmen geeignet ist.
2. **Definieren Sie klare Zielsetzungen**
    - Legen Sie SMART-Ziele (spezifisch/specific, messbar/measurable, erreichbar/achievable, relevant, zeitlich begrenzt/time bound) für Ihre KI-gestützte Automatisierungsbereitstellung fest. Ganz gleich, ob es um die Verbesserung der Effizienz, die Senkung der Kosten oder die Verbesserung der Produktqualität geht, klar definierte Ziele steuern Ihre KI-Implementierung. Sie helfen Ihnen dabei, zu beurteilen, was funktioniert und was nicht.
3. **Stellen Sie Ihr Team zusammen**
    - Stellen Sie ein Team zusammen, dem Personen aus dem gesamten Unternehmen angehören, insbesondere aus den Bereichen Operations und IT. Durch diese Zusammenarbeit wird sichergestellt, dass die Implementierung sowohl mit den technischen Anforderungen als auch mit den Produktionsanforderungen übereinstimmt.
4. **Beurteilen Sie Ihre Dateninfrastruktur**
    - Bewerten Sie die Verfügbarkeit und Qualität Ihrer Daten. KI ist auf große Datensätze zum Trainieren von Modellen angewiesen. Beginnen Sie damit, Daten aus allen Bereichen Ihrer Fertigungsprozesse zu sammeln.
5. **Arbeiten Sie mit Händlern zusammen**
    - Wenden Sie sich an Händler und Beratungsunternehmen, die sich auf Automatisierung spezialisiert haben. Sie können Ihnen Einblicke, Orientierungshilfen und möglicherweise maßgeschneiderte Lösungen für Ihre spezifischen Anforderungen bieten. Jede KI-Implementierung ist anders. Daher ist es wichtig, mit Händlern zusammenzuarbeiten, die die Zeit und die Ressourcen aufbringen können, um Ihre individuellen Herausforderungen zu verstehen.
6. **Entwerfen Sie Ihre Strategie zur kontinuierlichen Verbesserung**
    - Sie sollten Folgendes bewerten:
        - Erfüllt das KI-Modell seine Aufgabe gut genug?
        - Fängt es an, Dinge zu übersehen?
        - Muss es mit neuen Daten erneut trainiert werden?
        - Wie werden Sie im Laufe der Zeit neue Modelle einführen?

Hardware für Lösungen im Bereich der künstlichen Intelligenz

Was die KI-Hardwareoptionen angeht, so erfordern verschiedene Anwendungen unterschiedliche KI-Lösungen.

IoT-Gateways

Wenn sich Ihre KI-Lösung in der Cloud befindet, benötigen Sie ein IoT-Gateway. Dieser kleine, aber zuverlässige Computer ist das Bindeglied zwischen den von Embedded-Sensoren erfassten Daten und der Cloud. Sie spielen eine immer wichtigere Rolle in KI-Lösungen, da sie eingehende Daten erfassen, speichern und manchmal teilweise verarbeiten, bevor sie übertragen werden.

Beispielsweise haben wir den Karbon 410 so konzipiert, dass er selbst unter schwierigsten Installationsbedingungen zuverlässig arbeitet. Dazu gehören extreme Temperaturen und vibrationsanfällige Standorte.



KI im Edge mit integrierter GPU

Einige der neuesten Prozessoren mit integriertem Grafikprozessor können problemlos KI-Inferenzlösungen am Edge unterstützen. Der lüfterlose OnLogic Helix 511 wird beispielsweise von Intel-Prozessoren der 12. Generation mit Hybrid-Core-Architektur und DDR5-Speicher unterstützt. Dieses kompakte Kraftpaket bietet reichlich I/O, darunter auch ältere Verbindungen, und eine leistungsstarke Datenverarbeitung für eine KI-Lösung am Edge.



KI am Edge mit separater GPU

Wenn Sie eine robustere Lösung mit einem Intel Core™-Prozessor der 12. oder 13. Generation und einem diskreten Grafikprozessor implementieren möchten, ist der Karbon 804 geeignet, ein Fanless-Hybrid-PC mit hohen Rechenleistungen. Er verfügt über Optionen für GPU-Erweiterungskarten mit PCIe Gen 4.0-Funktionen. Der unglaubliche Durchsatz, den die PCIe Gen 4 bietet, bedeutet, dass Sie verschiedene Optionen haben, wenn es darum geht, zu entscheiden, welche Steckplätze und Lanes Sie für Ihre GPU verwenden möchten.

- Option 1: Ein einzelner PCIe-Riser mit x16-Steckplätzen und x16-Lanes, der einen Durchsatz von 256 GT/s ermöglicht.
- Option 2: Ein Dual-Riser, der zwei x16-Steckplätze mit jeweils x8 Lanes bietet. Sie könnten einen Steckplatz für die GPU verwenden. Dadurch erhalten Sie einen Durchsatz von 128 GT/s. Für viele KI-Anwendungen ist dieser Durchsatz mehr als ausreichend. Darüber hinaus verfügt der Computer dann über einen offenen Steckplatz, der Ihnen Flexibilität bei der Konfiguration für zusätzliche Erweiterungsoptionen bietet.





Benötigen Sie eine kleinere Alternative?

Unser Karbon 803 bietet einen etwas kleineren Formfaktor mit einem einzelnen PCIe Gen 4 x16 Steckplatz.



GPU-Server

Für komplexe Workloads, Deep Learning am Edge sowie Training und Inferenz vor Ort ist ein GPU-Server wie der AC101 eine hervorragende Lösung. Diese Plattform bietet Intel-Prozessoren der 13. Generation, fortschrittliche GPUs und DDR5-Speicher. Viele Unternehmen nutzen Edge-Server im Rahmen ihrer Strategien zur Rückführung in die Cloud, um Rechenressourcen vor Ort zu verlagern. Außerdem werden so Latenzzeiten vermieden und die Betriebskosten gesenkt.

Die wichtigsten Erkenntnisse

Wie unsere Beispiele zeigen, ist die KI dabei, nahezu jede Branche zu transformieren. Dadurch verändert sich die Art und Weise, wie wir leben, arbeiten und spielen. Jetzt ist es an der Zeit, Ihre technologische Infrastruktur zu stärken, damit Sie besser gerüstet sind, um von den Vorteilen der KI zu profitieren. Dazu gehören Verbesserungen bei Produktivität, Effizienz, Qualität und Kundenzufriedenheit.

Alles beginnt mit den folgenden 5 Schritten:

1. Führen Sie eine Bedarfsanalyse durch
2. Definieren Sie klare Zielsetzungen
3. Stellen Sie Ihr Team zusammen
4. Beurteilen Sie Ihre Dateninfrastruktur
5. Arbeiten Sie mit Händlern zusammen

Von dort aus können Sie ein Pilotprojekt für KI starten und schrittweise erweitern. Fangen Sie klein an, bauen Sie auf Ihren Erfolg auf und verbessern Sie kontinuierlich Ihre KI-Modelle und -Prozesse. Das OnLogic-Team ist bereit und hilft Ihnen dabei, Ihre Hardwareanforderungen zu verstehen. Gemeinsam finden wir heraus, wie ein Industrie-Computer Ihre anspruchsvollen Anforderungen erfüllen kann. Die Zukunft ist jetzt und der Einsatz von KI ist nicht mehr eine Frage des “ob”, sondern des “wann”. Wir sind für Sie da und erleichtern Ihnen den Einstieg.

Ein Glossar von Begriffen rund um KI

Algorithmus	Software, die eine Reihe von Anweisungen oder Regeln enthält, denen ein Computer folgt, um eine bestimmte Aufgabe zu erfüllen oder ein bestimmtes Problem zu lösen.
Algorithmus-Verzerrung	Unbeabsichtigte und systematische Fehler in einem Algorithmus, die zu ungerechten oder diskriminierenden Ergebnissen führen. Sie spiegeln häufig bestehende Verzerrungen in den Trainingsdaten wider.
Algorithmische Transparenz	Das Ausmaß, in dem die innere Funktionsweise eines Algorithmus offen und verständlich ist, insbesondere bei Anwendungen, bei denen Entscheidungen Auswirkungen auf den Einzelnen oder die Gesellschaft haben.
Big Data	Extrem große und komplexe Datensätze, die mit herkömmlichen Werkzeugen zur Datenverarbeitung nicht bewältigt werden können und oft erweiterte Analyseverfahren und Machine Learning erfordern.
Chatbot	Ein Computerprogramm, das eine Unterhaltung mit menschlichen Nutzer:innen simulieren soll, insbesondere über das Internet.
Cloud-Computing	Die Bereitstellung von Rechendiensten über das Internet, inklusive Speicherplatz, Rechenleistung und KI-Funktionen.
Computer Vision	Der Bereich der künstlichen Intelligenz, der Maschinen befähigt, Entscheidungen auf der Grundlage visueller Daten zu interpretieren und zu treffen. Häufig beinhaltet dies eine Bild- und Videoanalyse.
Datenwissenschaften	Der interdisziplinäre Bereich, der wissenschaftliche Methoden, Prozesse, Algorithmen und Systeme einsetzt, um Erkenntnisse und Wissen aus strukturierten und unstrukturierten Daten zu gewinnen.
Deep Learning	Eine Art von Machine Learning, bei der neuronale Netze mit vielen Schichten (tiefe neuronale Netze) zum Einsatz kommen, die es dem System erlauben, selbständig zu lernen und Entscheidungen zu treffen.
Digitaler Zwilling	Ein digitaler Zwilling ist eine virtuelle Version eines physischen Objekts, Prozesses oder Standorts und dient als digitales Gegenstück in Echtzeit.
Edge-Computing	Verarbeitung von Daten in der Nähe der Quelle der Datenerzeugung (z. B. IoT-Geräte), anstatt sich auf einen zentralen Cloud-Server zu verlassen.
Expertensystem	Ein Expertensystem ist eine Software, die KI einsetzt, um Probleme zu lösen und das Urteil menschlicher Expert:innen zu simulieren.
Generative KI	Generative KI, auch “GenAI” genannt, wird verwendet, um neue Inhalte (wie Text, Audio, Video, Bilder und sogar Code) auf der Grundlage von Trainingsdaten zu erstellen.

GPU	GPU (Graphics Processing Unit, Grafikverarbeitungseinheit oder Grafikprozessor) ist eine Computerkomponente, die ursprünglich für das Rendern von Bildern und Grafiken entwickelt wurde. Die Fähigkeiten eines Grafikprozessors zur parallelen Verarbeitung befähigen ihn, komplexe Probleme in Ströme von einzelnen Aufgaben aufzuteilen und diese parallel zu bearbeiten, sodass sie alle auf einmal abgeschlossen werden. Während GPUs ursprünglich zur Beschleunigung der Grafikverarbeitung entwickelt wurden, entdeckten Ingenieur:innen bald, dass GPUs auch die mathematischen Herausforderungen vieler anderer Aufgaben, einschließlich der KI, bewältigen können.
Handlungsempfehlende Wartung	Die handlungsempfehlende Wartung geht noch einen Schritt weiter als die prädiktive Wartung. Sie nutzt nicht nur Daten zur Vorhersage von Ausfällen, sondern schreibt auch die effektivsten Wartungsmaßnahmen vor.
Human-in-the-loop Machine Learning	Human-in-the-Loop Machine Learning, manchmal auch HITL genannt, ist ein Prozess, bei dem Menschen Eingaben in ein Machine-Learning-Modell liefern. So wird die Genauigkeit maximiert und die Effizienz von Anwendungen, die künstliche Intelligenz (AI) nutzen, wird gesteigert.
Industrie-Computer	Ein Industrie-PC ist ein Rugged-Computer, der für den Einsatz in einer industriellen Umgebung konzipiert ist, häufig für Anwendungen wie die Herstellung von Waren, die Gebäudeautomation und Logistikzentren. Ein Industrie-Computer weist in der Regel folgende Merkmale auf: lüfterlose und belüftungsfreie Konstruktion, Widerstandsfähigkeit gegenüber rauen Umgebungsbedingungen, hohe Konfigurierbarkeit, umfangreiche I/O-Optionen und ein langer Lebenszyklus.
Inferenzierung/Inferenz	Bei der KI-Inferenzierung wird ein trainiertes Modell verwendet, um Vorhersagen zu treffen und die Daten in verwertbare Erkenntnisse umzuwandeln.
Internet of Things (IoT)	Das Netzwerk miteinander verbundener Geräte, die miteinander kommunizieren und Daten austauschen und so dazu beitragen, große Mengen an Echtzeitinformationen für KI-Anwendungen zu sammeln.
Künstliche Intelligenz (KI)	Die Simulation menschlicher Intelligenzprozesse durch Maschinen, insbesondere durch Computersysteme, um Aufgaben zu erfüllen, die normalerweise menschliche Intelligenz erfordern.
Künstliche allgemeine Intelligenz (AGI)	Allgemeine KI (AGI) ist die Fähigkeit von Maschinen, zu denken, zu verstehen, zu lernen und Intelligenz anzuwenden, um komplexe Probleme zu lösen. So kann die AGI eine Lösung finden, wenn sie mit einer unbekannten Aufgabe konfrontiert wird.
Künstliche Intelligenz der Dinge (AIoT)	Das AIoT kombiniert KI mit dem IoT, um vernetzte Geräte in die Lage zu versetzen, Informationen zu verarbeiten und zu lernen.

Lernen durch Verstärkung	Eine Art von Machine Learning, bei der ein Akteur lernt, Entscheidungen zu treffen, indem er mit seiner Umgebung interagiert und Rückmeldungen in Form von Belohnungen oder Bestrafungen erhält.
Machine Learning (ML)	Eine Untergruppe der KI, die Systemen die Fähigkeit verleiht, automatisch zu lernen und sich aufgrund von Erfahrungen zu verbessern, ohne explizit programmiert zu werden.
Machine Vision	Machine Vision umfasst die Technologie, die Methoden, die Software und die Hardware, die bei der Verarbeitung visueller Eingaben zum Einsatz kommen.
Natural Language Processing (NLP)	Ein Bereich der künstlichen Intelligenz, der sich auf die Interaktion zwischen Computern und Menschen durch natürliche Sprache konzentriert und Maschinen in die Lage versetzt, menschliche Sprache zu verstehen, zu interpretieren und zu erzeugen.
Neuronales Netzwerk	Ein Computermodell, das sich an der Struktur und Funktionsweise des menschlichen Gehirns orientiert und aus miteinander verbundenen Knoten (Neuronen) besteht, die in Schichten organisiert sind.
NPU	Eine Neural Processing Unit (NPU) wurde speziell dafür entwickelt, die Ausführung neuronaler Netzwerke für Deep Learning zu beschleunigen. NPUs bieten hohe Leistung bei minimalem Stromverbrauch. Daher werden NPUs häufig in mobilen Geräten, Edge-Computing und anderen energiesensiblen Anwendungen eingesetzt.
Open Source	Der Begriff “Open Source” bezieht sich auf etwas, das von allen verändert und weitergegeben werden kann, weil das Design öffentlich zugänglich ist. Open-Source-Software ist beispielsweise Software mit Quellcode, den jede:r einsehen, verändern und verbessern kann.
Prädiktive KI	Prädiktive KI erstellt auf der Grundlage historischer Daten Vorhersagen über zukünftige Ereignisse, Verhaltensweisen und Trends.
Robotik	Die Robotik ist ein Bereich der Ingenieurwissenschaften, in dem künstliche Intelligenz eingesetzt wird, um Maschinen bei der Navigation und Steuerung ihrer Umgebung zu unterstützen.
Rugged-Computer	Sowohl Industrie- als auch Rugged-Computer bieten maximale Zuverlässigkeit und hohe Performance, um leistungsstarke Software auszuführen und komplexe Anwendungen zu steuern (siehe Definition des Begriffs Industrie-Computer.) Rugged-Computer verfügen über zusätzliche Design-Merkmale, die sie extrem zuverlässig und betriebsfähig in besonders rauen Umgebungen machen.
Spezialisierte oder schwache künstliche Intelligenz (ANI)	Schwache KI ist nur in der Lage, eine bestimmte, einzelne oder konzentrierte Aufgabe auszuführen.

Transfer Learning	Eine Technik des Machine Learning, bei der ein für eine Aufgabe trainiertes Modell für eine andere, verwandte Aufgabe angepasst wird, wobei das aus dem ursprünglichen Training gewonnene Wissen genutzt wird.
Trainingsdaten	Der Datensatz, der zum Trainieren eines KI-Modells verwendet wird. Er liefert Beispiele und Muster, aus denen das System lernen kann.
Tensor Processing Unit (TPU)	Eine Tensor Processing Unit (TPU) ist eine anwendungsspezifische integrierte Schaltung (ASIC), die von Google entwickelt wurde. Die Hauptaufgabe von TPUs ist die mathematische Matrixverarbeitung, die Machine-Learning-Workloads beschleunigt. In den Worten von Google: "TPUs können keine Textverarbeitungsprogramme ausführen, Raketentriebwerke steuern oder Banktransaktionen ausführen, aber umfangreiche Matrixoperationen in neuronalen Netzwerken mit hoher Geschwindigkeit verarbeiten."
Unüberwachtes Lernen	Eine Art von Machine Learning, bei der der Algorithmus unmarkierte Daten erhält und selbständig Muster und Beziehungen finden muss.
Überwachtes Lernen	Eine Art von Machine Learning, bei der der Algorithmus anhand eines markierten Datensatzes mit Input-Output-Paaren trainiert wird, um die Mapping-Funktion zu erlernen.
Vorausschauende Wartung	Ein Modell für die vorausschauende Wartung verwendet Daten aus einer Vielzahl von Quellen, um vorherzusagen, wann eine Anlage wahrscheinlich gewartet werden muss.