

# Pioniers van kunstmatige intelligentie met oog voor natuurkunde

## De Nobelprijs voor de Natuurkunde 2024

Kunstmatige neurale netwerken zoals ChatGPT helpen ons enorm in het dagelijks leven en wetenschappelijke toepassingen. De winnaars van de Nobelprijs voor de Natuurkunde 2024, John Hopfield en Geoffrey Hinton, benutten de analogie met spinmodellen in de statistische fysica voor een doorbraak in kunstmatige intelligentie.

**K**unstmatige neurale netwerken hebben zich ontwikkeld tot een veelzijdig en krachtig hulpmiddel voor zowel dagelijkse als geavanceerde wetenschappelijke toepassingen. De basis van deze netwerken bestaat uit een verzameling 'neuronen' of knopen, verbonden door 'synapsen' of gewogen verbindingen, die moeten worden getraind om specifieke taken uit te voeren. Deze structuur vertoont sterke parallellen met spinmodellen in de statistische fysica, zoals die worden toegepast in de studie van magnetisme. De koppelingen tussen de neuronen zijn vergelijkbaar met de wisselwerkingen tussen spins en kunnen specifieke patronen versterken of onderdrukken. Dit leidt tot een dynamisch en collectief gedrag dat sterk overeenkomt met dat van fysische systemen. John Hopfield (Chicago 1933) en Geoffrey Hinton (Wimbledon 1947) (figuur 1) gebruikten deze analogie voor een doorbraak in kunstmatige intelligentie en kregen daarvoor de Nobelprijs voor de Natuurkunde 2024.

### Historische achtergrond

De eerste elektronische computers, ontwikkeld in de jaren veertig van de vorige eeuw, waren bedoeld om berekeningen uit te voeren die moeilijk en tijdrovend waren. Vanaf de jaren vijftig ontstond er echter een nieuwe behoefte: het gebruik van computers voor patroonherkenning, een taak waarin mensen en andere dieren uitblinken. Deze behoefte leidde tot de ontwikkeling van kunstmatige intelligentie (AI) en in het bijzonder tot kunstmatige neurale netwerken. In 1943 stelden de neuroloog Warren McCulloch en de logicus Walter Pitts een model voor dat beschreef hoe neuronene in de hersenen samenwerken [1]. In dit model berekent een neuron een gewogen som van inkomende binaire (0 of 1) signalen, wat vervolgens een binaire uitvoer oplevert. Dit werk legde de basis voor verder onderzoek naar zowel biologische als kunstmatige neurale netwerken. Even later introduceerde de psycholoog Donald Hebb een mechanisme voor leren en geheugen, waarbij de herhaalde activatie van twee neuronene leidde tot

een versterking van de synaptische verbinding tussen hen: "Cells that fire together, wire together" [2].

In 1957 introduceerde Frank Rosenblatt zijn *perceptron* (figuur 2a), een eenvoudig neuraal netwerkmodel gebaseerd op het werk van McCulloch en Pitts dat echter, ondanks aanzienlijke belangstelling van neurowetenschappers, wiskundigen en computerwetenschappers, niet in staat bleek bepaalde fundamentele logische bewerkingen te kunnen uitvoeren zoals de exclusieve OR of XOR (de XOR-operatie geeft 'waar' als precies één van de invoerwaarden waar is, maar 'onwaar' als beide invoerwaarden hetzelfde zijn) [3]. Dit schijnbare onvermogen leidde tot de eerste AI-winter, een periode waarin de interesse in neurale netwerken sterk afnam. Hoewel het oorspronkelijke perceptron van Rosenblatt de XOR-operatie niet kon simuleren, bleken geavanceerdere netwerken, de zogenoemde *multi-layer perceptrons* (MLP's), dankzij hun meerdere verwerkingslagen wel in staat om de XOR-operatie uit te voeren en daarnaast een breed scala aan complexere



Figuur 1. Links: John Hopfield geeft college over neurale netwerken aan de universiteit of Illinois Urbana-Champaign in 1991. Foto: Department of Physics, University of Illinois at Urbana-Champaign, courtesy AIP Emilio Segrè Visual Archives. Rechts: Geoffrey Hinton achter de schermen tijdens de Collision 2024-conferentie. Foto: Vaughn Ridley/Collision via Sportsfile CC BY 2.0.

problemen op te lossen. Deze netwerken vormen de basis van zogenaemde *feed-forward* neurale netwerken. Hierbij stroomt de informatie van de invoer naar de uitvoer zonder terugkoppeling, wat hen krachtig maakt voor taken zoals patroonherkenning en classificatie. In figuur 2b wordt het concept van een *feed-forward*-netwerk uitgelegd, waarin elke laag de informatie van de vorige laag verwerkt.

Tegelijkertijd waren andere onderzoekers bezig modellen te ontwikkelen die geïnspireerd waren door magnetische systemen voor het beschrijven van zogenaemde terugkoppelende neurale netwerken (*recurrent neural networks*, RNN's). Dit is een type neuraal netwerk waarbij verbindingen terugkoppelen, zodat informatie uit eerdere stappen kan worden gebruikt om toekomstige beslissingen te beïnvloeden, ideaal voor het verwerken van sequentiële data.

### De doorbraak van John Hopfield

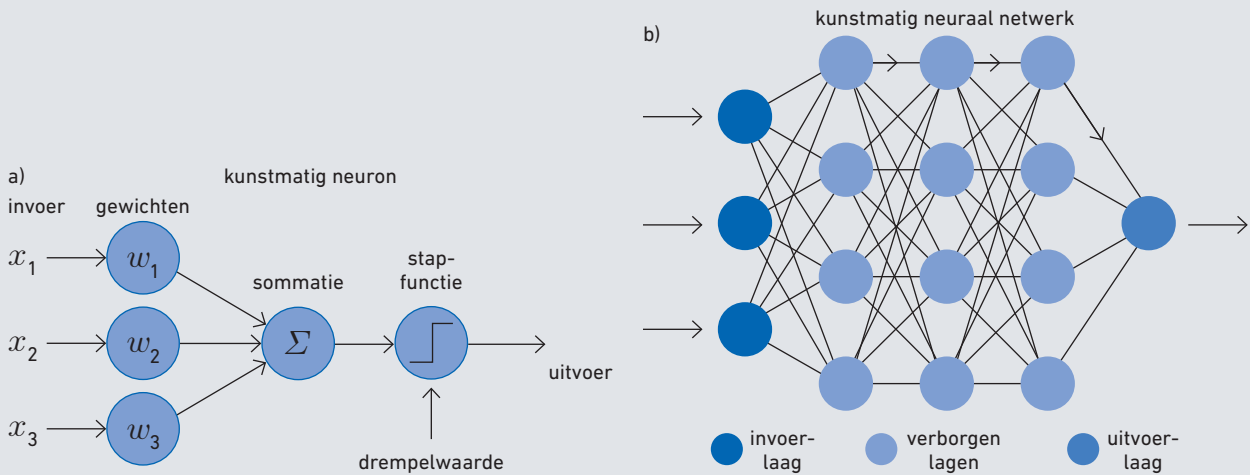
John Hopfield, theoretisch natuurkundige, leverde een beslissende bijdrage door in 1982 een dynamisch model voor associatief geheugen te introduceren, gebaseerd op een eenvoudig terugkop-

pelend neuraal netwerk [4]. Hij vroeg zich af of er *emergente* collectieve verschijnselen in grote verzamelingen neuronon konden optreden die rekenkundige mogelijkheden boden. Hopfields model – het Hopfield-netwerk – bevat  $N$  binaire neuronon die asynchroon worden geactualiseerd afhankelijk van de toestanden van de andere neuronon en de respectievelijke onderlinge koppelingen (zie figuur 3). De dynamiek van het model wordt bepaald door een energiefunctie, waarin opgeslagen herinneringen (zoals afbeeldingen) overeenkomen met energieminima.

Het Hopfield-netwerk is een model voor een associatief geheugen en kan patrooncorrectie uitvoeren. Zo kan het model patronen aanvullen of corrigeren (figuur 4), een functie die later toepasbaar bleek in allerlei andere AI-toepassingen. Het Hopfield-netwerk is direct geïnspireerd op het Ising-model dat in de natuurkunde wordt gebruikt om het gedrag van magnetische systemen te bestuderen, in het bijzonder de wisselwerking tussen spins in een rooster. Hopfield toonde dus aan hoe fysische principes toegepast konden worden bij de ontwikkeling van neurale netwerken.



**Wilfred van der Wiel** is hoogleraar nanoelektronica en co-directeur van het Center for Brain-Inspired Nano Systems (BRAINS) aan de Universiteit Twente. Zijn onderzoek richt zich op het gebruik van complexe fysische systemen om kunstmatige intelligentie efficiënter te maken. [w.g.vanderwiel@utwente.nl](mailto:w.g.vanderwiel@utwente.nl)



Figuur 2. Schematische weergave van een kunstmatig neuron en een kunstmatig neuraal netwerk. a) In een kunstmatig neuron, of *perceptron*, worden de invoerwaarden  $x_1, x_2, x_3, \dots$  na vermenigvuldiging met de gewichten  $w_1, w_2, w_3, \dots$  gesommeerd. Deze som vormt de invoer voor een sterk niet-lineaire stapfunctie met een instelbare drempelwaarde. De waarde van de stapfunctie dient als invoer voor kunstmatige neuronen in de volgende laag. b) Voorbeeld van een kunstmatig neuraal netwerk of *multi-layer perceptron* opgebouwd uit kunstmatige neuronen (gekleurde cirkels). In dit *feed-forward*-netwerk stroomt de informatie van links naar rechts. De zwarte lijnen stellen de gewichten  $w$  voor die getraind dienen te worden om de gewenste functionaliteit te realiseren.

## Geoffrey Hinton: de Boltzmann-machine

In de jaren 1983-1985 ontwikkelden Geoffrey Hinton en collega's een uitbreiding van Hopfields model, de Boltzmann-machine [5,6]. De naam refereert aan het feit dat de toestand van Hinton's neurale netwerk niet deterministisch is, zoals in het geval van het Hopfield-netwerk, maar beschreven wordt door een waarschijnlijkheid volgens de Boltzmann-verdeling (figuur 5). Ludwig Boltzmann (1844-1906) was een Oostenrijkse natuurkundige die de basis legde voor de statistische mechanica, waarmee hij het verband aantoonde tussen de microscopische bewegingen van deeltjes en de macroscopische thermodynamische eigenschappen van materie. Met Hopfield en Hinton deed deze statistische aanpak ook zijn intrede in de wereld van neurale netwerken. Door gebruik te maken van waarschijnlijkheden kan de Boltzmann-machine complexe patronen en relaties leren. Hoewel de oorspronkelijke Boltzmann-machine vanwege zijn rekenkundige complexiteit beperkt bruikbaar bleek, ontwikkelde Hinton later de *restricted Boltzmann-machine* (RBM), die veel efficiënter is en nieuwe toepassingen mogelijk maakt [7].

## Van backpropagation tot deep learning

De doorbraken van Hopfield en Hinton leidden tot succesvolle toepassingen in patroonherkenning, waaronder beeld- en spraakherkenning. In 1986 introduceerden David Rumelhart, Geoffrey Hinton en Ronald Williams het *backpropagation*-algoritme, waarmee *feed-forward*-netwerken met verborgen lagen getraind konden worden [8]. Dit algoritme maakte het mogelijk om neurale netwerken nauwkeuriger af te stemmen op complexe taken. Het werk van Hinton en zijn team in de jaren 2000 resulteerde in de implementatie van 'diepe' neurale netwerken. Door de *restricted Boltzmann-machine* als voortraject te gebruiken, worden de kunstmatige neuronen laag voor laag getraind, wat bijdraagt aan een significant verbeterde prestatie van neurale netwerken. Dit legde de basis voor wat tegenwoordig bekend staat als *deep learning*, dat nu in een breed scala aan toepassingen wordt gebruikt.

## Neurale netwerken als krachtig hulpmiddel

Naast hun bijdrage aan het theoretische fundament van kunstmatige neurale netwerken hebben Hopfield en Hinton methoden ontwikkeld die wor-

den toegepast in verschillende wetenschappelijke disciplines, waaronder de natuurkunde, scheikunde en biologie. Neurale netwerken worden gebruikt om de energieën van materiaalfasen en de sterkte van interatomaire krachten te voorspellen, met een nauwkeurigheid die vergelijkbaar is met die van quantummechanische modellen. Dit maakt het mogelijk om de stabiliteit en dynamiek van nieuwe materialen te bestuderen. Het Nobelcomité refereerde in zijn toelichting bijvoorbeeld aan de studie aan faseovergangen door de Twentse wetenschapper Menno Bokdam en zijn collega's [9]. Kunstmatige neurale netwerken worden ook in de deeltjesfysica ingezet voor gegevensanalyse. In experimenten die hebben geleid tot de ontdekking van het higgsdeeltje bij CERN heeft het gebruik van neurale netwerken de gevoeligheid van de analyses verhoogd. Ook in de astrofysica worden neurale netwerken gebruikt voor het analyseren van gegevens uit grote detectors, zoals de IceCube-neutrino-detector.

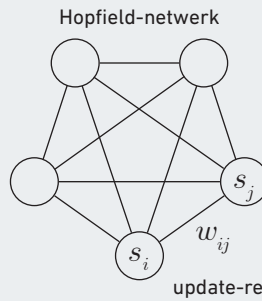
In het dagelijks leven zijn toepassingen van AI alomtegenwoordig, van beeldherkenning op smartphones tot taalgeneratie in chatbots en vertaaltoepassingen. Binnen de gezondheids-

zorg worden kunstmatige neurale netwerken ingezet als hulpmiddel bij diagnoses, bijvoorbeeld bij de detectie van borstkanker in mammogrammen of bewegingscorrectie in MRI-scans. Deze toepassingen laten zien hoe de methoden van Hopfield en Hinton hun weg hebben gevonden in alledaagse technologieën.

### Uitdagingen voor de toekomst

Het snel toenemende gebruik van AI komt echter ook met enorme uitdagingen. Het trainen van het *large language model* GPT-4 (bekend van ChatGPT) kostte maar liefst  $2,4 \cdot 10^{25}$  FLOPS (*floating point operations*), wat overeenkomt met het continu laten draaien van 10.000 GPU's (*graphic processing units*) gedurende zes maanden [10]. AI gebruikt nu al net zo veel energie als een klein land en het einde is nog lang niet in zicht [11]. Het ziet er naar uit dat informatie- en communicatietechnologie al in 2030 meer dan 20% van de totale wereldelektriciteitsproductie zal opslokken [12]. En mogelijk zal in 2040 de wereldenergieproductie niet meer toereikend zijn [13]. Dit vraagt om fundamentele oplossingen voor hoe we AI toepassen, om slimmere, efficiëntere algoritmes en vooral om veel efficiëntere hardware. Aangezien kunstmatige neurale netwerken geïnspireerd zijn op biologische neurale netwerken en gebruikt worden voor taken waarin de hersenen uitblinken, ligt het voor de hand naar het brein te kijken voor inspiratie voor nieuwe hardware. Het brein is een enorm krachtige computer en verbruikt slechts 20 watt, ordes van grootte minder dan supercomputers. Een sterk opkomende tak van wetenschap die zich hiermee bezighoudt is *brain-inspired* of *neuromorphic computing* met een sterke vertegenwoordiging in Nederland [14].

En net zoals Hopfield en Hinton hun inspiratie uit de natuurkunde halen voor het ontwikkelen van hun revolutionaire modellen, kunnen we ook naar de natuurkunde kijken voor het ontwikkelen van nieuwe, efficiëntere hardware voor diezelfde soort modellen. Het ligt immers voor de hand dat complexe fysische systemen van nature geschikt zijn om modellen



- neuron  $s_i$  is of 'aan' (1) of 'uit' (-1)
- elk neuron  $s_i$  is verbonden met alle andere neuronen  $s_j$  via  $w_{ij}$
- elk neuron  $s_j$  heeft een drempelwaarde  $\theta_i$
- de toestand van het netwerk is gegeven door 'energie'  $E$

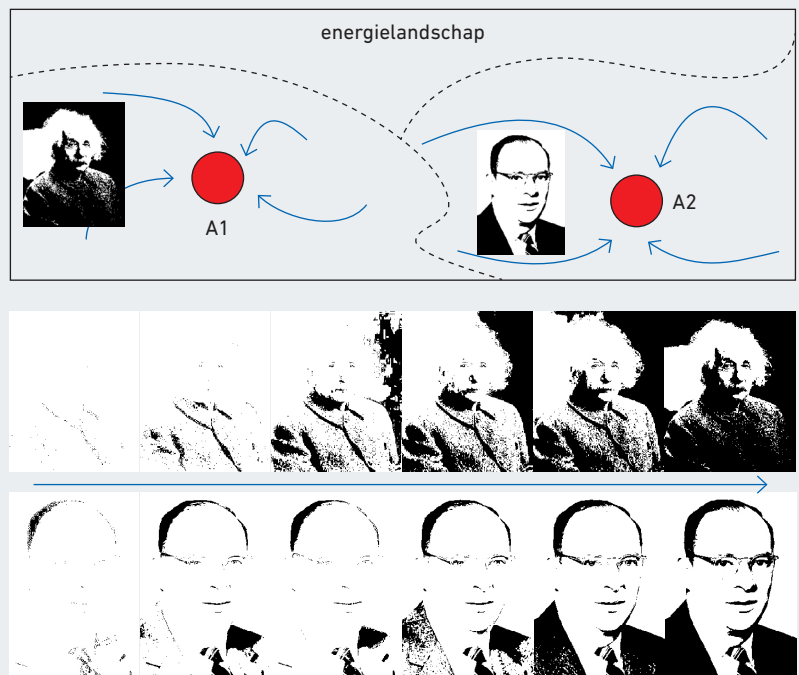
update-regel

$$s_i \leftarrow \begin{cases} +1 & \text{als } \sum_j w_{ij}s_j \geq \theta_i, \\ -1 & \text{anders.} \end{cases}$$

'energie'

$$E = -\frac{1}{2} \sum_{i,j} w_{ij}s_i s_j - \sum_i \theta_i s_i$$

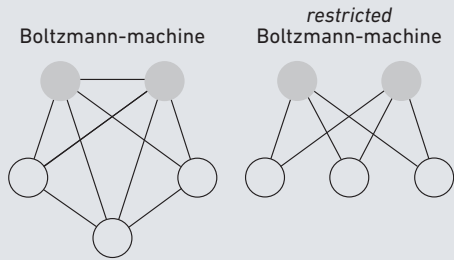
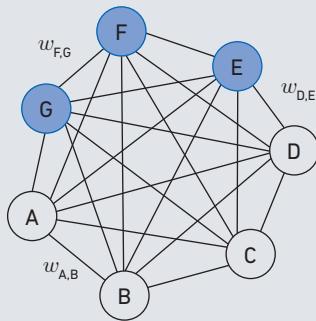
Figuur 3. Het Hopfield-netwerk is een vorm van een terugkoppelend (*recurrent*) neuraal netwerk, waarbij informatie teruggekoppeld kan worden in het systeem. In dit type netwerk kunnen patronen of beelden worden opgeslagen en gereconstrueerd, wat het een model voor associatief geheugen maakt. Elk neuron  $s_i$  in het netwerk heeft een toestand die 'aan' (1) of 'uit' (-1) kan zijn, en is verbonden met alle andere neuronen  $s_j$  via een symmetrisch gewicht  $w_{ij}$ . Daarnaast heeft elk neuron een drempelwaarde  $\theta_i$  die bepaalt of het neuron actief wordt of niet. De toestand van het netwerk wordt bepaald door de 'energie'  $E$ , waarbij het netwerk streeft naar een energiminimum, vergelijkbaar met hoe spijnsystemen in de fysica zich gedragen. (De koppelingsparameters  $w_{ij}$  zijn niet onafhankelijk, zoals bij het Sherrington-Kirkpatrick-model voor spinglazen, maar worden geconstrueerd op basis van patronen.)



Figuur 4. Hopfield-netwerk als associatief geheugen. De rode punten A1 en A2 zijn attractoren die corresponderen met verschillende opgeslagen patronen (foto's van Albert Einstein en John Bardeen). Binnen dit energielandschap trekken de attractoren de toestanden van het netwerk aan totdat dit convergeert naar één van de opgeslagen beelden. De blauwe pijlen geven de dynamische evolutie aan van een initiële, verstoorde invoer naar de dichtstbijzijnde attractor. De onderstaande panelen laten de opeenvolgende reconstructie van een vervaagde afbeelding zien, waarbij het netwerk stapsgewijs convergeert naar een volledig gereconstrueerd patroon. Afbeelding met dank aan A.A. Khajetoorians.

te implementeren die op diezelfde complexe fysische systemen geënt zijn.

Zo hebben Alexander Khajetoorians en collega's aan de Radboud Universiteit



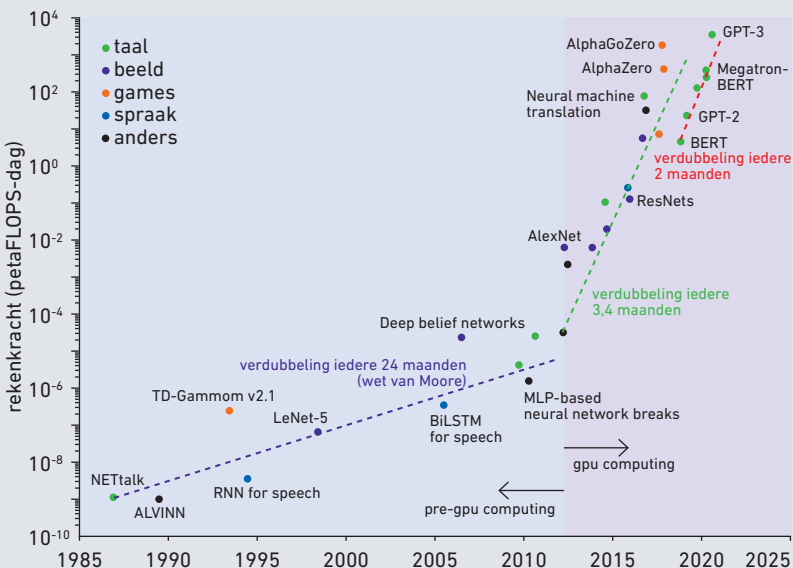
$$E = - \left( \sum_{i < j} w_{ij} s_i s_j + \sum_i \theta_i s_i \right)$$



Ludwig Boltzmann (1844-1906)

- $s_i$  is de toestand  $\{0,1\}$  van unit  $i$
- elk neuron  $s_i$  is verbonden met alle andere neuronen  $s_j$  via gewichten  $w_{ij}$
- elk neuron  $s_i$  heeft een drempelwaarde  $\theta_i$
- toestand  $s$  van het netwerk is gegeven door 'energie'  $E$
- waarschijnlijkheid van netwerktoestand  $P(s) = e^{-E/T}$
- $T$  is analoog aan de temperatuur

Figuur 5. De Boltzmann-machine is een stochastische uitbreiding van het Hopfield-model dat gebruikmaakt van de Boltzmann-verdeling. In tegenstelling tot het Hopfield-netwerk, dat een deterministisch, reproductief model is dat opgeslagen patronen ophaalt, fungeert de Boltzmann-machine als een generatief model dat nieuwe data kan creëren en patronen en structuren kan ontdekken. Dit komt doordat de Boltzmann-machine gebruikmaakt van probabilistische neuronen in plaats van deterministische neuronen, en beschikt over verborgen neuronen (blauw) die het model flexibeler maken voor complexere problemen. De *restricted Boltzmann-machine* (RBM) is een vereenvoudigde versie die eenvoudiger te trainen is dankzij technieken zoals contrastieve divergentie (ontwikkeld door Hinton), wat uiteindelijk de ontwikkeling van *deep learning* mogelijk heeft gemaakt.



Figuur 6. AI als energieslurper. De toename in de vraag naar rekenkracht over de afgelopen vier decennia, uitgedrukt in petaFLOPS-dagen. Tot 2012 verdubbelde de vraag naar rekenkracht elke 24 maanden, min of meer gelijke tred houdend met de bekende wet van Moore; recentelijk is dit verkort tot ongeveer elke twee maanden. PetaFLOPS staat voor een biljard ( $10^{15}$ ) *floating point*-bewerkingen per seconde (FLOPS). Deze meeteenheid beschrijft de snelheid van supercomputers en andere krachtige computersystemen. PetaFLOPS-dag betekent het aantal operaties dat in een dag wordt uitgevoerd bij een rekenkracht van 1 petaFLOPS ( $= 86,4 \cdot 10^{18}$  operaties, [17]).

een atomaire Boltzmann-machine generaliseerd door het manipuleren van enkele kobaltatomen [15]. En in ons

eigen Center for Brain-Inspired Nano Systems (BRAINS) aan de Universiteit Twente wordt gebruikgemaakt van de

unieke elektronische eigenschappen van wanordelijke atomaire netwerken in silicium voor het realiseren van kleine neurale netwerken voor taken als beeld- en spraakherkenning [16]. Dit soort systemen is niet alleen uitermate compact, maar kan ook een bijdrage leveren aan het oplossen van de enorme energiehonger van AI. De toekenning van de Nobelprijs aan John Hopfield en Geoffrey Hinton erkent hun visionaire werk, dat niet alleen de wetenschap heeft verrijkt, maar ook de weg heeft vrijgemaakt voor toekomstige innovaties die de samenleving ten goede kunnen komen. Nieuwe, efficiëntere algoritmes en hardware bieden een kans om de energieconsumptie van AI te verlagen en tegelijk de mogelijkheden van deze technologieën verder uit te breiden.

REFERENTIES EN BRONNEN

- 1 W.S. McCulloch en W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.* **5**, 115 (1943).
- 2 D.O. Hebb, *The organization of behavior* (1949).
- 3 F. Rosenblatt, *Principles of neurodynamics: Perceptrons and theory of brain mechanisms* Spartan Book (1962).
- 4 J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- 5 S.E. Fahlman, G.E. Hinton en T.J. Sejnowski. Massively Parallel Architectures for AI: NETL, Thistle, and Boltzmann Machines I. *Proceedings of the AAAI-83 conference*, 109-113 (1983).
- 6 D.H. Ackley, G.E. Hinton en T.J. Sejnowski, A learning algorithm for Boltzmann machines, *Cogn. Sci.* **9**, 147-169 (1985).
- 7 G.E. Hinton, Training products of experts by minimizing contrastive divergence, *Neural Comput.* **14**, 1771 (2002).
- 8 D.E. Rumelhart, G.E. Hinton en R.J. Williams, Learning representations by back-propagating errors, *Nature* **323**, 533-536 (1986).
- 9 R. Jinnouchi et al., Phase Transitions of Hybrid Perovskites Simulated by Machine-Learning Force Fields Trained on the Fly with Bayesian Inference, *Phys. Rev. Lett.* **122**, 225701 (2019).
- 10 Stephen M. Walker II, Everything We Know About GPT-4, <https://klu.ai/blog/gpt-4-llm> (2023).
- 11 Brian Calvert, AI already uses as much energy as a small country. It's only the beginning, *Vox* (2024).
- 12 N. Jones, How to stop data centers from gobbling up the world's electricity, *Nature* **561**, 163-166 (2018).
- 13 *Decadal plan for Semiconductors - Full report*, Semiconductor Industry Association (SIA) & Semiconductor Research Corporation (SRC) (januari 2021).
- 14 M. Zolfagharijnejad et al., Brain-inspired computing systems: a systematic literature review, *Eur. Phys. J. B* **97**, 70 (2024).
- 15 B. Kiraly et al., An atomic Boltzmann machine capable of self-adaption, *Nat. Nanotechnol.* **16**, 414-420 (2021).
- 16 T. Chen et al., Classification with a disordered dopant-atom network in silicon, *Nature* **577**, 341-345 (2020).
- 17 A. Mehonic en A. J. Kenyon, Brain-inspired computing needs a master plan, *Nature* **604**, 255-260 (2022).
- 18 [www.nobelprize.org/prizes/physics/2024/advanced-information](http://www.nobelprize.org/prizes/physics/2024/advanced-information).
- 19 Hopfield network en Boltzmann machine, zie [en.wikipedia.org](http://en.wikipedia.org).