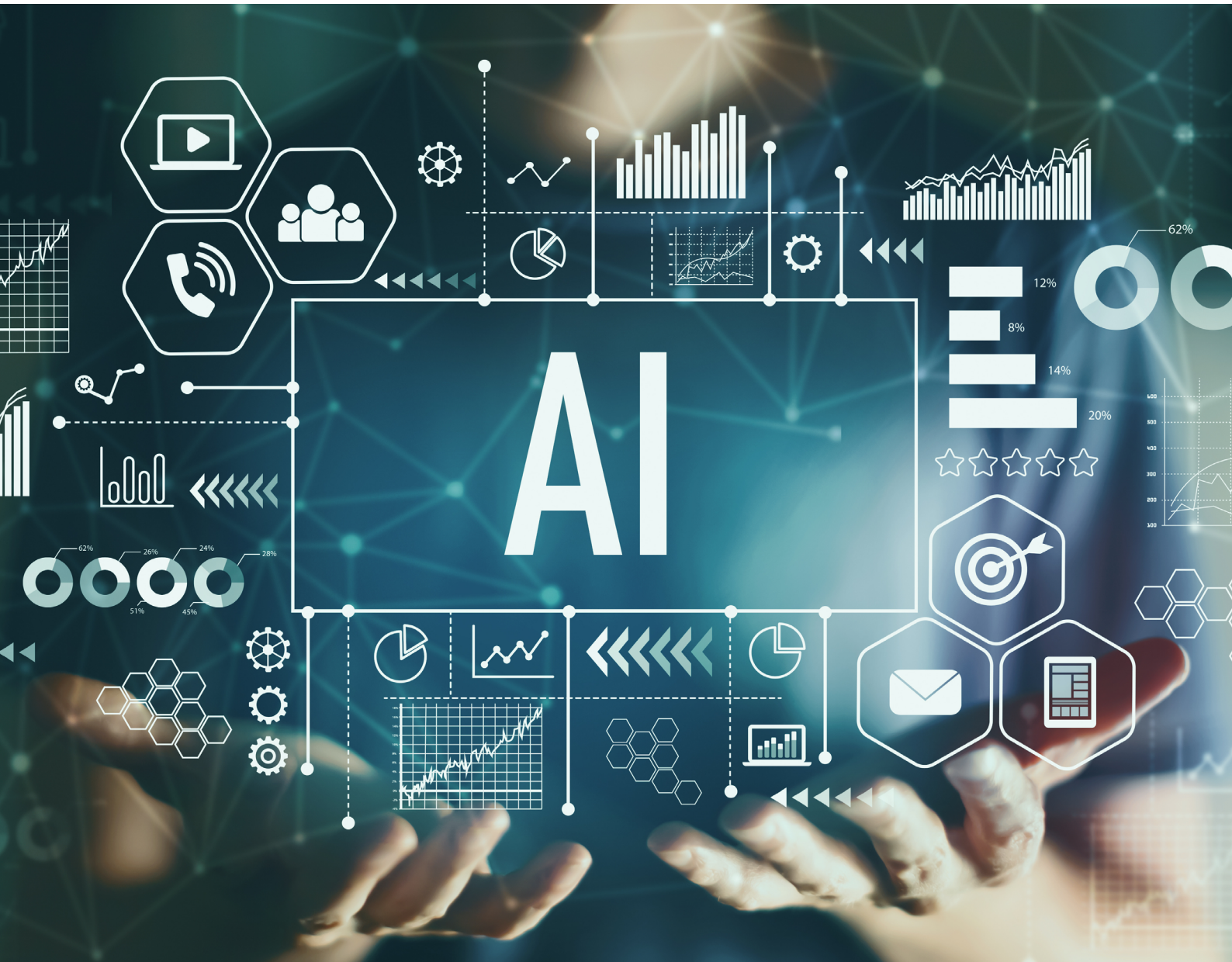
ONLOGIC

Alles over AI

Alles wat je altijd al wilde weten over AI, maar niet aan ChatGPT durfde te vragen



ONLOGIC

Inhoudsopgave

Introductie 03
Wat is AI? 03
Hoe werkt AI? 04
- AI-training 04
- AI-inferentie 04

AI-toepassingen 05
- Machine Learning (ML) 05
- Deep Learning 05
- Natural language processing (NLP) 05
- Expert system 05
- Robotica 05
- Machine Vision 06

Voordelen van AI in productie automatisering 06
- Langere levensduur van apparatuur 06
- Minder uitval 06
- Veiligere werkomgevingen 06
- Verbeterde productiviteit 06
- Beter beheer van voorraad en vraagprognose 07
- Gestroomlijnde fabrieks layout 07

AI aan de Edge 07
- Voorbeeld van AI aan de edge 07

Technologie voor AI-oplossingen 07
- GPU versus CPU 09
- Waarvoor wordt een GPU gebruikt? 09

Soorten GPU's 10
- Geïntegreerde GPU (iGPU) 10
- Zelfstandige GPU 10
- TPU 10
- NPU 10

Industriële computers en AI 10
Overwegingen voor een GPU in een industriële computer 11
- iGPU voor industriële AI 11
- Zelfstandige GPU voor industriële AI 11

Aan de slag met een AI-oplossing - een stapsgewijze gids 12
Hardware voor AI-oplossingen 13
Belangrijke conclusies 15
Woordenlijst van AI-termen 16

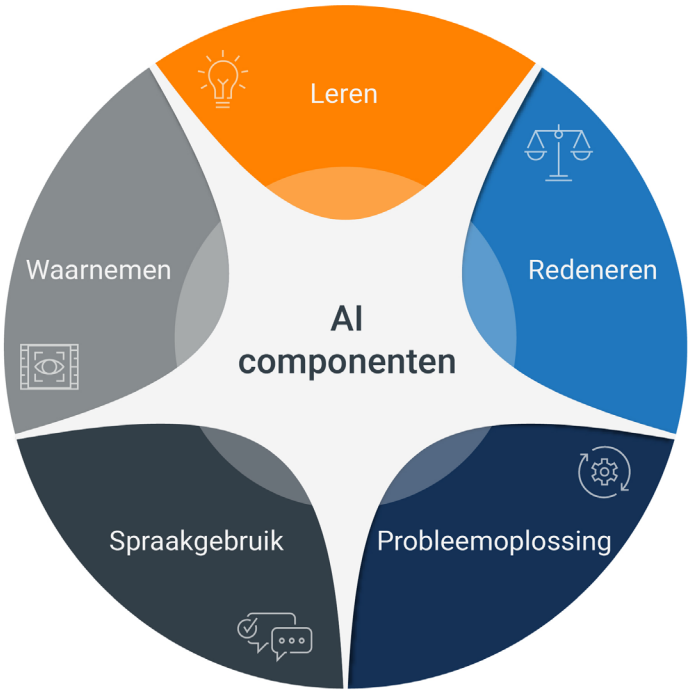
Alles over AI

AI heeft in no time de hoogste Buzz Word-status gekregen. Onder meer omdat het elke industrie werkelijk radicaal kan veranderen. Als je nog niet precies weet wat AI is, wat het kan doen, of hoe het je leven of je bedrijf kan beïnvloeden, ben je hier op de juiste plaats.

Lees verder voor een spoedcursus in AI en inzicht in de elementen en mogelijkheden die AI zo krachtig maken. We geven je ook een beknopt overzicht van de voordelen en mogelijkheden van AI en wat je nu moet doen om je voor te bereiden op de implementatie van AI-oplossingen, inclusief een stap-voor-stap gids. Tot slot hebben we een woordenlijst van AI-termen toegevoegd om je wegwijs te maken in de wereld van AI.

Wat is AI?

AI, afkorting van Artificial Intelligence, richt zich op technologieën die proberen de resultaten of output van menselijke intelligentie te repliceren. Daarvoor moet AI zich een weg banen door complexe informatieverwerking, waaronder: leren, redeneren, probleemoplossing, taalgebruik en perceptie.



Hoe werkt AI?

AI is niet slechts één computerprogramma of toepassing; het is een heel gebied van de informatica. Globaal gezien zijn er twee fasen van AI: training en inferentie.

AI-training

AI-training is het proces waarbij een AI-systeem getraind wordt om gegevens te identificeren, te verwerken en van gegevens te leren. Het doel van het trainen van het systeem is inferentie (deductie) mogelijk te maken zodra het systeem is geïmplementeerd. Tijdens de training wordt de input geanalyseerd zodat het model een bibliotheek van gestructureerde patronen in de gegevens kan opbouwen. Het is de bedoeling dat het AI-model geïnformeerde voorspellingen kan doen wanneer nieuwe gegevens worden verstrekt. Effectieve AI-training vereist enorme hoeveelheden gegevens en veel rekenkracht, vaak met behulp van multicoreprocessors en GPU's (hierover zo dadelijk meer).

Wat is een GPU?

GPU staat voor Graphics Processing Unit. Soms worden ze grafische kaarten of videokaarten genoemd. Zoals de naam al aangeeft, zijn GPU's oorspronkelijk ontworpen om afbeeldingen en video's te verwerken, zodat de CPU zich op andere taken kan richten.

GPU's zijn ontworpen voor parallelle verwerking. Dit stelt ze in staat om complexe problemen op te splitsen in stromen van afzonderlijke taken en deze parallel uit te voeren, zodat ze allemaal tegelijk klaar zijn. In plaats van één voor één zoals een CPU.

Het blijkt dat dit ontwerp geweldig is voor meer dan alleen het renderen van afbeeldingen en video's. De parallelle verwerkingsmogelijkheden hebben ervoor gezorgd dat ze zeer effectief zijn voor kunstmatige intelligentie (AI), machine learning, computer vision en nog veel meer. GPU's elimineren mogelijke knelpunten in de verwerking voor snellere resultaten, meer mogelijkheden en een verbeterde gebruikerservaring.

AI-inferentie

AI-inferentie is het proces waarbij nieuwe data in een getraind model wordt ingevoerd en de data wordt omgezet in bruikbare inzichten. AI-inferentie is waar de waarde van AI wordt gerealiseerd, aangezien de modellen die gebouwd worden slechts zo nuttig zijn als de output die wordt afgeleid van inferentie.

In tegenstelling tot het trainen van AI-modellen heeft inferentie niet altijd speciale GPU's nodig. Het model wordt immers alleen toegepast en gerefereerd, niet gebouwd. Dit is natuurlijk afhankelijk van de hoeveelheid nieuwe data die geanalyseerd moet worden.

AI-toepassingen

Het gebied van kunstmatige intelligentie is groot en wordt met de dag groter. Als je het toch wilt opsplitsen, zou je de toepassingen van AI kunnen onderverdelen in onder andere: Machine Learning, Deep Learning, NLP (natuurlijke taalverwerking), Expert Systems, Robotica en Machine Vision.

Machine Learning (ML)

Machine learning-modellen gebruiken gegevens en algoritmen om specifieke taken uit te voeren zonder expliciet te zijn geprogrammeerd. De nauwkeurigheid van deze algoritmen voor machine learning verbetert geleidelijk in de loop der tijd.

- Zo ontwikkelt Plus One Robotics machine learning-oplossingen voor warehousing en distributie.

In het algemeen zijn er drie grote subcategorieën van machine learning:

- Machine learning-algoritme onder toezicht
    - Datawetenschappers leveren gelabelde trainingsgegevens. Machines gebruiken de gegevens om uitkomsten te voorspellen.
- Machine learning-algoritme zonder toezicht
    - Algoritmen zoeken naar patronen in ongelabelde gegevens.
- Reinforcement machine learning
    - Het algoritme leert door interactie met zijn omgeving en het verbeteren van gedrag op basis van positieve of negatieve response op acties.

Deep Learning

Deep Learning is een soort machine learning dat algoritmen in lagen structureert om een 'kunstmatig neuraal netwerk' te creëren. Deep learning is een geavanceerdere benadering van machine learning en kan worden gebruikt om complexere problemen op te lossen.

- Bear Flag Robotics ontwikkelt bijvoorbeeld zelfrijdende autonome tractoren, gevoed door deep learning

Natural language processing (NLP)

NLP-algoritmen worden gebruikt om geschreven of gesproken menselijke taal te verwerken. Het wordt gebruikt voor vertaling, samenvatting of om een actie uit te voeren.

- Google Translate is een voorbeeld van NLP om tekst van de ene taal naar de andere te vertalen.

Expert system

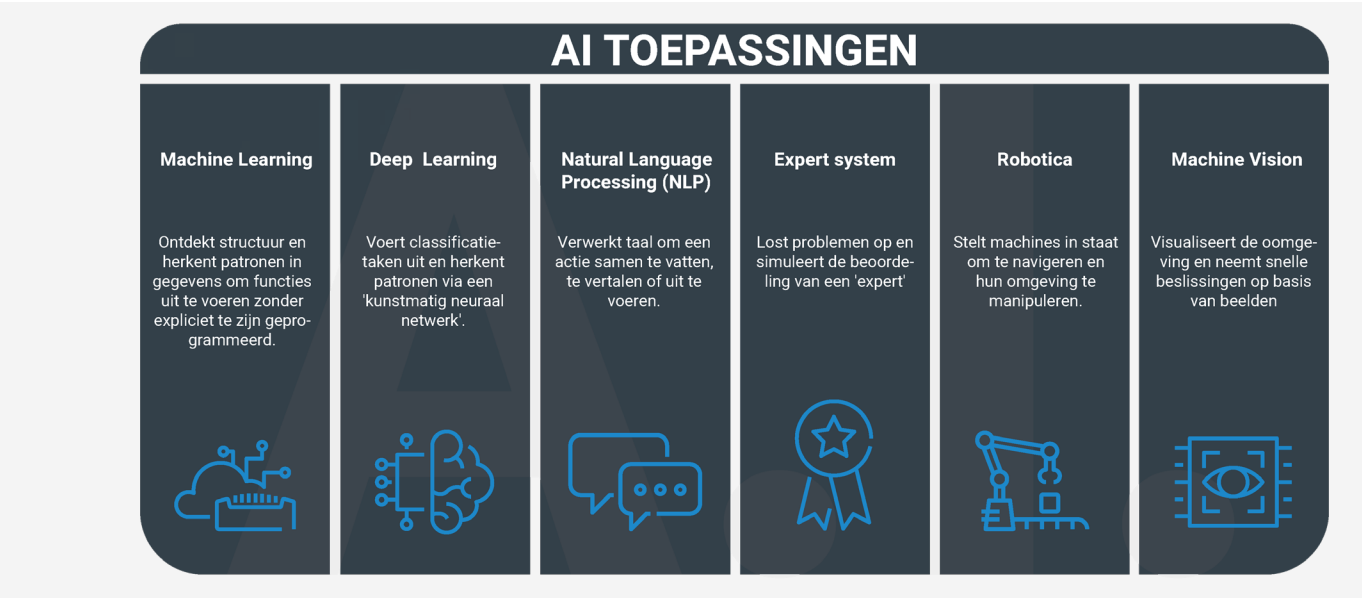
Een expert system is software die AI gebruikt om problemen op te lossen en het oordeel van een menselijke expert te simuleren.

- MYCIN, een programma dat in de jaren 70 van de vorige eeuw door Stanford is gemaakt, is een voorbeeld van een expert system. Het wordt gebruikt om artsen te helpen bij de diagnose en behandeling van infecties.

Robotica

Robotica is een onderdeel van engineering dat AI gebruikt om machines te helpen navigeren en hun omgeving te manipuleren.

- advanced.farm maakt bijvoorbeeld fruitoogstrobots die helpen bij het herkennen en plukken van rijpe groenten en fruit.



Machine Vision

Machine Vision gebruikt de nieuwste AI-technologieën om industriële apparatuur de mogelijkheid te geven om zijn omgeving te visualiseren en snelle beslissingen te nemen op basis van wat deze 'ziet'.

- Artemis Vision creëert bijvoorbeeld machine vision-oplossingen voor kwaliteitsinspectie in een productieomgeving om de kleinste details vast te leggen, die door het menselijk oog kunnen worden gemist.

Voordelen van AI in productie automatisering

AI heeft een groot aantal voordelen voor productie automatisering. Ze bieden mogelijkheden om kosten te verlagen en producten van hogere kwaliteit te produceren. Dit zijn volgens ons de zes grootste voordelen.

Langere levensduur van apparatuur

- Voorspellend onderhoud door AI maakt gebruik van sensorgegevens van machines om belangrijke kenmerken zoals bijvoorbeeld temperatuur, trillingen en druk te detecteren. Op basis daarvan kan worden bepaald wanneer onderhoud aan de apparatuur moet worden uitgevoerd. Dit kan ongeplande uitvaltijd van machines minimaliseren, onderhoudskosten verlagen en zelfs de levensduur van machines verlengen.

Minder uitval

- In een productielijn kunnen meerdere punten voor kwaliteitscontrole worden opgenomen die kwaliteitsproblemen kunnen registreren die door het menselijk oog worden gemist. Daardoor kan de nauwkeurigheid van de inspectie aanzienlijk worden verbeterd. Als defecten in een vroeg stadium van het productieproces worden ontdekt, kan de hoeveelheid afval worden verminderd.

Veiligere werkomgevingen

- AI-aangedreven oplossingen zoals kwaliteitscontrole kunnen worden geplaatst op locaties die mogelijk niet veilig zijn voor mensen, zoals direct op de productielijn.
- Robotoplossingen kunnen worden ontwikkeld om de moeilijkste en/of repetitieve taken op een productievloer uit te voeren.

Verbeterde productiviteit

- AI-gestuurde automatisering, bijvoorbeeld met behulp van robotica, kan het werk 24 uur per dag mogelijk maken om taken, zelfs repetitieve taken, uit te voeren met een hoge mate van nauwkeurigheid.

Beter beheer van voorraad en vraagprognose

- AI kan trends in voorraad en vraag identificeren om voorraadniveaus te optimaliseren, vraag te voorspellen en zelfs verzendingen in een complexe wereldwijde toeleveringsketen te volgen. Het kan er ook voor zorgen dat beslissingen over de productie gebaseerd worden op de meest recente en actuele gegevens.

Gestroomlijnde fabrieks layout

- AI kan inefficiënties in de productieworkflow identificeren, knelpunten verhelpen en de uitvoer op de werkvloer verhogen. Een manier om dit te doen is door een digital twin van de fabrieksvloer te maken. Aanpassingen kunnen eerst op de twin worden uitgeprobeerd om de resultaten te bekijken.

AI aan de edge

Om bijna real-time besluitvorming mogelijk te maken, halen veel bedrijven AI-oplossingen weg uit de cloud en verplaatsen ze deze naar de edge, dichter bij de bron van de systemen die de gegevens genereren, om zo realtime bruikbare inzichten te kunnen verstrekken.

Het bouwen van een AI-model en inferentie kunnen natuurlijk plaatsvinden in een gecentraliseerde cloud of externe datacenters. Maar door AI-mogelijkheden aan de rand van het netwerk te plaatsen, kun je sneller beslissingen nemen en handelen op basis van de informatie die van sensoren op locatie is verzameld. De edge, oftewel de rand van het netwerk, kan zich bevinden in een magazijn, op een productielijn, op een vorkheftruck of zelfs op een verafgelegen, externe locatie.

Voorbeeld van AI aan de edge

Autonome voertuigen, zoals de tractoren, gemaakt door Bear Flag Robotics, zijn een voorbeeld van AI aan de edge. Informatie van een groot aantal inputs, waaronder: 3D camera's, LiDAR- en positiesensoren worden verzameld en verwerkt om voertuigen te informeren over waar ze moeten keren om een oneffenheden in de bodem te omzeilen, hoe apparatuur moet worden afgesteld en wanneer ze moeten remmen om obstakels te vermijden.

Dit zijn allemaal beslissingen die snel moeten worden genomen. Er is geen tijd om de gegevens naar de cloud te sturen, te verwerken en vervolgens instructies naar de tractor terug te sturen. Doordat de gegevens worden ingelezen en verwerkt door een computer die op de tractor is bevestigd, in plaats van in de cloud, wordt deze vertraging, latentie genoemd, drastisch verminderd.

Om nog maar te zwijgen over het feit dat de toegang tot een internetverbinding best een uitdaging kan zijn op een landbouwterrein. Met AI aan de edge kan het computergebruik gewoon doorgaan, ongeacht waar die gegevens worden geproduceerd.

Technologie voor AI-oplossingen

De explosieve groei van AI en de daaruit voortvloeiende waarde zijn nauw verbonden met de explosie van beschikbare gegevens, open-source tools en de vooruitgang van de technologie om de verzamelde informatie te verwerken en daarop te reageren. Enkele van de belangrijkste verbeteringen zijn:

Grotere gegevenssets

- Het Internet of Things (IoT) biedt de mogelijkheid om gegevens te verzamelen van een breed scala aan verbonden apparaten en open gegevenssets.
- Data-centric AI is een opkomende wetenschap die zich richt op het verbeteren van datasets voor betere prestaties in praktische ML-toepassingen.

Wat is het internet of things (IoT)?

Het IoT verwijst naar de verzameling apparaten die met elkaar communiceren via het internet. Veel voorkomende IoT-hardware-apparaten zijn draagbare apparaten, slimme thuisapparaten en industriële sensoren die met de cloud verbonden zijn.

Wat is het AIoT?

Het AIoT combineert AI met IoT om verbonden apparaten in staat te stellen informatie te verwerken en te leren. Dit kan helpen om toekomstige patronen of gebeurtenissen te voorspellen waardoor beslissingen over hoe te reageren mogelijk worden.

Wat is open source?

De term open source verwijst naar iets dat mensen kunnen aanpassen en delen omdat het ontwerp openbaar toegankelijk is. Open source software is bijvoorbeeld software met broncode die iedereen kan inspecteren, wijzigen en verbeteren.

Voordelen van open source AI-tools zijn onder andere transparantie, samenwerking, snelle prototyping en gemeenschapsgerichte ontwikkeling. Het kan leiden tot snellere vooruitgang omdat ontwikkelaars van elkaar leren en voortbouwen op elkaars werk. Sommigen zien het ook als een manier om bronnen van vooroordelen te identificeren die zich in AI kunnen ontwikkelen en om een AI-monopolie van bedrijven te voorkomen.

Beschikbaarheid van open-source AI-tools

- Open-source AI-frameworks, toolkits en vooraf getrainde modellen hebben snelle AI adoptie mogelijk gemaakt. Ze bieden de bouwstenen voor het ontwerpen, trainen, valideren en implementeren van AI-oplossingen. Doordat handmatige codering en traditionele methoden voor softwareontwikkeling niet meer nodig zijn, kunnen deze toolkits de kosten aanzienlijk verlagen en de implementatietijd verkorten.
- Enkele voorbeelden van toolkits zijn OpenVino, TensorFlow, Apache, Hugging Face en PyTorch.

Verbeteringen van de processor

- Multicore CPU's, geïntegreerde GPU's en discrete GPU's (Graphical Processing Units) zijn de basis voor de stijgende waarde van AI. Ze leveren de rekenkracht om gegevenssets te verwerken en te interpreteren om een AI-algoritme te bouwen.

Wat is een CPU?

CPU staat voor Central Processing Unit en wordt vaak het brein van de computer genoemd. Simpel gezegd is de CPU een elektronisch circuit dat gegevens regelt, inclusief invoer, uitvoer en opslag. De CPU zorgt ervoor dat het besturingssysteem van de computer en de applicaties allemaal werken.

GPU versus CPU

We hebben de definities van een GPU en CPU geschetst, maar wat is het verschil? Zowel GPU als CPU zijn microprocessors op basis van silicium die gegevens verwerken. De verschillen liggen in de manier waarop ze gegevens verwerken en hun fysieke verwerkingseenheden – hun cores.

Gegevensverwerking

- CPU's voeren computeropdrachten serieel uit. Ze verwerken gegevens van de ene taak na de andere.
- GPU's voeren computeropdrachten parallel uit. Ze verwerken veel berekeningen tegelijk.

De cores

- CPU's hebben een handvol cores.
    - Intel® 13e generatie Hybride-core processors zijn bijvoorbeeld beschikbaar met maximaal 24 cores.
    - Intel heeft hybride-core technologie geïntroduceerd met 12e generatie. Meer daarover kun je hier lezen.
- GPU's kunnen 100 tot 1000 cores hebben. Ze lijken op CPU-cores, maar zijn kleiner.
    - Zo biedt Nvidia's RTX A4000 GPU bijvoorbeeld meer dan 6300 cores, waaronder: 48 tweede generatie RT-cores, 192 derde generatie Tensor-cores en 6144 CUDA-cores.

Waarvoor wordt een GPU gebruikt?

Zowel CPU's als GPU's kunnen echt goed rekenen. Maar zoals hierboven vermeld, verwerken CPU's de vergelijkingen taak voor taak. Een GPU kan de vergelijkingen daarentegen allemaal tegelijk verwerken. Daarom bieden GPU's buitengewoon rekenvermogen voor complexe wiskundige berekeningen.

De weergave van afbeeldingen en grafieken komt neer op een reeks wiskundige problemen. Zoals al gezegd, blinken GPU's daarin uit. In feite versnellen ze deze workloads voor superieure grafische prestaties. Daarom worden ze 'versnellers' genoemd.

Hoewel GPU's oorspronkelijk zijn ontworpen om de grafische verwerking te versnellen, ontdekten engineers al snel dat GPU's de geavanceerde uitdagingen van kunstmatige intelligentie aankonden. Ze konden bijvoorbeeld veel gegevens opnemen, classificeren en deze informatie gebruiken om gevolgtrekkingen en voorspellingen te maken voor toekomstige resultaten.

Soorten GPU's

Er zijn twee primaire typen GPU's: geïntegreerde en zelfstandige GPU's

Geïntegreerde GPU (iGPU)

Een geïntegreerde GPU, ook wel iGPU genoemd, is rechtstreeks ingebouwd in de processor of CPU van de computer. Deze deelt voeding en systeemgeheugen met de CPU. Dat betekent dat computers met geïntegreerde GPU's energiezuinig zijn en over het algemeen lichter en kleiner. Omdat het geheugen echter wordt gedeeld, kunnen zware GPU-verwerkingstaken vertraging veroorzaken bij algemene CPU-verwerking.

Zelfstandige GPU

Een zelfstandige of speciale GPU is een grafische processor die volledig los staat van de CPU en een eigen geheugen heeft, ook wel videogeheugen of VRAM genoemd. Dankzij dit eigen geheugen hoeft deze GPU niet te vertrouwen op het RAM van de computer. Afzonderlijke grafische kaarten leveren betere prestaties, maar verbruiken meer vermogen en genereren daardoor veel warmte.

De koeling van een zelfstandige GPU is een grote ontwerpoverweging om maximale prestaties te bereiken. Daarom hebben ingenieurs deze GPU's ontworpen met eigen koelventilatoren. Deze GPU's zijn daardoor groter en verbruiken meer stroom.

Lees onze Wat is een GPU? Een gids voor beginners voor meer informatie

TPU

Een Tensor Processing Unit (TPU) is een toepassingsspecifiek geïntegreerd circuit (ASIC) ontwikkeld door Google. De primaire taak voor TPU's is wiskundige matrixverwerking die machine learning workloads versnelt. In de woorden van Google: "TPU's kunnen geen tekstverwerkers aansturen, raketmotoren besturen of banktransacties uitvoeren, maar ze kunnen enorme matrixbewerkingen die worden gebruikt in neurale netwerken met hoge snelheden verwerken."

NPU

Een Neural Processing Unit (NPU) is speciaal gebouwd om de uitvoering van neurale netwerken voor deep learning te versnellen. NPU's leveren hoge prestaties terwijl het stroomverbruik tot een minimum wordt beperkt. Daarom worden NPU's vaak gebruikt in mobiele apparaten, edge computing en andere energiegevoelige toepassingen.

Industriële computers en AI

Industriële computers kunnen worden uitgerust met GPU's voor AI en andere deep learning-toepassingen. Ingenieurs hebben deze computers zorgvuldig ontworpen zodat ze stand houden in extreme omgevingen. Ze hebben bedrijven in staat gesteld om AI op vrijwel elke gewenste plaats toe te passen.

Ze worden bijvoorbeeld gebruikt voor het aansturen van robots in productiefaciliteiten, magazijnen en zelfs in de landbouw waar ze kunnen worden blootgesteld aan trillingen en extreme temperaturen. Ze worden ook gebruikt voor het aansturen van kwaliteitsinspecties, soms direct op een productielijn waar de omgeving niet veilig is voor mensen, laat staan voor een traditionele desktopcomputer.



Overwegingen voor een GPU in een industriële computer

Industriële computers zijn ontworpen zodat ze kunnen functioneren in extreme omgevingen. Een gemeenschappelijk ontwerpkenmerk voor industriële computers is hun fanless en geheel afgesloten ontwerp. Hoewel het veel voordelen heeft, werkt dit ontwerp niet goed voor alle GPU's. Sommige GPU's hebben namelijk een aanzienlijk hoger energieverbruik dan CPU's. Er moet dus veel meer warmte worden afgevoerd.

iGPU voor industriële AI

Het gebruik en de koeling van geïntegreerde grafische kaarten die een eenvoudige oplossing voor beeld- en objectherkenning mogelijk maken, is volledig mogelijk met een fanless ontwerp. Vooral als dit wordt gedaan met een aantal van de nieuwste en ongelooflijk krachtige processors.

Aparte GPU's zijn niet altijd nodig en kunnen de kosten van een computer aanzienlijk opdrijven. Door AI-inferentie uit te voeren op een CPU of geïntegreerde GPU (iGPU) kun je de kosten van een effectieve AI-implementatie verlagen.

Als het gaat om industriële computers voor AI, zorgt een geheel afgesloten ontwerp voor een klein formaat. Dit biedt meer flexibiliteit bij de installatie. Aangezien de computer geen ventilator nodig heeft om lucht in het chassis te brengen, kunnen deze fanless computers worden geïnstalleerd op plekken met veel stof of vuil in de lucht en/of op plekken die onderhevig zijn aan trillingen.

Zelfstandige GPU voor industriële AI

Voor een complexere AI- of ML-oplossing is misschien een afzonderlijke of speciale grafische kaart vereist. Zoals hierboven vermeld, genereren GPU's veel warmte. Dat betekent dat de koeling van een GPU een belangrijke overweging is. De meeste losse GPU's worden geconfigureerd met hun eigen ventilatoren.

Eén optie is een hybride fanless computer met een geïsoleerd uitbreidingsslot. Een GPU kan in dit slot worden geïnstalleerd en actief worden gekoeld met een ventilator. Het moederbord en andere gevoelige interne onderdelen worden beschermd in het geheel afgesloten gedeelte van het chassis. Als je de GPU moet vervangen, is deze gemakkelijk toegankelijk.

Aan de slag met een AI-oplossing - een stapsgewijze gids

Oké, je hebt nu bepaald dat de implementatie van AI je bedrijf ten goede kan komen en je bent klaar om aan de slag te gaan. Maar waar begin je? Natuurlijk is het verleidelijk om direct op zoek te gaan naar de juiste hardware en software, maar er zijn een paar stappen die je echt eerst moet overwegen.

1. Analyseer je behoeften

- Identificeer de gebieden in het bedrijf die kunnen profiteren van automatisering. Enkele veelvoorkomende voorbeelden zijn:
  - Repetitieve taken
  - Kwaliteitscontrole
  - Voorspellend onderhoud
  - Supply chain management.

2. Definieer duidelijke doelstellingen

- Stel SMART-doelen (Specific, Measurable, Achievable, Relevant en Time bound) vast voor de op AI gebaseerde automatisering. Of het nu gaat om de verbetering van de efficiëntie, kostenbesparing of de verbetering van de productkwaliteit, duidelijk gedefinieerde doelstellingen zullen je AI-implementatie sturen en zullen je helpen te beoordelen wat werkt en wat misschien niet.

3. Stel een team samen

- Stel een team samen met vertegenwoordigers uit de hele organisatie, met name Operations en IT. Gezamenlijke inspanningen zorgen ervoor dat de implementatie in overeenstemming is met zowel de technische eisen als de productiebehoeften.

4. Beoordeel de gegevensinfrastructuur

- Evalueer de beschikbaarheid en kwaliteit van je gegevens. AI werkt met grote datasets voor trainingsmodellen. Begin met het verzamelen van gegevens uit alle onderdelen van de productieprocessen.

5. Werk samen met leveranciers

- Overleg met leveranciers en consultants die gespecialiseerd zijn in automatisering. Misschien bieden ze inzichten, begeleiding en mogelijk aangepaste oplossingen die zijn afgestemd op jouw specifieke behoeften. Elke AI-implementatie is anders, dus het is belangrijk om samen te werken met leveranciers die de moeite doen om de unieke uitdagingen van elk project te begrijpen.

6. Ontwerp een strategie voor continuous improvement

- Wat wordt beoordeeld:
  - Doet het AI-model zijn werk goed genoeg?
  - Begint het dingen te missen?
  - Moet het opnieuw worden getraind met nieuwe gegevens?
  - Hoe ga je na verloop van tijd nieuwe modellen ontwikkelen?

Hardware voor AI-oplossingen

Als het gaat om hardwareopties voor AI, zien we dat verschillende toepassingen verschillende AI-oplossingen vereisen.

IoT-gateways

Of een AI-oplossing nu in de cloud of op een server op locatie wordt gebruikt, je hebt altijd een IoT-gateway nodig. Deze kleine maar betrouwbare computer is de schakel tussen gegevens die worden verzameld door geïntegreerde sensoren en de cloud. Ze spelen een steeds belangrijkere rol in AI-oplossingen voor het verzamelen, opslaan en soms gedeeltelijk verwerken van binnenkomende gegevens voordat deze worden verzonden.

Zo hebben we de Karbon 410 ontworpen voor betrouwbaarheid onder zelfs de meest uitdagende installatieomstandigheden, waaronder extreme temperaturen en trillingsgevoelige locaties.



AI aan de edge met een geïntegreerde GPU

Een aantal van de nieuwste processors met hun geïntegreerde GPU kunnen eenvoudig AI-inferentieoplossingen aan de edge ondersteunen. De fanless Helix 511 van OnLogic bijvoorbeeld wordt aangedreven door Intel 12e generatie processors met hybride-core architectuur en DDR5-geheugen. Deze compacte krachtige computer biedt een overvloed aan I/O, connectiviteit voor legacy apparatuur en krachtige verwerking voor een geavanceerde AI-oplossing.



AI aan de edge met zelfstandige GPU

Wil je een robuustere oplossing implementeren, aangedreven door 12e of 13e generatie Intel Core™-processors met zelfstandige GPU? De Karbon 804, een fanless hybride pc, biedt high-performance computing en heeft opties voor GPU-uitbreidingskaarten met PCIe gen 4.0 mogelijkheden. De ongelooflijke doorvoer die PCIe Gen 4 biedt, betekent dat je opties hebt als het gaat om de keuze van de slots en lanes die je voor je GPU wilt gebruiken.

- Optie 1: Een enkele PCIe-riser met x16 slots en x16 lanes voor een doorvoer van 256GT/s.
- Optie 2: Een dubbele riser met 2 x16 slots met elk x8 lanes. Als je één slot voor de GPU gebruikt, heb je een doorvoer van 128GT/s. Voor veel AI-toepassingen is die doorvoer meer dan genoeg. Bovendien heeft de computer dan een slot over. Die kun je naar behoefte configureren voor extra uitbreidingsopties.





Heb je een kleiner alternatief nodig?

Onze Karbon 803 heeft een iets kleiner formaat met één PCIe gen 4 x16 slot.



GPU-server

Voor complexe workloads, deep learning aan de edge en training en inferentie op locatie, is een GPU-server zoals de AC101 een geweldige oplossing. Dit platform biedt Intel 13e generatie processors en geavanceerde GPU's en DDR5-geheugen. Veel bedrijven gebruiken edge servers in hun strategieën voor cloudrepatriëring, waarbij ze computerresources naar de edge verplaatsen om latentie te voorkomen en de operationele kosten te verlagen.

Belangrijke conclusies

Zoals onze voorbeelden hebben aangetoond, transformeert AI bijna elke bedrijfstak. AI verandert onze manier van leven, werken en spelen. Nu is het tijd om de infrastructuur van je technologie te versterken om beter gepositioneerd te kunnen profiteren van de voordelen van AI. Denk hierbij aan verbeteringen in productiviteit, efficiëntie, kwaliteit en klanttevredenheid.

Het begint allemaal met de volgende 5 stappen:

1. Analyseer je behoeften
2. Definieer duidelijke doelstellingen
3. Stel een team samen
4. Beoordeel de gegevensinfrastructuur
5. Werk samen met leveranciers

Van daaruit kun je een AI-project starten en stapsgewijs schalen. Begin klein en bouw verder op je succes. Verbeter je AI-modellen en processen voortdurend. Het OnLogic-team staat klaar om je te helpen je hardwarebehoeften te begrijpen en met je mee te denken over hoe een industriële computer aan je uitdagende eisen kan voldoen. De toekomst is nu en het gebruik van AI is niet langer een kwestie van "of" maar van "wanneer". Wij helpen je graag op weg.

Woordenlijst van AI-termen

Algemene AI (AGI)	Algemene AI (AGI) is het vermogen van machines om na te denken, te begrijpen, te leren en intelligentie toe te passen om complexe problemen op te lossen, zodat de AGI bij een onbekende taak een oplossing kan vinden.
Algoritme	Software die een reeks instructies of regels bevat die door een computer worden gevolgd om een specifieke taak uit te voeren of een bepaald probleem op te lossen.
Algoritme bias	Onbedoelde en systematische fouten in een algoritme die leiden tot oneerlijke of discriminerende resultaten, vaak als gevolg van bestaande vooroordelen in de trainingsgegevens.
Algoritmische transparantie	De mate waarin de innerlijke werking van een algoritme open en begrijpelijk is, vooral in toepassingen waar beslissingen invloed hebben op individuen of de samenleving.
Artificial Intelligence of Things (AIoT)	De AIoT combineert AI met IoT om verbonden apparaten in staat te stellen informatie te verwerken en te leren.
Big data	Extreem grote en complexe datasets die niet verwerkt kunnen worden door traditionele gegevensverwerkingstools. Hiervoor zijn vaak geavanceerde analyses en machine learning nodig.
Chatbot	Een computerprogramma dat is ontworpen om gesprekken met menselijke gebruikers te simuleren, vooral via internet.
Cloud computing	De levering van computerdiensten, waaronder opslag, verwerkingskracht en AI-mogelijkheden, via internet.
Computer Vision	Het gebied van AI dat machines in staat stelt om visuele gegevens te interpreteren en op basis van visuele gegevens beslissingen te nemen, vaak met beeld- en videoanalyse.
Datawetenschap	Het interdisciplinaire veld dat wetenschappelijke methoden, processen, algoritmen en systemen gebruikt om inzichten en kennis te extraheren uit gestructureerde en ongestructureerde data.
Deep Learning	Een soort machine learning dat neurale netwerken met vele lagen (diepe neurale netwerken) omvat, waardoor het systeem zelf in staat is om te leren en beslissingen te nemen.
Digital twin	Een digital twin is een virtuele versie van een fysiek object, proces of locatie die dient als een real-time digitale tegenhanger.
Edge computing	Het verwerken van gegevens in de buurt van de plek waar de gegevens worden gegenereerd (bijv. IoT-apparaten) in plaats van te vertrouwen op een gecentraliseerde cloudserver.
Expert System	Een expert system is software die AI gebruikt om problemen op te lossen en het oordeel van een menselijke expert te simuleren.
Generatieve AI	Generatieve AI, ook wel GenAI genoemd, wordt gebruikt om nieuwe inhoud (zoals tekst, audio, video, beelden en zelfs code) te creëren op basis van trainingsgegevens.

GPU	GPU (Graphics Processing Unit) is een computercomponent die oorspronkelijk is ontworpen om afbeeldingen en grafische elementen weer te geven. De parallelle verwerkingsmogelijkheden van een GPU maken het mogelijk om complexe problemen op te splitsen in stromen van afzonderlijke taken en ze parallel uit te werken, zodat ze allemaal tegelijk worden voltooid. Hoewel GPU's oorspronkelijk zijn ontworpen om de grafische verwerking te versnellen, ontdekten ingenieurs al snel dat GPU's de wiskundige uitdagingen van vele andere taken, waaronder AI, aankonden.
Human-in-the-loop machine leren	Human-in-the-loop machine learning, soms HITL genoemd, is een proces waarbij mensen input leveren aan een machine learning model om de nauwkeurigheid te maximaliseren en de efficiëntie te verhogen voor toepassingen die kunstmatige intelligentie (AI) gebruiken.
Industriële computer	Een industriële pc is een robuuste computer die is ontworpen voor gebruik in een industriële omgeving, vaak voor toepassingen zoals de productie van goederen, gebouwautomatisering en logistieke centra. Een industriële computer heeft doorgaans de volgende kenmerken: fanless en geheel gesloten ontwerp, bestand tegen zware omgevingsomstandigheden, zeer configureerbaar, uitgebreide I/O-opties en een lange levensduur.
Inferentie	AI-inferentie is het proces waarbij het getrainde model wordt gebruikt om voorspellingen te doen en de data om te zetten in bruikbare inzichten.
Internet of Things (IoT)	Het netwerk van onderling verbonden apparaten die communiceren en gegevens uitwisselen. Dit bevordert de verzameling van enorme hoeveelheden real-time informatie voor AI-toepassingen.
Kunstmatige intelligentie (AI)	De simulatie van menselijke-intelligentieprocessen door machines, met name computersystemen, om taken uit te voeren die meestal menselijke intelligentie vereisen.
Machine Learning (ML)	Een subset van AI die systemen de mogelijkheid biedt om automatisch te leren en te verbeteren op basis van ervaring, zonder expliciet geprogrammeerd te moeten worden.
Machine Vision	Machine vision omvat de technologie, methoden, software en hardware die betrokken zijn bij de verwerking van visuele input,
Natuurlijke taalverwerking (NLP)	Een gebied van AI dat zich richt op de interactie tussen computers en mensen door middel van natuurlijke taal, waardoor machines menselijke taal kunnen begrijpen, interpreteren en genereren.
Neuraal netwerk	Een rekenmodel geïnspireerd door de structuur en werking van het menselijk brein, bestaande uit onderling verbonden knooppunten (neuronen) georganiseerd in lagen.

Open Source	De term open source verwijst naar iets dat mensen kunnen wijzigen en delen omdat het ontwerp openbaar toegankelijk is. Open-source software is bijvoorbeeld software met broncode die iedereen kan bekijken, wijzigen en verbeteren
Reinforcement Learning	Een soort machine learning waarbij een agent leert beslissingen te nemen door een wisselwerking met zijn omgeving en feedback te ontvangen in de vorm van beloningen of boetes.
Robotica	Robotica is een onderdeel van engineering dat AI gebruikt om machines te helpen navigeren en hun omgeving te manipuleren.
Rugged computer	Zowel industriële als rugged computers bieden maximale betrouwbaarheid en hoge prestaties voor de uitvoering van krachtige software en de besturing van complexe toepassingen. (Zie definitie van industriële computer.) Rugged computers hebben extra ontwerpfuncties waardoor ze uiterst betrouwbaar en bruikbaar zijn in bijzonder veeleisende omgevingsomstandigheden.
Supervised Learning	Een type machine learning waarbij het algoritme is getraind in een gelabelde gegevensset, met ingang/uitgang-paren om de toewijzingsfunctie te leren.
Trainingsgegevens	De gegevensset die wordt gebruikt om een AI-model te trainen, met voorbeelden en patronen waar het systeem van kan leren.
Voorspellende AI	Voorspellende AI genereert voorspellingen over toekomstige gebeurtenissen, gedragingen en trends op basis van historische gegevens.
Voorspellend onderhoud	Een model voor voorspellend onderhoud gebruikt gegevens uit een groot aantal bronnen om te voorspellen wanneer apparatuur waarschijnlijk onderhoud nodig heeft.
Voorschrijvend onderhoud	Voorschrijvend onderhoud brengt voorspellend onderhoud een stap verder. Naast het gebruik van gegevens om storingen te voorspellen, worden ook het meest effectieve onderhoud voorgeschreven.
Zwakke AI (ANI)	Zwakke AI is alleen in staat om een specifieke, enkelvoudige of gerichte taak uit te voeren.