

05

REKEN MEE

Een echte
parameterupdate

10

ARCHITECTUUR

Attention
van binnenuit

29

WETENSCHAP

Neurale netwerken
ontmoeten fysica

■ AI ONTLEED

Wat een model werkelijk leert

Van trainingsdata tot toetsbare voorspelling.



De kaart van deze editie

VECTOR onderzoekt lerende modellen: training, attention, kansen en onafhankelijke metingen. De fysica maakt aannamen concreet.

BEGRIPPEN

Zakwoordenboek

Acht begrippen die steeds terugkomen.

Parameter Instelbaar getal in een model; training past het aan.

Verliesfunctie Vat de afwijking tussen voorspelling en uitkomst samen in één getal.

Gradiënt Vector van partiële afgeleiden; wijst naar de steilste stijging.

Backpropagation Rekent met de kettingregel laag voor laag terug en vindt zo de gradiënt voor elke parameter.

Kalibratie Kansscores kloppen: van alle 80%-voorspellingen komt ongeveer 80% uit.

Randvoorwaarde Eis aan een oplossing op de rand of bij de start, zoals $u(0) = 1$.

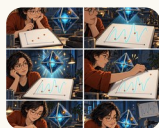
Aleatorische onzekerheid Variatie die als toeval overblijft, zoals meetruis.

Epistemische onzekerheid Ontbrekende kennis, ook over de modelkeuze; neemt af met meer gegevens of een beter model.

Bronnen: 2, 10, 11, 52

LEESAFSpraak

Elk getal krijgt een herkomst: gemeten, berekend (theorie met gemeten constanten), rekenvoorbeeld (zelfgekozen getallen) of model (een bewuste vereenvoudiging).



3-9

Leren en proberen

Strip op pagina 3; leren op pagina 4-5; sensorquiz en puzzels op pagina 6-7; weetjes op pagina 9.



10-15

Taal en positie

Attention op pagina 10-11; sensorfusie op pagina 12-13; denkproeven op pagina 14; leerstap op pagina 15.



16-21

Kansen en getuigen

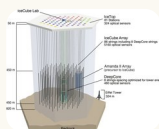
Kansscores op pagina 16-17; nieuwe strip en meetweetjes op pagina 18-19; methodecheck op pagina 21.



22-27

Werelden en foutbanden

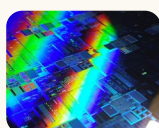
Wereldmodel en quantum op pagina 22-25; onzekerheidsquiz op pagina 26; halve-pagina-meetbegroting op pagina 27.



28-33

Klokken en werkplaats

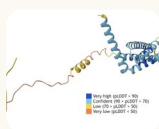
Golven en Nobel op pagina 28-29; lichtklok en GPS op pagina 30-31; puzzels en logboek op pagina 32-33.



34-39

Bits en kopieën

Rekenen en onderzoek op pagina 34-35; Shannon en Landauer op pagina 36-37; demoquiz en gedachten op pagina 38-39.



40-46

Toetsbare wetenschap

AlphaFold op pagina 40; modellen op pagina 41-43; OpenAI-casus op pagina 44; voetbal en papers op pagina 45-46.



47-50

Antwoorden en bronnen

Bestaande uitwerkingen op pagina 47; nieuwe puzzels en denkproeven op pagina 48; nieuwe quizen op pagina 49; bronnen en vervolg op pagina 50.

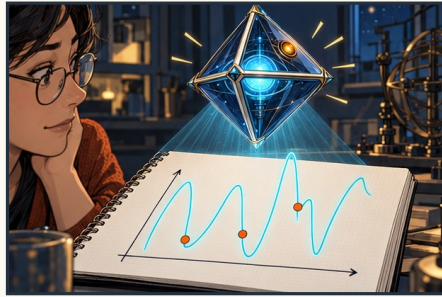
DE LABSTRIP · AFLEVERING 01

AXI en de perfecte voorspeller

Drie meetpunten. Eén schitterende curve. AXI is tevreden – tot de onderzoeker nog een meting doet.



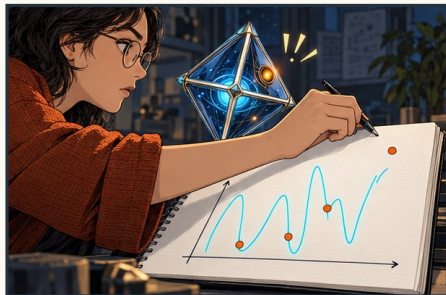
1 Onderzoeker Drie meetpunten. Wat kun je daarmee voorspellen?



2 AXI Kijk! Mijn curve raakt ze alle drie. Geen enkele fout!



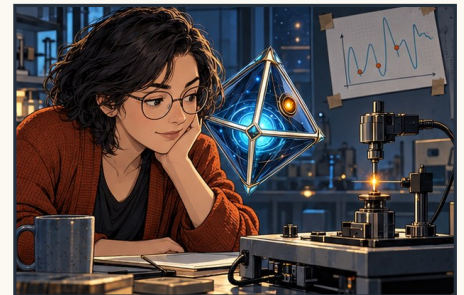
3 AXI Ik ben klaar voor elke volgende meting. Waarschijnlijk.



4 Onderzoeker Hier is nummer vier. Je waarschijnlijkheid heeft een probleem.



5 AXI Dan wil ik ook een eenvoudiger model en een eerlijke test proberen.



6 Onderzoeker Die oude curve krijgt een onderschrift: "Perfect op gisteren".

AXI is een fictief onderzoeksmaatje: een glazen octaëder met een cyaankleurige kern. De strip is een illustratie.

Waarom nul fout niet genoeg is

Een curve kan drie meetpunten exact raken en toch verkeerd voorspellen. Hoe meer knoppen een model heeft, hoe makkelijker het ook toevallige afwijkingen volgt: dat heet overfitting. Eén afwijkend punt kan ook meetruijs zijn, dus controleer eerst de meting. Daarna beslist een toets op gegevens die het model nooit zag. Nul trainingsfout is een uitspraak over gisteren, geen belofte over morgen.



AXI VRAAGT

Zou jij liever nul trainingsfout hebben of een kleine, goed getoetste fout op nieuwe metingen? Waarom?



LEREN • MODEL, ALGORITME, ARCHITECTUUR

Een verstelbare relatie

Door (1, 2), (2, 4) en (3, 6) gaat precies één rechte lijn: $a = 2$ en $b = 0$. Wat weet die lijn eigenlijk over $x = 4$?

Drie punten, twee parameters, één voorspelling

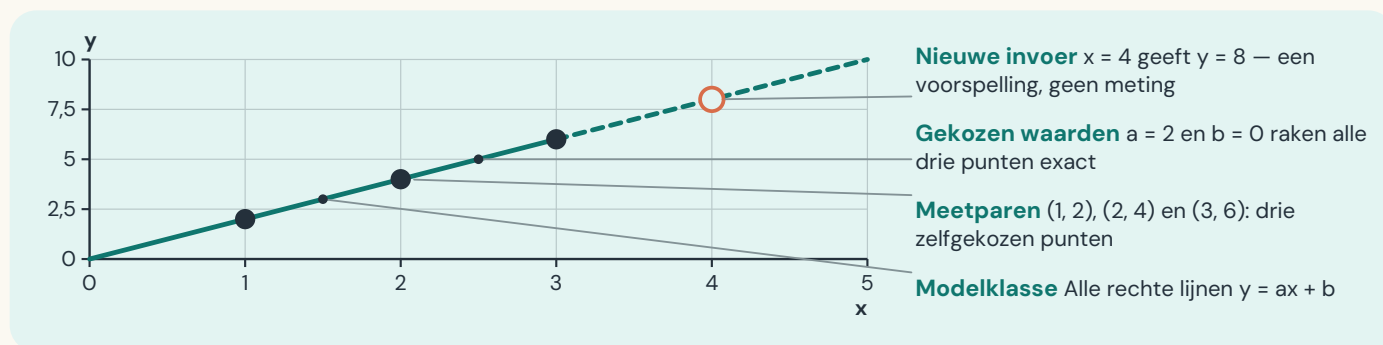


FIG. 1 Rekenvoorbeeld met zelfgekozen punten; de holle stip is een voorspelling, geen meting.

Drie dingen die vaak hetzelfde worden genoemd

Een model is een verstelbare relatie. Neem de zelfgekozen meetparen (1, 2), (2, 4) en (3, 6). Binnen de klasse $y = ax + b$ passen $a = 2$ en $b = 0$ exact, dus voorspelt de lijn bij $x = 4$ de waarde 8. Dat is een voorspelling, geen meting: buiten het interval van 1 tot 3 weten we alleen wat de lijn aanneemt. Het algoritme is de procedure die a en b koos; de architectuur bepaalt welke lijnen überhaupt mogelijk zijn. Gaat er iets mis, vraag dan apart naar invoer, parameters, modelklasse en toetsopzet.

Waarom niet gewoon meer parameters?

Extra parameters helpen een ingewikkelde relatie te benaderen, maar volgen net zo graag toeval. Heeft een sensor op dinsdag een afwijking en op woensdag niet, dan kan een rijk model dinsdag perfect beschrijven zonder het proces te begrijpen. Daarom splitsen onderzoekers hun gegevens: trainingsdata stellen de parameters af, validatiedata helpen kiezen tussen ontwerpen, en een onafhankelijke eindtest beoordeelt de keuze. Die eindtest is pas eerlijk als hij werkelijk nieuw is — zie 'Het lek in de toets'.



AXI Acht bij $x = 4$? Ik zet er pas geld op na een vierde meting.

KORTE BESCHOUWING

Het lek in de toets

Een model scoort bijna perfect op de testset. Hoe weet je of het de ziekte herkent — of de patiënt?

Splits op de juiste eenheid

Rekenvoorbeeld: 1.000 röntgenfoto's van 200 patiënten, vijf per patiënt. Verdeel je de foto's willekeurig 80/20 over training en test, dan heeft vrijwel elke testfoto een 'broertje' in de training: de

kans dat alle vier andere foto's van dezelfde patiënt in de test belanden is $0,2^4 = 0,16\%$. Het model kan dan scoren door de patiënt te herkennen in plaats van de aandoening. Splits daarom per patiënt, per meetdag of per instrument: de eenheid waarover je straks iets wilt beweren.

Bron: 8



LEREN · DE KETTINGREGEL

De stap onder de motorkap

Eén meetpaar geeft twee afgeleiden. De kettingregel laat je precies volgen hoe a en b veranderen.

Van helling naar update

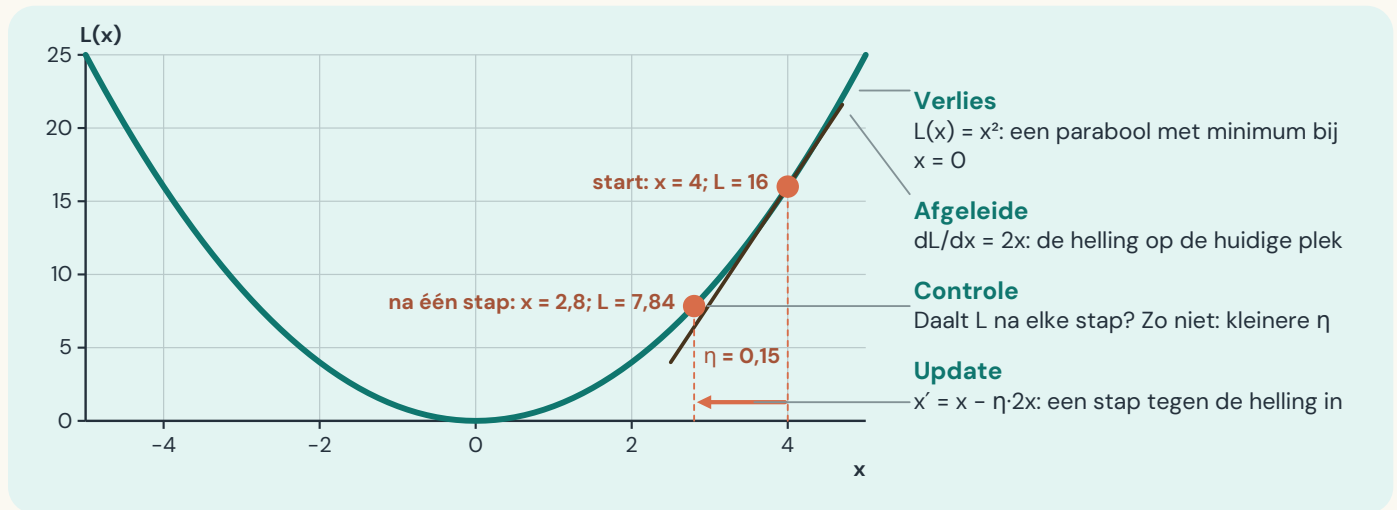


FIG. 2 Model: één parameter, verlies $L(x) = x^2$.

Het verlies is een ontwerpkeuze

Een verliesfunctie vat alle fouten samen in één getal. Je kunt gekwadrateerde afwijkingen optellen, of beoordelen hoeveel kans een model aan het juiste antwoord gaf. Elke keuze bepaalt welke fouten training het hardst bestrijdt. Gekwadrateerde fouten straffen één grote misser zwaarder dan tien kleine; een gemiddelde kan dalen terwijl de zeldzame, belangrijke gevallen slechter worden. Kies dus eerst welke fout voor de toepassing telt, en controleer dan of de verliesfunctie die zichtbaar houdt.

Volg de kettingregel

Neem het lijnmodel $\hat{y} = ax + b$ met verlies $L = (\hat{y} - y)^2$. De kettingregel geeft $\partial L / \partial a = 2(\hat{y} - y) \cdot x$ en $\partial L / \partial b = 2(\hat{y} - y)$. Met $x = 2$, $y = 5$, $a = 2$ en $b = 0$ is de voorspelling 4 en de fout -1 : de afgeleiden zijn -4 en -2 . Bij leerstap 0,1 worden $a = 2 + 0,4 = 2,4$ en $b = 0 + 0,2 = 0,2$. De nieuwe voorspelling is precies 5. Backpropagation doet hetzelfde in een netwerk met miljoenen parameters: één keer vooruit rekenen, één keer achteruit. ♦

WAAROM EEN PARABOOL?

Met één parameter kun je elke stap exact narekenen. Grote netwerken volgen hetzelfde recept in miljoenen dimensies tegelijk, op een veel grilliger verlieslandschap.

IN DE TIJD

Van kettingregel naar framework

Gradient descent volgt de helling; backpropagation levert die helling efficiënt voor alle parameters tegelijk.

1970 • Linnainmaa

Beschrijft in zijn masterscriptie het automatisch achterwaarts terugrekenen van afgeleiden (reverse mode).

1986 • Rumelhart, Hinton & Williams

Backpropagation laat verborgen lagen bruikbare interne representaties leren.

2017 • De Transformer

Attention vervangt recurrente lagen; ook dit netwerk leert met backpropagation.

2019 • PyTorch

Het PyTorch-artikel beschrijft automatische differentiatie als standaardonderdeel van een open framework.

Bronnen: 2, 4, 30, 53

Quiz: welke getuige weegt zwaarder?

Vier nieuwe vragen over de railrobot. De getallen zijn aangenomen; beargumenteer eerst en kijk daarna naar de uitwerkingen.

Gebruik de opgegeven miniwereld

De relevante formules en hun voorwaarden staan op pagina 12–13. De volledige oplossingen staan op pagina 49. Geef varianties in m^2 en posities in meter: een verwisseling verandert de betekenis.

1 Voor $x^- = 2,00 \text{ m}$, $P^- = 0,04 \text{ m}^2$, $z = 2,30 \text{ m}$ en $R = 0,01 \text{ m}^2$: wat is K ?

- (A) 0,2
- (B) 0,8
- (C) 4,0

Waarom?

2 Wat wordt x^+ met precies die waarden?

- (A) 2,24 m
- (B) 2,15 m
- (C) 2,80 m

Waarom?

3 Nu is $R = 0,09 \text{ m}^2$, terwijl $P^- = 0,04 \text{ m}^2$ blijft. Welke aanwijzing krijgt meer gewicht?

- (A) De sensor, want R is groter.
- (B) Beide precies evenveel.
- (C) De voorspelling; K is $4/13$.

Waarom?

4 Twee signalen delen dezelfde systematische tijdsfout. Welke conclusie volgt?

- (A) De eenvoudige onafhankelijkheidsaannname vraagt controle.
- (B) Twee signalen bewijzen automatisch de juiste positie.
- (C) Een neuralen netwerk maakt de fout zeker onafhankelijk.

Waarom?

Jouw antwoorden

1 2 3 4

De gewichtenkast

Drie oplosvragen met sluitende voorwaarden. Maak eerst een rekenspoor; de volledige bewijzen staan op pagina 48.

1 Twee posities $A = 1$ m en $B = 4$ m worden gemengd als $x = wA + (1-w)B$, met $0 \leq w \leq 1$. De uitkomst is 2 m. Welk gewicht w is uniek mogelijk?

Mijn antwoord

2 Een sensor meet $x+b = 7$. Zowel x als b is onbekend en ieder reëel getal is toegestaan. Kun je x uniek bepalen? Geef twee verschillende passende situaties.

Mijn antwoord

Alleen de aangegeven regels

Ieder raadsel is een eigen miniwereld. Een gewogen gemiddelde hoeft niet het werkelijke onbekende

3 Model A voorspelt $x(t)=t$; model B $x(t)=t^2$, met t in seconden en zelfgekozen bijpassende coëfficiënten. De enige foutloze waarneming is $x(1)=1$. Je mag vervolgens $t=0$ of $t=2$ meten. Welke tijd onderscheidt de modellen?

Mijn antwoord

getal te zijn. Bij de laatste opgave zijn precies de twee genoemde modellen toegestaan.

Zes vragen over geleerd gedrag

Kies één antwoord per vraag. De foute opties zijn misvattingen die je in het echt tegenkomt. Antwoorden en uitleg op pagina 47.

1 Het trainingsverlies van een model daalt van 0,40 naar 0,05. Wat weet je nu zeker?

- (A) Dat ook de fout op nieuwe gegevens vrijwel zeker is gedaald
- (B) Dat het model beter past op de trainingsgegevens, gemeten met de gekozen maat
- (C) Dat het model nu waarschijnlijk minder overfit dan eerst

2 Wat levert backpropagation op?

- (A) De nieuwe waarden van alle gewichten
- (B) De richting naar het globale minimum van het verlies
- (C) De afgeleide van het verlies naar elk gewicht

3 Een beeldclassifier heeft 95% van de testbeelden goed en zegt bijna altijd "99% zeker". Wat is er aan de hand?

- (A) Hij is overmoedig: bij 99% zekerheid hoort ongeveer 99% goed
- (B) Niets: een hoge nauwkeurigheid betekent betrouwbare kanscores
- (C) De testset is te klein om iets te concluderen

4 Het totale verlies van een physics-informed neural network is bijna nul. Wat kan er nog mis zijn?

- (A) Tussen de getoetste punten kan de vergelijking toch niet kloppen
- (B) Niets: een verlies van nul betekent dat de vergelijking overal klopt
- (C) Alleen de afrondingsfouten van de computer

5 AlphaFold 2 haalde in CASP14 een mediane afwijking van 0,96 Å. Wat zegt dat?

- (A) Dat elke voorspelde structuur tot op 1 Å nauwkeurig is
- (B) Dat ongeveer de helft van de voorspelde domeinen binnen 0,96 Å van het experiment lag
- (C) Dat AlphaFold 2 ook voorspelt hoe eiwitten bewegen en binden

6 Taalmodel X scoort 81% en model Y 78% op dezelfde benchmark. Wat moet je minimaal weten voordat je X kiest voor jouw toepassing?

- (A) Welk model de meeste parameters heeft
- (B) Of een tweede benchmark dezelfde volgorde geeft
- (C) Of taak, gegevens en beoordeling van de benchmark op jouw toepassing lijken

Jouw antwoorden

1	2	3
4	5	6

+ LEES JE EIGEN REDENERING

Een goed antwoord zonder reden is minder waard dan een fout antwoord waarvan je de aanname kunt aanwijzen.

■ WOWFEITEN · ELK GETAL HEEFT EEN REKENING

Zes getallen die je intuïtie tarten

Zes schaalvergelijkingen met hun rekensom, bron en grens.

DATAPUNT

10^{12} kg

NASA's orde van grootte voor een suikerklontje neutronenster

1 • Een suikerklontje met bergmassa

Volgens NASA weegt een suikerklontje neutronenstermaterie op aarde ongeveer 10^{12} kilogram: een miljard ton, zoveel als een berg. Het is een dichtheidsbeeld, geen monster dat op aarde bewaard kan blijven. Het zegt hoe compact zo'n ster is: zwaarder dan onze zon, maar zo klein als een stad.

DATAPUNT

0,300 mm

lichtweg in één picoseconde

2 • In een picoseconde legt licht drie tiende millimeter af

Een picoseconde is 10^{-12} seconde. Met $c = 299.792.458$ m/s legt licht in die tijd ongeveer 0,300 mm af in vacuüm. Kleine tijdsverschillen krijgen zo een ruimtelijke maat: een klokfout van één nanoseconde is al dertig centimeter.

DATAPUNT

$2,08 \times 10^{170}$

legale stellingen op een Go-bord van 19×19

3 • Meer Go-stellingen dan je kunt opschrijven

Op een Go-bord van 19×19 zijn ongeveer $2,08 \times 10^{170}$ legale stellingen mogelijk: een getal van 171 cijfers, exact uitgerekend in 2016 door John Tromp en collega's. Geen computer kan ze allemaal bekijken; een Go-programma moet dus leren welke stellingen de moeite waard zijn.

DATAPUNT

350 GB

alleen de gewichten van GPT-3 in 16 bits

4 • 350 gigabyte voordat er één woord valt

GPT-3 (2020) heeft 175 miljard parameters. Rekenvoorbeeld: sla je elk gewicht op als 16-bits getal van twee bytes, dan is dat 350 gigabyte, alleen aan gewichten. Training vraagt daarbovenop opslag voor gradiënten, optimizergegevens en tussenresultaten: een veelvoud daarvan.

DATAPUNT

$1,11 \times 10^{-17}$ kg

massa-equivalent van één joule

5 • Een joule heeft een massaschaduw

Volgens $\Delta m = \Delta E/c^2$ hoort bij één joule extra rustenergie ongeveer $1,11 \times 10^{-17}$ kilogram extra massa. Een opgeladen batterij is dus een fractie zwaarder dan een lege, ver onder elke weegschaal. Geeft een apparaat warmte af, dan verhuist die massa mee naar de omgeving.

DATAPUNT

> 200 miljoen

voorspelde structuren in de AlphaFold-database

6 • Tweehonderd miljoen voorspelde eiwitten

De AlphaFold-database bevat ruim 200 miljoen voorspelde eiwitstructuren. Experimenteel waren er in 2021 zo'n 100.000 unieke eiwitstructuren bepaald, na decennia werk. Elke voorspelling draagt wel haar eigen pLDDT-kleuren: niet elke structuur is even betrouwbaar.



AXI VRAAGT

Welke vergelijking heeft jou verrast – en welke voorwaarde houdt haar eerlijk?

TAAL

Woorden worden gewogen informatie

Conceptillustratie: drie informatieroutes komen gewogen samen. Geen afbeelding van een echt netwerk.



DOSSIER · TAAL

Een attentionstap van binnen

Scores 2, 1 en 0 worden gewichten. Pas daarna wordt informatie met waarden 10, 20 en 30 gecombineerd.

Een kleine attentionberekening

Scores 2, 1 en 0: hoe goed elke key bij de query past			
2	1	0	
Gewichten softmax: $0,665 + 0,245 + 0,090 = 1$			
0,665	0,245	0,090	
Values 10, 20 en 30: de informatie die meereist			
× 10	× 20	× 30	
Uitvoer $0,665 \cdot 10 + 0,245 \cdot 20 + 0,090 \cdot 30 \approx 14,25$			
6,65	4,90	2,70	$\Sigma \approx 14,25$

FIG. 1 Rekenvoorbeeld met zelfgekozen getallen; de balken zijn de softmaxgewichten.

Van tekens naar vectoren

Tekst wordt opgeknipt in tokens: woorden, woorddelen of leestekens. Het model rekent met een vector per token. Attention bepaalt per positie hoeveel informatie elke andere positie bijdraagt: een query wordt vergeleken met keys,

en de resulterende gewichten mengen values. Zo krijgt 'bank' in 'we zitten op de bank' een andere vector dan in 'geld naar de bank brengen': de burens kleuren de betekenis. Gemeten: de oorspronkelijke Transformer (2017) haalde 28,4 BLEU (een score voor overlap met menselijke vertalingen) op Engels-Duits, ruim twee punten boven alle eerdere systemen.

Een gewicht is nog geen verklaring

Een hoog attentiongewicht zegt welke informatie wordt verplaatst, niet waarom het antwoord klopt: wat latere lagen ermee doen, telt ook. Mechanistisch onderzoek aan kleine modellen vond wel concrete circuits, zoals 'induction heads' die een eerder patroon in de tekst kopiëren (Elhage e.a., 2021). Een groot taalmodel is daarmee nog lang niet doorgelicht.

Reken één attentionstap na

In formulevorm: uitvoer = $\text{softmax}(QK^T/\sqrt{d_k}) \cdot V$, met de queries, keys en values in Q, K en V. Ons rekenvoorbeeld laat $\sqrt{d_k}$ weg en neemt de scores 2, 1 en 0. Softmax maakt daar $e^2/(e^2 + e + 1)$, $e/(e^2 + e + 1)$ en $1/(e^2 + e + 1)$ van: ongeveer 0,665, 0,245 en 0,090, samen precies 1. Met values 10, 20 en 30 wordt de uitvoer $0,665 \times 10 + 0,245 \times 20 + 0,090 \times 30 \approx 14,25$. Echte modellen doen dit met vectoren, in tientallen heads en lagen tegelijk.



AXI 'Bank' om op te zitten of om geld te brengen? Dat weet ik pas als ik de burens heb gewogen.



Scores worden gewichten en gewichten combineren informatie. Met drie getallen reken je één attentionstap na. Daarna de lastigere vraag: hoe toets je of een vloeiend antwoord ook klopt?

VERVOLG • TAAL

Van antwoord naar toets

Acht goede antwoorden kunnen de hoogste score zijn — tot twee extra vragen van een ander type de rangorde omdraaien.

Een score heeft een onderzoeksofzet nodig

Een evaluatie bestaat uit taak, gegevens, beoordelingsprocedure en maatstaf. Verander er één en het percentage betekent iets anders. HELM (Holistic Evaluation of Language Models, 2022) nam dat serieus: 30 taalmodellen op 42 scenario's, gemeten op zeven eigenschappen, van nauwkeurigheid tot kalibratie en toxiciteit. Gemeten: vóór HELM waren modellen gemiddeld op slechts 17,9% van dezelfde kernscenario's getest; daarna op 96%. Pas dan is een vergelijking eerlijk.

Feitelijkheid vraagt een eigen toets

Een taalmodel leert tekst voortzetten, niet zinnen controleren. TruthfulQA legde modellen 817 vragen voor waarop mensen vaak een misvatting herhalen. Gemeten: het beste model was op 58%

van de vragen waarheidsgetrouw, mensen op 94%, en de grootste modellen scoorden vaak het slechtst: ze imiteerden de misvattingen beter. Dat gaat over de modellen van 2021, maar de les blijft: stijl en feitelijkheid vragen aparte toetsen. Open bij een verwijzing het document en zoek de claim.

Een proef die je zelf kunt organiseren

Rekenvoorbeeld: tien korte feitenvragen. Systeem A heeft er acht goed, B zeven. Voeg twee vragen toe waarvoor broncontrole nodig is: A mist beide, B heeft beide goed. De eindstand wordt 8/12 tegen 9/12. Geen systeem veranderde, alleen de taakmix. Daarom horen aantallen, vraagtypen en beoordelingsregels bij elk percentage. ♦

VIER PIJLERS VAN EEN EVALUATIE

Taak • gegevens • procedure • maatstaf. Zonder die vier weet je niet goed wat een percentage beweert.

“Een taalmodel leert tekst voortzetten, niet zinnen controleren.”



Een robotpositie met twee getuigen

Een railrobot rijdt achter een scherm. De motor voorspelt waar hij is; een sensor ziet hem opnieuw. Hun verschil wordt de volgende stap van een rekensom.



FIG. 1 AI-illustratie • sensorfusie van een railrobot; wolken verbeelden aangenomen onzekerheid. AXI en de onderzoeker zijn fictief.

Een expliciete miniwereld

Ons fictieve karretje beweegt langs één rechte rail. De toestand is alleen de positie x in meter. Veronderstel een lineaire beweging, bekende gemiddelde stappen en onafhankelijke, nulgemiddelde fouten. Na de motorstap is de voorspelling $x^- = 2,00$ m. De bijbehorende variantie $P^- = 0,04$ m² zegt hoeveel spreiding het model verwacht; zij is geen gemeten fout van dit ene karretje.

De nieuwe sensor spreekt tegen

De sensor geeft $z = 2,30$ m, met aangenomen ruisvariantie $R = 0,01$ m². Het verschil $z - x^-$ heet de innovatie: hier $0,30$ m. Een klassieke scalaire Kalmanstap gebruikt $K = P^- / (P^- + R) = 0,04 / 0,05 = 0,8$. De correctie is K maal het verschil: $0,24$ m. Zo volgt $x^+ = 2,24$ m. Het resultaat blijft tussen beide aanwijzingen, dichterbij de aangenomen preciezere sensor.

Eerst vragen, dan vertrouwen

Kalman beschreef in 1960 een recursieve schatter voor lineaire systemen. Ons eigen voorbeeld reduceert die tot één getal. Er is geen neurale netwerk nodig om deze stap uit te rekenen. AI wordt pas een aanvullende keuze wanneer onderdelen van de dynamica of de schatting uit data worden geleerd. Schrijf daarom bij ieder resultaat op wat gegeven, aangenomen of geleerd is. De volgende pagina rekent door wat een tegenvallende aanname doet.

REKENKAART

$x^+ = x^- + K(z - x^-)$. Alle posities, varianten en sensorwaarden in dit dossier zijn aangenomen voorbeelden.



■ VECTOR • ONDERZOEK MET AANNAMEN

De foutband groeit mee met de volgende stap

De positie 2,24 m is één uitkomst. De belangrijkste vervolgvragen gaan over spreiding, afhankelijkheid en een beweging die het model niet verwachtte.

Na de correctie

Onder dezelfde scalaire aannamen is $P^+ = (1 - K)P^- = 0,008 \text{ m}^2$. De standaardafwijking is $\sqrt{0,008} \approx 0,0894 \text{ m}$. Wordt de volgende motorstap 0,50 m en voegt beweging een onafhankelijke variantie $Q = 0,002 \text{ m}^2$ toe, dan voorspellen we 2,74 m met $P^- \text{ nieuw} = 0,010 \text{ m}^2$. De foutband wordt dus eerst smaller door meten en daarna weer breder door bewegen.

Wat als de rail ineens glad wordt?

Een vergeten slip kan een bias veroorzaken: alle volgende voorspellingen wijken dezelfde kant op. De papieren variantie kan dan te optimistisch blijven. Bewaar daarom innovaties over de tijd, bekijk hun gemiddelde en test nieuwe snelheden. Wanneer model en sensor dezelfde tijdstempelfout gebruiken, zijn hun fouten bovendien afhankelijk; de eenvoudige formule kan dat verband niet wegtoveren.

Waar leren iets kan toevoegen

KalmanNet combineert de structuur van toestandschatting met een recurrent neurale onderdeel bij deels bekende dynamica. Dat onderzoeksresultaat betekent niet dat een geleerd filter iedere klassieke schatter verslaat. Voor onze rail zouden we een volledig uitgeschreven baseline, trainingsritten en aparte slipritten nodig hebben. Vergelijk positiefout én de dekking van gerapporteerde intervallen. Een kleine fout en een te stellige foutband kunnen tegelijk voorkomen. ♦

DATAPUNT

2,24 m

gecorrigeerde positie

DATAPUNT

0,008 m²

variantie na meten

DATAPUNT

2,74 m

volgende voorspelling

DATAPUNT

0,010 m²

variantie na bewegen

Drie beloftes die te mooi zijn

AXI krijgt drie aantrekkelijke aanbiedingen. Leg eerst de aannames vast, kies jouw antwoord en vergelijk daarna met de uitwerking op pagina 47.

Jouw keuze en redenering

1 • De demon met de volle schijf

Een poortwachter sorteert gasdeeltjes en noteert elk besluit als een bit die met gelijke kans 0 of 1 is. Zijn geheugen is eindig.

Kan hij zijn hele machine steeds terugzetten zonder thermodynamische rekening?

Denk aan:

- Wat gebeurt bij vol geheugen?
- Is extra geheugen gratis?

Jouw keuze en redenering

2 • De universele quantumkopieerder

Een apparaat belooft bij elke onbekende zuivere qubittoestand een perfecte extra kopie, met behoud van het origineel.

Waarom is die universele belofte onmogelijk?

Denk aan:

- Kopieer eerst $|0\rangle$ en $|1\rangle$.
- Wat volgt voor hun superpositie?

Jouw keuze en redenering

3 • Twee perfecte curves

Bij $x = 0, 1$ en 2 zijn de oefenwaarden $0, 1$ en 4 . A gebruikt $y = x^2$; B gebruikt $y = x^2 + x(x - 1)(x - 2)$.

Welke waarde bij $x = 3$ bewijzen die punten?

Denk aan:

- Test de productterm bij $x = 0, 1$ en 2 .
- Reken A en B uit bij $x = 3$.



AXI VRAAGT

Welke verborgen hulpbron, extra aanname of ontbrekende meting zou elke belofte wél waarmaken?

SIMULATIE 01 • GRADIENT DESCENT

Hoe groot mag de leerstap zijn?

Kies een leerstap η , voorspel wat er gebeurt en controleer het met de rekenregel $x' = (1 - 2\eta) \cdot x$. Eén parameter, elke stap exact na te rekenen.

Vier leerstappen vanaf $x = 4$

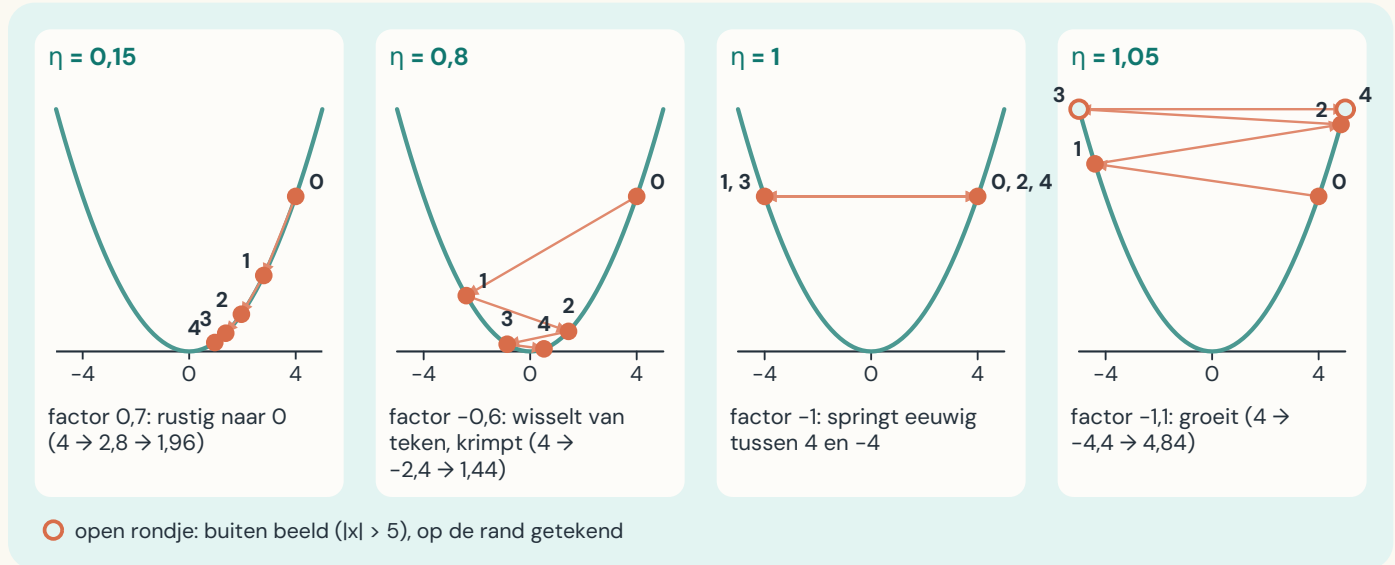


FIG. 1 Berekend met $x' = (1 - 2\eta) \cdot x$; model, geen gemeten gegevens.

Voorspel eerst, reken daarna

Begin bij $x = 4$ en $\eta = 0,15$. De update $x' = x - 0,15 \times 2x$ vermenigvuldigt x telkens met 0,7. Na één stap is $x = 2,8$ en $L = 7,84$; na twaalf stappen is $x \approx 0,0554$ en $L \approx 0,00307$. Bij $\eta = 0,8$ wordt de factor -0,6: x wisselt van teken maar krimpt ($4 \rightarrow -2,4 \rightarrow 1,44$). Heen en weer springen is hier dus geen instabiliteit. Kijk naar de grootte van x en het verlies. Een start bij 0 blijft 0: daar is de helling nul.

De grens zelf afleiden

Omdat $x' = (1 - 2\eta) \cdot x$, convergeert elke start $x_0 \neq 0$ precies als $|1 - 2\eta| < 1$, dus voor $0 < \eta < 1$. Bij $\eta = 0,5$ is de factor 0: in één stap op het minimum. Bij $\eta = 1$ springt x eeuwig tussen 4 en -4. Boven 1 groeien x en verlies ($4 \rightarrow -4,4 \rightarrow 4,84$ bij $\eta = 1,05$). Deze grens hoort bij deze parabool: een steilere kromming verlaagt de toegestane leerstap. Juist omdat het model zo klein is, kun je de grens exact afleiden.

Wat je meeneemt

Er zijn vier regimes: rustig dalen ($0 < \eta < 0,5$), in één stap raak ($\eta = 0,5$), wisselend krimpen ($0,5 < \eta < 1$) en niet meer convergeren ($\eta \geq 1$). Positieve en negatieve starts van dezelfde grootte geven gespiegelde waarden en hetzelfde verlies. Minibatches en meerdere parameters vallen buiten dit model.

Gradient descent op een parabool

$L(x) = x^2$; $x' = x - \eta \cdot 2x$. Het minimum ligt bij $x = 0$.



AXI Bij $\eta = 0,5$ sta ik in één stap op het minimum. Waarom werkt die truc alleen op deze parabool?

ONZEKERHEID · DE REGEL VAN BAYES

Wat een signaal werkelijk zegt

Negen echte en tien valse signalen. Die negentien gevallen vormen de noemer na een detectie.

Een detector in een denkbeeldige fabriek

	signaal	geen signaal	totaal
Afwijking	9 Afwijking, signaal 9 terechte treffers	1 Afwijking, geen signaal 1 gemiste afwijking	10
Goed	10 Goed, signaal 10 valse alarmen	980 Goed, geen signaal 980 terecht stil	990
	totaal 19	totaal 981	1.000

9 van 19 ≈ 47 %

FIG. 1 Rekenvoorbeeld: hypothetische aantallen per 1.000 onderdelen.

Draai de voorwaarde niet om

$P(\text{signaal} \mid \text{afwijking})$ is iets anders dan $P(\text{afwijking} \mid \text{signaal})$. Rekenvoorbeeld: een fabriek maakt duizend onderdelen, waarvan tien met een afwijking. De detector vindt er negen, maar slaat ook aan bij tien van de 990 goede onderdelen. Er zijn dus negentien signalen: negen echte en tien valse. Na een signaal is de kans op een afwijking $9/19$, ongeveer 47,4%. De regel van Bayes vat dat samen: $P(A \mid S) = P(S \mid A) \cdot P(A) / P(S) = 0,9 \times 0,01 / 0,019 \approx 0,474$. Dat is geen tegenspraak met de trefkans van 90%: de twee percentages hebben een andere noemer.

De uitgangskans bepaalt veel

Dezelfde detector betekent iets heel anders als afwijkingen vaker of zeldzamer zijn (zie het kader). Baseer de uitgangskans daarom op een meting of een beschreven populatie, en reken bij twijfel meerdere waarden door: een gevoeligheidsanalyse. Ook de trefkans en het aantal valse alarmen zijn zelf geschat in een test. Houd die twee onzekerheden apart, anders lijkt één precies percentage meer te weten dan je weet.

ZO ZIT DAT Dezelfde detector, een andere uitgangskans

Trefkans 90% en vals alarm bij 1,01% van de goede onderdelen blijven gelijk. Alleen het aandeel afwijkingen verandert.

0,1% afwijkingen

Na een signaal is ongeveer 8% echt.

1% afwijkingen

Na een signaal is $9/19 \approx 47,4\%$ echt.

10% afwijkingen

Na een signaal is ongeveer 91% echt.

Bron: 10

DATAPUNT

9/19

terecht na signaal in het didactische voorbeeld



AXI VRAAGT

Waarom delen we negen door negentien, en niet door duizend? Wijs de gekozen groep aan in de tabel.

■ ONZEKERHEID · KALIBRATIE

Zekerheid vraagt een populatie

Twee groepen kunnen elk verkeerd gekalibreerd zijn terwijl hun totaal precies bij 80% zekerheid past.

Het totaal verbergt twee fouten

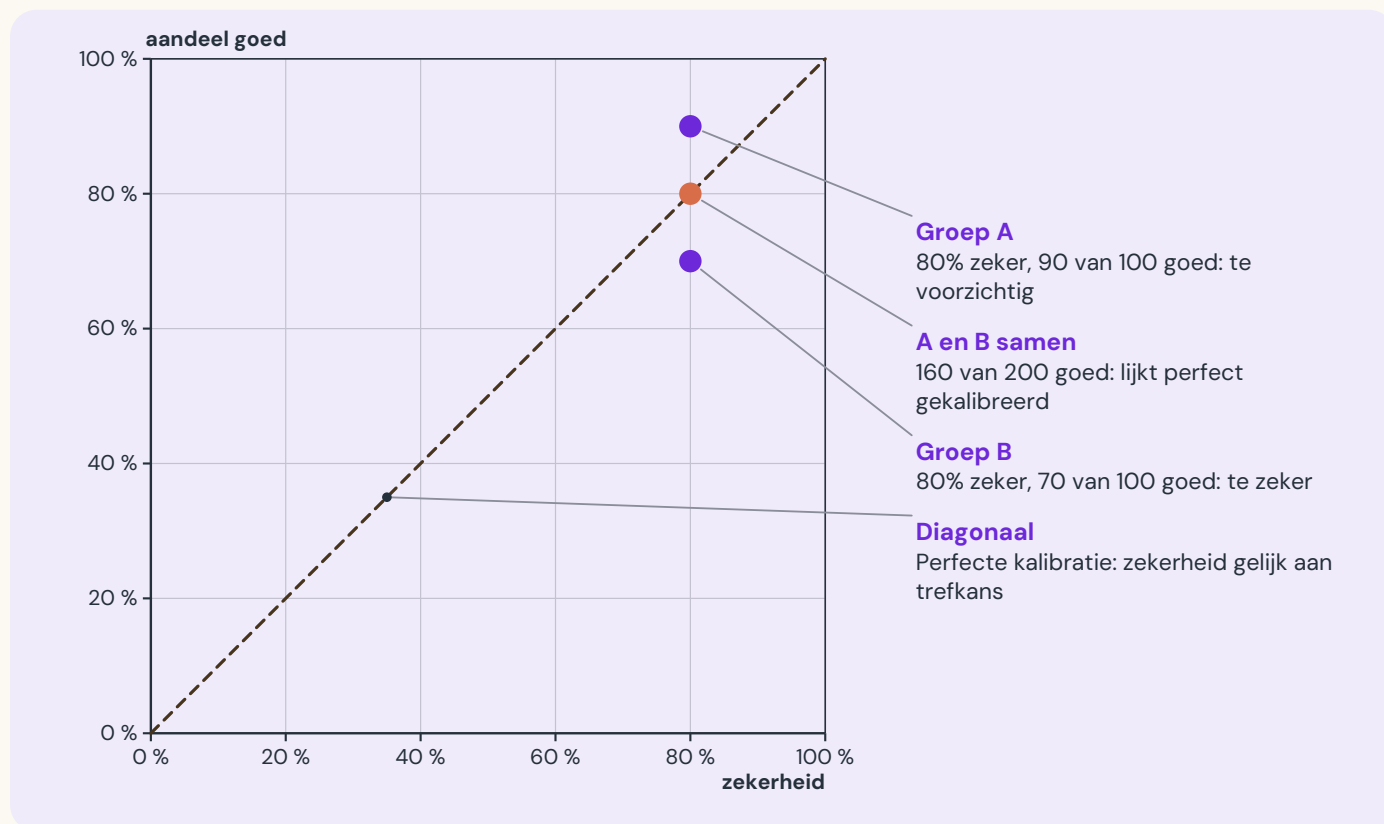


FIG. 2 Rekenvoorbeeld: twee zelfgekozen groepen van 100 voorspellingen met 80% zekerheid.

Goed kiezen en goed twijfelen?

Nauwkeurigheid en kalibratie zijn verschillende eigenschappen. Gemeten: Guo en collega's (2017) vergeleken een vijflaags LeNet uit 1998 met een ResNet van 110 lagen op de beeldenset CIFAR-100. De ResNet maakte minder fouten (30,6% tegen 44,9%), maar zijn kansscores weken gemiddeld 16,5 procentpunt af van de werkelijke trefkans; bij LeNet was dat 4,9. Eén extra getal bracht dat terug naar 1,3: de temperatuur T in $p = \text{softmax}(z/T)$, afgesteld op een validatieset. Met $T > 1$ worden de scores minder extreem.

Een betrouwbaarheidsdiagram?

Een betrouwbaarheidsdiagram (reliability diagram) groepeerde voorspellingen op hun zekerheid en zet die af tegen het aandeel dat klopte. Perfecte kalibratie ligt op de diagonaal.

Hoe je de groepen kiest, beïnvloedt wat zichtbaar wordt, en kleine groepen geven grillige punten. Vermeld daarom aantallen per groep.

Een goed totaal kan twee slechte groepen verbergen?

Rekenvoorbeeld: twee groepen van elk honderd voorspellingen met steeds 80% zekerheid. In groep A kloppen er negentig, in B zeventig. Samen zijn dat 160 van 200: precies 80%. Het totaal lijkt perfect gekalibreerd, terwijl A te voorzichtig en B te zeker is. Kijk dus ook naar relevante deelgroepen – bijvoorbeeld beelden uit een ander ziekenhuis. ♦



AXI Ik ben 99% zeker dat ik goed gekalibreerd ben. Hoe zou jij dat bij mij controleren?

AXI en de twee perfecte thermometers

Twee gelijke aanwijzingen verleiden AXI tot een conclusie die de ijking nooit heeft beloofd.



1 **AXI** Twee thermometers. Allebei 24,0. Dubbel bewijs!



2 **Onderzoeker** En tegen welke referentie zijn ze geijkt?



3 **AXI** Dezelfde. Die stond twee graden verkeerd.



4 **Onderzoeker** Kijk: in het gecontroleerde bad verschuiven beide samen.



5 **AXI** Mijn tweede scherm was geen tweede oorsprong.



6 **Onderzoeker** Precies. Controleer ook de gedeelde geschiedenis.

AI-illustratie • Zes fictieve scènes met AXI en een onderzoeker. Dialoog staat onder ieder kader; de beelden leveren geen meetgegevens.

Een gemeenschappelijke voorgeschiedenis

De strip gebruikt fictieve instrumenten. Hun gelijke aanwijzingen kunnen voortkomen uit dezelfde

vershoven referentie. De overeenkomst is echt, maar de afleiding "dus juist" volgt niet. Een tweede onafhankelijke controle gaat over de referentie, niet over het uiterlijk van de cijfers.

■ VECTOR · ONDERZOEK MET AANNAMEN

Zes kleine cijfers, zes grote verschillen

Nieuwe meetweetjes uit zelfgekozen voorbeelden. Iedere kaart vermeldt wat wordt geteld en welke stap het resultaat draagt.

DATAPUNT

1

oorspronkelijke waarneming

1 • Een ruisbron verdwijnt niet door kopiëren

Kopieer één meting honderd keer: je hebt honderd records maar één waarneming. Gemiddeldes van de kopieën zijn exact gelijk aan het oorspronkelijke getal. Dit is een informatierekening, geen eigenschap van een bepaald AI-product.

DATAPUNT

0,10 m

ruis van het gemiddelde

2 • Vier onafhankelijke herhalingen

Neem vier onafhankelijke metingen met dezelfde nulgemiddelde standaardafwijking 0,20 m. Het gemiddelde heeft onder die aannamen standaardafwijking $0,20/\sqrt{4} = 0,10$ m. De vaste bias zou blijven staan.

DATAPUNT

3 mm

verplaatsing in 1 ms

3 • Een milliseconde beweging

Bij aangenomen constante snelheid 3 m/s is 0,001 s tijdsverschil goed voor 0,003 m verplaatsing. De klokaanname wordt zo onderdeel van de positiebegroting.

DATAPUNT

×4

variantie bij dubbele spreiding

4 • Een vergeten kwadraat

Verdubbel een standaardafwijking van 0,10 naar 0,20 m. De variantie groeit van 0,01 naar 0,04 m²: een factor vier. Het ene soort foutgetal leest dus anders dan het andere.

DATAPUNT

16

vierbits patronen

5 • Vier bits voor zestien labels

Een vaste binaire code met vier bits heeft $2^4 = 16$ mogelijke patronen. Zijn ze slechts meetlabels, dan leveren meer bits geen betere sensor. Een codecapaciteit en meetnauwkeurigheid zijn aparte vragen.

DATAPUNT

geen

eenheid van de gain

6 • Een gain zonder eenheid

P en R hebben beide eenheid m². In $K = P/(P+R)$ vallen die eenheden weg. De dimensieloze gain kan daarom een verschil in meter vermenigvuldigen en een correctie in meter geven.

Zes vragen over de wereld achter het model

Denk eerst na over de voorwaarde in de vraag: systeemgrens, referentiestelsel of meetopzet. Antwoorden en uitleg op pagina 47.

1 Een slee glijdt met constante snelheid rechtdoor over vlak, wrijvingsloos ijs. Wat is de resulterende kracht op de slee?

- (A) Een kleine constante kracht in de bewegingsrichting
- (B) Het gewicht van de slee, recht naar beneden
- (C) Nul

Waarom?

2 Een auto versnelt van 30 naar 90 km/h. Hoeveel keer zo groot wordt zijn kinetische energie?

- (A) 9 keer
- (B) 3 keer
- (C) 6 keer

Waarom?

3 Alice en Bob meten verstrengelde fotonen en vinden $S = 2,4$, zonder mazen in de opzet. Wat volgt daaruit?

- (A) De meting van Alice stuurt Bobs uitkomst, dus ze kunnen er berichten mee seinen
- (B) Geen model met vooraf vastgelegde, lokale uitkomsten kan deze correlaties verklaren
- (C) De detectoren zijn slecht afgesteld, want correlaties kunnen niet boven 2 komen

Waarom?

4 Wat zou er gebeuren als GPS-klokken niet op relativiteit waren afgesteld?

- (A) Ze zouden per dag ongeveer 38 μ s voorlopen op klokken op de grond
- (B) Ze zouden achterlopen, omdat bewegende klokken altijd trager tikken
- (C) Niets merkbaars: het effect is kleiner dan een nanoseconde per dag

Waarom?

5 Een koelkast koelt zijn binnenkant af: de entropie daar daalt. Klopt dat met de tweede hoofdwet?

- (A) Nee, entropie kan nergens dalen
- (B) Ja, want de keuken wint meer entropie dan de binnenkant verliest
- (C) Ja, want de tweede hoofdwet geldt niet voor apparaten met een motor

Waarom?

6 Wat beschrijft de Landauergrens $kT \ln 2$?

- (A) De minimale energie per token van een taalmodel
- (B) De minimale warmte bij het wissen van één bit, bij temperatuur T
- (C) De minimale energie van elke rekenstap, ook een omkeerbare

Waarom?

Jouw antwoorden

1	2	3
4	5	6

PAPER-CHECKLIST · NEGEN VRAGEN DIE EEN CLAIM PRECIES MAKEN

Lees de methode vóór het applaus

Een abstract heeft weinig ruimte. Een opvallend resultaat wordt pas interpreteerbaar wanneer taak, gegevens en vergelijkingsmethode duidelijk zijn.

01 Zoek de zin met de echte claim

Onderstreep de smalste beschrijving van wat het onderzoek aantoonde. Is het een verbetering op één dataset, een nieuwe numerieke methode of een fysische waarneming? Herschrijf die zin in je eigen woorden, zonder versterkende termen. “Sneller op deze drie vergelijkingen, bij deze foutgrens” is informatief; “AI vervangt fysica” niet. Een precieze claim kan nog steeds indrukwekkend zijn. De checklist is er om de grens te vinden, niet om elk resultaat af te kraken.

02 Controleer de tegenverklaring

Rekenvoorbeeld: een paper meldt “30% minder fout”. Vraag eerst naar de baseline: van 0,10 naar 0,07 is een absolute winst van 0,03 en een relatieve van 30%. Beide kloppen; noem bewust welke je bedoelt. Kregen model en baseline dezelfde gegevens, hetzelfde afstembudget voor hun instellingen (hyperparameters) en dezelfde eindtest? Een ablatiestudie haalt één onderdeel weg om zijn bijdrage te meten. Veranderden rekenbudget en architectuur tegelijk, dan weet je nog niet welke van de twee de winst gaf.

Zo doe je het

- 1 Welke taak wordt precies uitgevoerd, met welke invoer en uitvoer?
- 2 Welke gegevens zijn gebruikt voor training, selectie en eindtest?
- 3 Wat is de relevante baseline en heeft die een eerlijk budget?
- 4 Welke foutmaat past bij de inhoudelijke vraag?
- 5 Zijn onzekerheid, spreiding en verschillende runs gerapporteerd?
- 6 Welke claim draagt de testopzet en welke bredere claim nog niet?
- 7 Welke onderdelen zijn afzonderlijk onderzocht of weggehaald?
- 8 Zijn code, data en instellingen voldoende beschreven om de uitkomst te controleren?
- 9 Welke nieuwe situatie zou het resultaat overtuigend op de proef stellen?



AXI IN DE MARGE

Oefen op deze zin: “Ons model verlaagt de fout met 30% en vervangt dure simulaties.” Welke drie vragen uit de lijst stel je eerst?



AXI Mijn tiende vraag: staat ‘state-of-the-art’ in de titel? Dan lees ik de baseline twee keer.



WERELDMODELLEN · TOESTAND EN UPDATE

Een wereldmodel voorspelt de volgende stap

Een geleerd model kan van de huidige toestand naar een volgende toestand rekenen. Het modelontwerp bepaalt wat een toestand bevat en welke veranderingen zichtbaar kunnen worden.

Van kracht naar traject

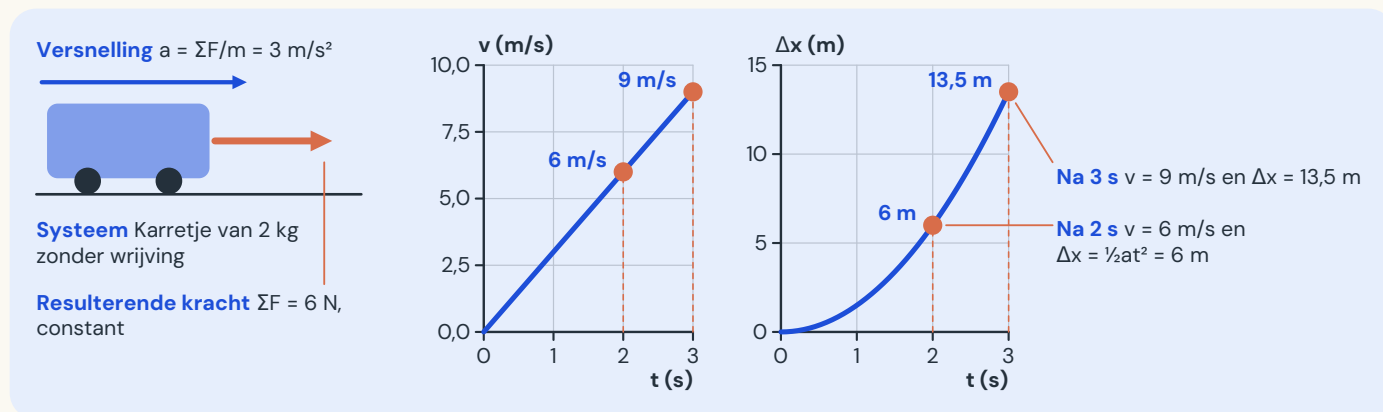


FIG. 1 Rekenvoorbeeld: $m = 2 \text{ kg}$, $\Sigma F = 6 \text{ N}$, vanuit rust, zonder wrijving.

01 Wat krijgt de leerder te zien?

Een bewegend karretje kan worden beschreven met positie en snelheid. Een cameramodel kan in plaats daarvan pixels krijgen. Die keuzes vragen andere verwerking: een losse foto vertelt niet meteen hoe snel het karretje beweegt. Een geleerde simulator voorspelt vervolgens de verandering over een vaste tijdstap. Onderzoek met graafnetwerken ordent zulke informatie rond objecten en relaties. De structuur helpt hergebruik van lokale regels; zij bewijst niet dat alle relevante krachten uit de gegevens volgen.

02 Een exact referentievoorbeeld

Onze fysische controle blijft eenvoudig. Voor constante massa 2 kg , resulterende kracht 6 N en geen wrijving geldt $a = \Sigma F/m = 3 \text{ m/s}^2$. Vanuit rust is na 2 s de snelheid 6 m/s en de verplaatsing 6 m ; na 3 s zijn dat 9 m/s en $13,5 \text{ m}$. Dezelfde getallen in verschillende eenheden betekenen verschillende grootheden. Een geleerd traject kan met deze referentie worden vergeleken voordat we ingewikkelder omstandigheden onderzoeken.

1 stap

lokale nauwkeurigheid is geen garantie voor een heel traject

03 Wat betekent een gunstige fout?

Een model kan één stap nauwkeurig voorspellen en bij herhaald vooruitrekenen afdrijven. Toets daarom zowel de lokale fout als een volledig traject op nieuwe beginsituaties. Behoud een eenvoudige referentiemethode naast het netwerk. Het woord wereldmodel is hier een taakbeschrijving: een model van gekozen dynamica, geen bewijs dat het systeem de complete wereld begrijpt.

ZO ZIT DAT Een wet uit vijf metingen

Rekenvoorbeeld met verzonnen meetwaarden. Duw hetzelfde karretje met vijf verschillende krachten en meet de versnelling.

Metingen

$F = 2, 4, 6, 8 \text{ en } 10 \text{ N}$ geven $a = 1,0; 2,1; 2,9; 4,0 \text{ en } 5,1 \text{ m/s}^2$.

Model

$a = kF$, met één parameter $k = 1/m$.

Leren

Minimaliseer $\Sigma(a - kF)^2$, met gradiëntstappen als op pagina 5: $k \approx 110,8/220 \approx 0,504 \text{ per kg}$.

Uitkomst

$m = 1/k \approx 1,99 \text{ kg}$. De geleerde parameter heeft een fysische betekenis.

Bron: 12



■ WERELDMODELLEN · EEN EXTRA REKENING

Energie als controle op een geleerde beweging

Een voorspeld traject kan er aannemelijk uitzien terwijl het energie uit het niets maakt. Een energierekening levert een aparte, controleerbare vraag.

Verdubbel v , verviervoudig E_k

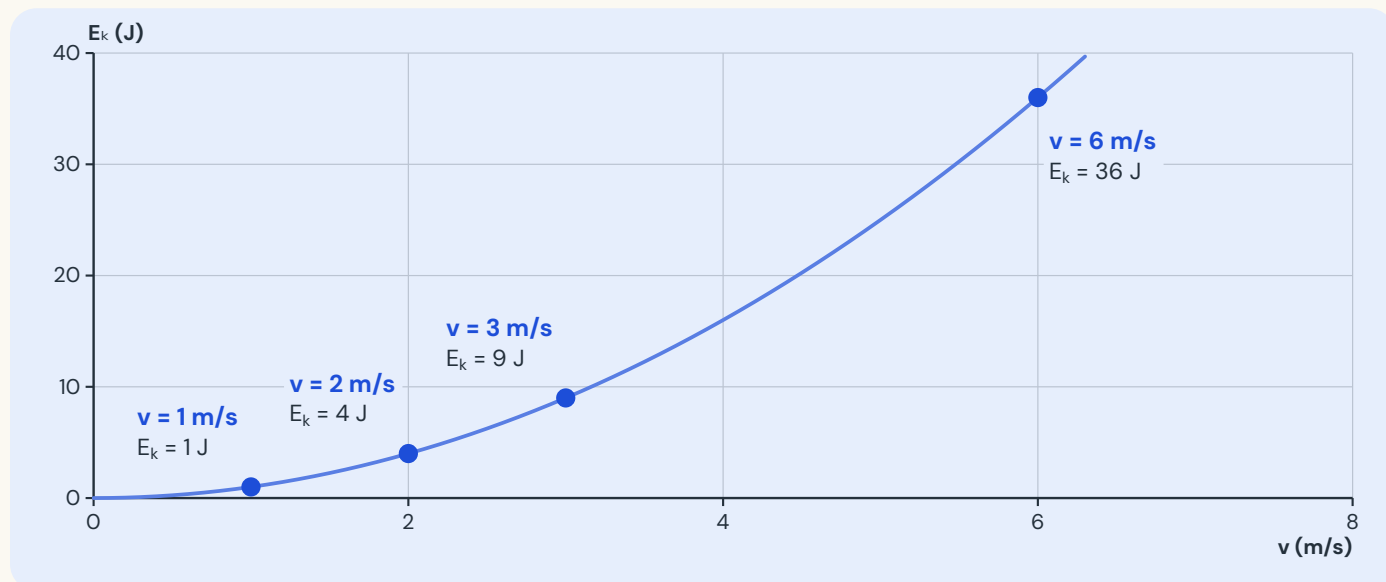


FIG. 3 Rekenvoorbeeld met $m = 2$ kg en de klassieke formule $E_k = \frac{1}{2}mv^2$.



FIG. 2 Russell Falls, Tasmanië: zwaartekrachtenergie wordt bewegingsenergie en uiteindelijk warmte.

Wat hoort het model te behouden?

Voor een klassiek voorwerp is $E_k = \frac{1}{2}mv^2$. Twee kilogram met 3 m/s heeft 9 J; met 6 m/s is dat 36 J. Een verdubbelde snelheid geeft dus viermaal zoveel kinetische energie. Een AI-simulator die zonder externe arbeid het eerste traject in het tweede verandert, moet verklaren waar die extra energie vandaan komt. Een scherp beeld of kleine

gemiddelde positiefout beantwoordt die vraag niet.

Kies de systeemgrens

Wrijving kan mechanische energie in interne energie omzetten. Een model hoeft dus niet altijd dezelfde mechanische energie te behouden. De controle moet passen bij wat binnen het systeem wordt gerekend en welke stromen zijn toegestaan. Een beperking die voor een ideaal wrijvingsloos model klopt, kan bij een ander systeem verkeerd zijn. Leg die keuze vast voordat je energiefouten tot netwerkfouten verklaart.

Een tweede route naar de snelheid

Twee kilogram die 3 m daalt verliest bij $g \approx 9,81$ m/s² ongeveer 58,9 J zwaartekrachtenergie. Zonder weerstand volgt $v = \sqrt{2gh} \approx 7,67$ m/s vanuit rust. Dat is een onafhankelijke rekenroute voor hetzelfde ideale proces. Een geleerd model kan ertegen worden getoetst. De energiebalans controleert een relevante eigenschap; zij bewijst niet dat het volledige traject of elk nieuw contact juist is.

QUANTUM

Een toestand is meer dan haar losse kansen

Conceptillustratie van een bron en twee meetstations met polarisatoren en detectoren; er loopt geen berichtkanaal tussen de stations.



DOSSIER • QUANTUM

Amplituden, basis en fase

Een AI die meetuitkomsten verwerkt krijgt niet automatisch alle informatie over een quantumtoestand. Fase en meetbasis laten zien waarom losse kansen onvoldoende kunnen zijn.

Dezelfde 0/1-kansen, een andere fase

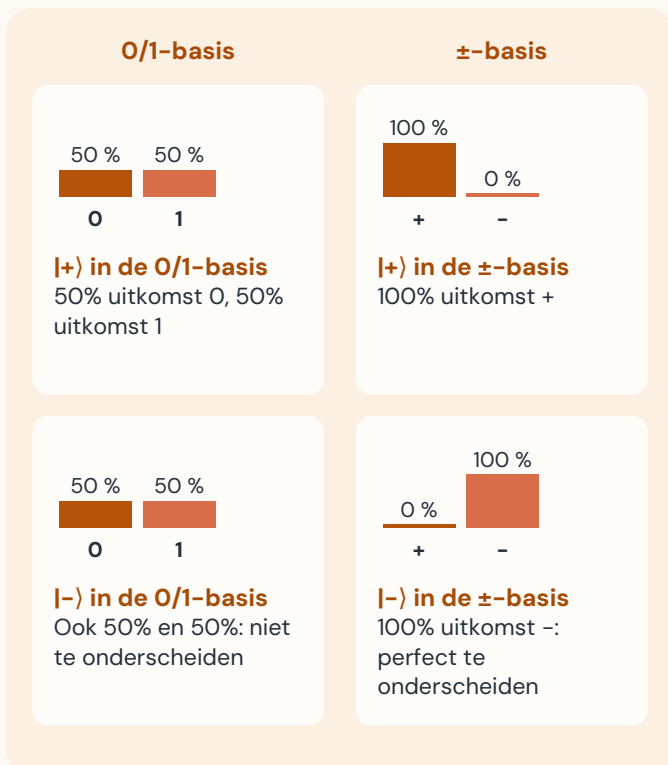


FIG. 4 Model: twee ideale zuivere toestanden; de meetbasis bepaalt welke vraag je stelt.

Amplitude is nog geen waarschijnlijkheid

Een quantumtoestand wordt beschreven met amplituden: complexe getallen waarvan het kwadraat van de absolute waarde een kans geeft. Voor een qubit $a|0\rangle + b|1\rangle$ geldt $|a|^2 + |b|^2 = 1$. Bij een deeltje in de ruimte geeft $|\psi|^2$ een kansdichtheid; de kans voor een gebied volgt door te integreren. Het verschil met gewone kansen: amplituden kunnen elkaar versterken of uitdoven. Daarom bewaren losse kansen niet alle informatie.

De meetopzet bepaalt de vraag

Toestand en meetopzet bepalen samen de voorspelling. Een meting is geen blik in een vaste lijst eigenschappen, maar een fysische koppeling tussen instrument en systeem die een uitkomst oplevert. Over wat er tussen voorbereiding en uitkomst 'echt' gebeurt, verschillen interpretaties. Over de voorspelde frequenties niet: die zijn met grote precisie getoetst.

Rekenoefening: één poort maakt de fase zichtbaar

Neem $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ en $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$. Meet je in de O/1-basis, dan geven beide 50% uitkomst 0 en 50% uitkomst 1. Pas nu eerst een Hadamard-poort H toe: $H|+\rangle = |0\rangle$ en $H|-\rangle = |1\rangle$. Een meting in de O/1-basis geeft nu met zekerheid 0 respectievelijk 1. Het minteken was dus geen negatieve kans, maar een relatieve fase die in een andere basis de hele uitkomst bepaalt.



Een geleerd wereldmodel voorspelt uit beschikbare waarnemingen. De quantumfysica maakt de informatiegrens scherp: dezelfde 0/1-kansen kunnen bij verschillende toestanden horen. Volg amplituden, basis en correlaties; extra rekenkracht kan een oninformatieve meetopzet niet vanzelf vervangen.

VERVOLG • QUANTUM

Correlaties voorbij klassieke kaarten

Een model moet passen bij de soort correlatie die wordt gemeten. De CHSH-grens maakt dat onderscheid concreet: 2 voor vooraf vastgelegde lokale uitkomsten, $2\sqrt{2}$ voor de quantumtheorie.

Drie getallen op één lijn

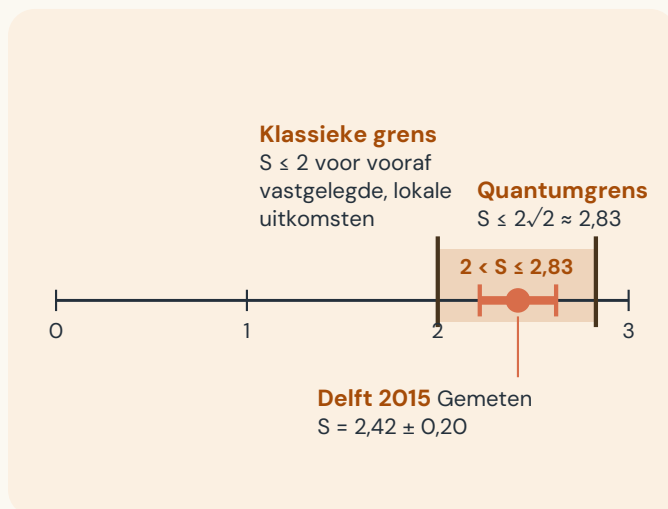


FIG. 5 Klassieke grens en quantumgrens berekend; Delft 2015 gemeten ($\pm$ één standaardafwijking).

Waarom klassiek bij 2 ophoudt

Twee meetstations kiezen elk tussen twee instellingen (a of a', b of b') en noteren +1 of -1. CHSH combineert vier correlaties:

$S = E(a, b) - E(a, b') + E(a', b) + E(a', b')$. Liggen alle uitkomsten vooraf vast, zoals op gedeelde kaartjes, dan is per paar $S = A(B - B') + A'(B + B')$. Eén haakje is 0, het andere ± 2 : elk paar levert ± 2 , dus het gemiddelde blijft tussen -2 en 2 (Bell 1964; Clauser, Horne, Shimony en Holt 1969).

Reken $2\sqrt{2}$ zelf na

Voor fotonparen in de toestand $(|HH\rangle + |VV\rangle)/\sqrt{2}$ voorspelt de quantumtheorie $E(\alpha, \beta) = \cos(2(\alpha - \beta))$. Kies $a = 0^\circ$, $a' = 45^\circ$, $b = 22,5^\circ$ en $b' = 67,5^\circ$. Dan is $E(a, b) = \cos(-45^\circ) \approx 0,707$, $E(a, b') = \cos(-135^\circ) \approx -0,707$, $E(a', b) = \cos(45^\circ) \approx 0,707$ en $E(a', b') = \cos(-45^\circ) \approx 0,707$. Samen: $S = 4 \times 0,707 = 2\sqrt{2} \approx 2,83$.

Gemeten in Delft

Oudere Bell-tests lieten mazen open: detectoren misten deeltjes, of de stations konden nog een lichtsignaal uitwisselen voordat de meting klaar was. In 2015 sloot een team van QuTech (TU Delft) beide tegelijk, met elektronspins op 1,3 km afstand. Gemeten: $S = 2,42 \pm 0,20$ in 245 pogingen ($p = 0,039$). Lokale uitkomsten blijven daarbij willekeurig: de correlatie verschijnt pas als beide lijsten via een gewoon kanaal worden vergeleken. Er reist dus geen bericht sneller dan het licht. ♦



AXI VRAAGT

Alice ziet alleen ruis, Bob ook. Waar zit dan de correlatie – en waarom kan niemand er een bericht mee sturen?

Quiz: vier onzekerheden uit elkaar houden

Een vierkante meter in de variantie is geen vierkante meter vloer. Lees iedere grootheid met haar rol en eenheid.

Drie manieren van mislezen

Standaardafwijking, variantie en een interval zijn verschillende objecten. De meetbegroting op pagina 27 en de foutband op pagina 13 geven genoeg informatie voor alle vier vragen; uitwerkingen op pagina 49.

1 Twee onafhankelijke standaardonzekerheden zijn 0,03 m en 0,04 m. Wat is de gecombineerde standaardonzekerheid?

- ☐ (A) 0,01 m
- ☐ (B) 0,07 m
- ☐ (C) 0,05 m

Waarom?

2 De variantie is 0,008 m². Welke standaardafwijking hoort erbij?

- ☐ (A) Ongeveer 0,0894 m.
- ☐ (B) 0,008 m.
- ☐ (C) 0,000064 m.

Waarom?

3 Een motorstap voegt onafhankelijk $Q = 0,002 \text{ m}^2$ toe aan $P = 0,008 \text{ m}^2$. Wat wordt de variantie?

- ☐ (A) 0,006 m²
- ☐ (B) 0,010 m²
- ☐ (C) 0,0894 m²

Waarom?

4 Een sensor heeft een vaste nulpuntsfout. Wat doet honderd keer dezelfde meting middelen gegarandeerd?

- ☐ (A) Het verwijdert de bias.
- ☐ (B) Het maakt de werkelijkheid exact.
- ☐ (C) Het verwijdert die vaste bias niet.

Waarom?

Jouw antwoorden

1 2 3 4

■ VECTOR • ONDERZOEK MET AANNAMEN

Een meetbegroting past op een halve pagina

Een robotmeting hoeft geen mysterieus kwaliteitscijfer te zijn. Houd de herkomst van onzekerheid in aparte vakjes.

ACHTERGROND

Ruis, bias en resolutie

Ruis wisselt tussen metingen. Bias verschuift een gemiddelde. Resolutie is de stapgrootte van de uitlezing. Een instrument kan fijn afgeronde cijfers tonen en toch een grote bias hebben. NIST TN 1297 behandelt combinaties van onzekerheidscomponenten; een begroting blijft afhankelijk van een geschikt meetmodel.

Bron: 60

Twee aangenomen bijdragen

Stel dat plaatsbepaling een onafhankelijke standaardonzekerheid van 0,03 m uit uitleesruis en 0,04 m uit timing heeft. In dit eigen voorbeeld tellen de varianten op: $u^2 = 0,03^2 + 0,04^2 = 0,0025 \text{ m}^2$, dus $u = 0,05 \text{ m}$. “Onafhankelijk” is het contract waaronder de korte som geldig is.

De vergeten rij

Een vaste nulpuntsafwijking hoort niet stilzwijgend bij nulgemiddelde ruis. Corrigeer een bekende afwijking en neem de onzekerheid van die correctie mee. Als timing en positie uit hetzelfde onzekere kloksignaal ontstaan, is een covariantieterm relevant. De begroting moet dat verband beschrijven voordat je de wortel neemt.

KORTE MEETNOTITIE

Afronden is een redactionele keuze

2,240000 meter oogt preciezer dan 2,24 meter, maar voegt geen waarneming toe.

Bewaar binnenin, kies buitenop

Reken intern met genoeg cijfers om afrondfout klein te houden. Rapporteer vervolgens een hoeveelheid cijfers die bij de onzekerheid past. Schrijf de meeteenheid en de betekenis van de band erbij. Een extra decimaal verandert noch de sensor, noch de steekproef.

Bron: 60

SIMULATIE 02 • FASE EN SUPERPOSITIE

Twée golven, één resultaat

Twée gelijke golven, één faseverschil φ . Voorspel hoe hoog hun som wordt – en ontdek waarom “twee keer een signaal” niet altijd “twee keer de uitwijking” betekent.

Drie faseverschillen bij $A = 1$

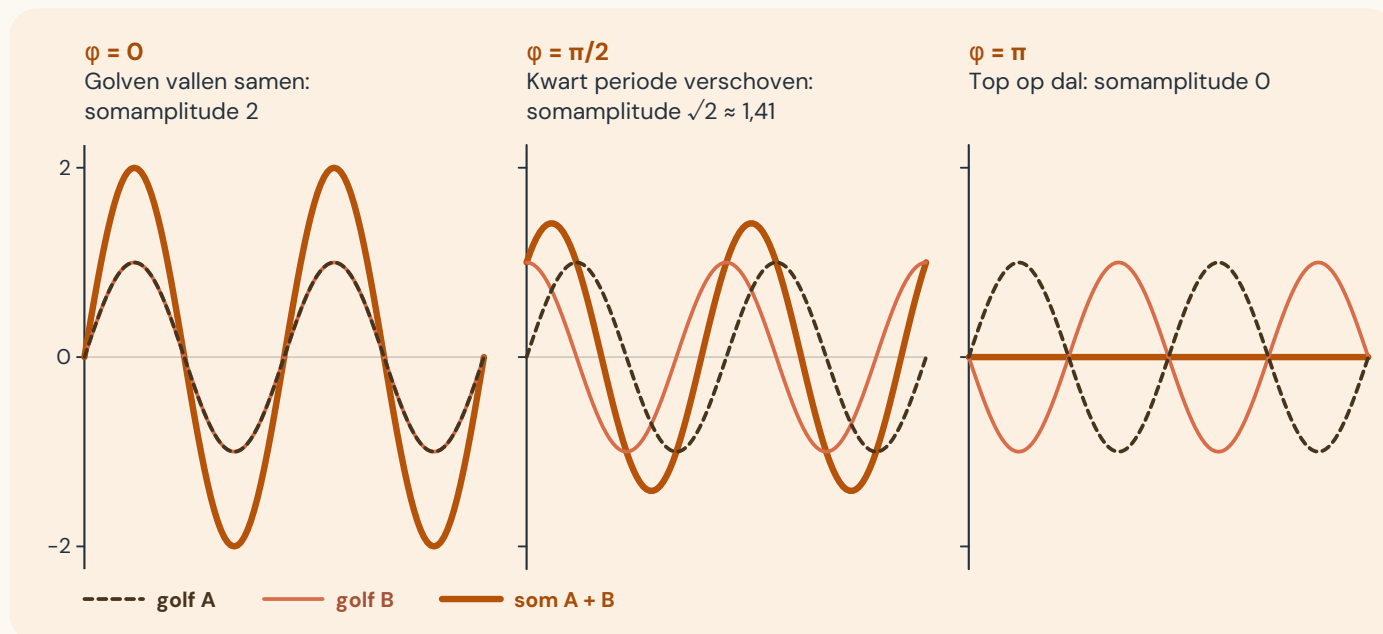


FIG. 1 Berekende somamplitudes $2A \cdot |\cos(\varphi/2)|$ voor ideale golven met dezelfde frequentie.

Superpositie van twee sinusgolven

$y_1 = A \cdot \sin(fx)$, $y_2 = A \cdot \sin(fx + \varphi)$; x en f zijn in dit model dimensieloos.

Wat de drie instellingen doen

De golven zijn $y_1 = A \cdot \sin(fx)$ en $y_2 = A \cdot \sin(fx + \varphi)$, met dezelfde amplitude A en frequentieparameter f . x en f zijn dimensieloos; φ is de fase in radialen. Bij $\varphi = 0$ vallen de toppen samen en is de somamplitude $2A$. Bij $\varphi = \pi/2$ is die $\sqrt{2} \cdot A \approx 1,41 \cdot A$. Bij $\varphi = \pi$ heffen de golven elkaar overall op: de som is nul. Met $A = 1$ en $\varphi \approx 1,57$ ($\pi/2$) zijn beide golven zichtbaar en is hun som 1,41 hoog, niet 2.

De formule achter het beeld

Met een somregel uit de goniometrie wordt $y_1 + y_2 = 2A \cdot \cos(\varphi/2) \cdot \sin(fx + \varphi/2)$. De somamplitude is dus $2A \cdot |\cos(\varphi/2)|$. Een minteken verschuift de fase, niet de grootte. Verhoog f bij dezelfde φ : er verschijnen meer perioden, maar de somamplitude blijft gelijk. Zo scheid je golflengte, fase en hoogte met één vergelijking.

Een precies voorbeeld met grenzen

Het model kent geen damping en geen verschillende frequenties; zwevingen vragen een andere opzet. Destructieve interferentie op één plek laat bovendien geen energie verdwijnen: elders tellen de golven juist op. Wie dat wil volgen, moet bronnen, veld en energiestromen samen beschrijven. Juist de kleine, zichtbare aannames maken deze demonstratie helder.

■ HISTORIE · VAN TURING TOT DE NOBELPRIJZEN

Ideeën groeien op elkaars schouders

Een selectie mijlpalen uit dit nummer: AI, natuurkunde en het werk waarin ze elkaar raken. In 2024 bekroonde de Nobelprijs voor de Natuurkunde de basis van neurale netwerken; de helft van de prijs voor de Scheikunde ging naar de makers van AlphaFold.

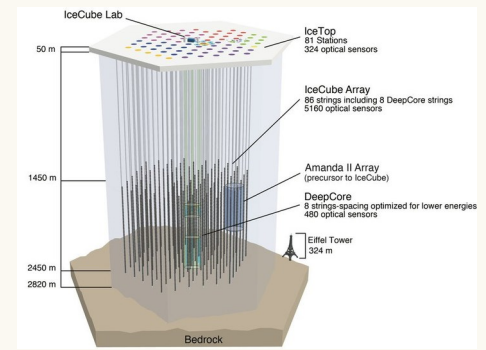


FIG. 1 IceCube: 86 kabels met 5.160 lichtsensoren, tussen 1.450 en 2.450 meter diep in het ijs. Schema: Karen Andeen & Matthias Plum / IceCube Collaboration, CC BY 4.0.

Tien mijlpalen

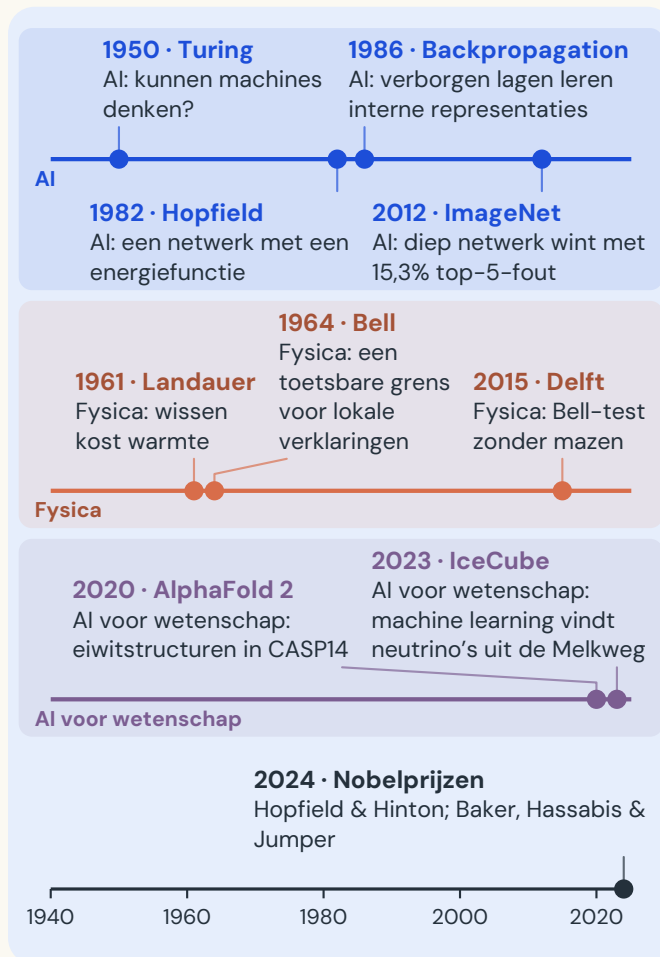


FIG. 2 Jaar van publicatie, wedstrijd of prijs; een selectie, geen volledige geschiedenis.

Een methode is ouder dan haar beroemdste naam

Turing vroeg in 1950 hoe je machine-intelligentie kunt onderzoeken. Backpropagation (1986) maakte leren met verborgen lagen populair, al was het achterwaarts rekenen van afgeleiden al in

1970 beschreven. Gemeten: in 2012 haalde een diep netwerk in de ImageNet-wedstrijd een top-5-fout van 15,3%, tegen 26,2% voor de nummer twee.

2024: neurale netwerken in twee Nobelprijzen

John Hopfield gaf een netwerk van aan-uitneuronen in 1982 een energie, $E = -\frac{1}{2} \sum w_{ij} s_i s_j$, net als een systeem van magnetische spins. Opgeslagen patronen zijn dalen in dat landschap: een vervormd beeld rolt terug naar het dichtstbijzijnde patroon. Geoffrey Hinton bouwde daarop de Boltzmannmachine, met kansen uit de statistische fysica. Samen kregen ze de Nobelprijs voor de Natuurkunde 2024. Een dag later ging de helft van de prijs voor de Scheikunde naar Hassabis en Jumper voor AlphaFold, de andere helft naar David Baker voor computationeel eiwitontwerp.

En de meetpraktijk gaat door

IceCube, een kubieke kilometer ijs met lichtsensoren onder de Zuidpool, rapporteerde in 2023 een neutrinosignaal uit het melkwegvlak met een significantie van $4,5\sigma$. Machine learning hielp de analyse van waarnemingen. Het resultaat hoort bij de gepubliceerde meet- en analysemethode; de detectorafbeelding maakt die fysieke meetroute zichtbaar. ♦



AXI Een netwerk dat zijn energie minimaliseert, kreeg een Nobelprijs natuurkunde. Is AI nu fysica — of was fysica altijd al een beetje AI?



RUIMTETIJD · DE LICHTKLOK

De geometrie van een tik

Een lichtklok verbindt beweging met tijd. Bij $0,6c$ hoort $\gamma = 1,25$: vier seconden op de bewegende klok tegen vijf op de klok van de waarnemer.

Eén tik, twee lichtwegen

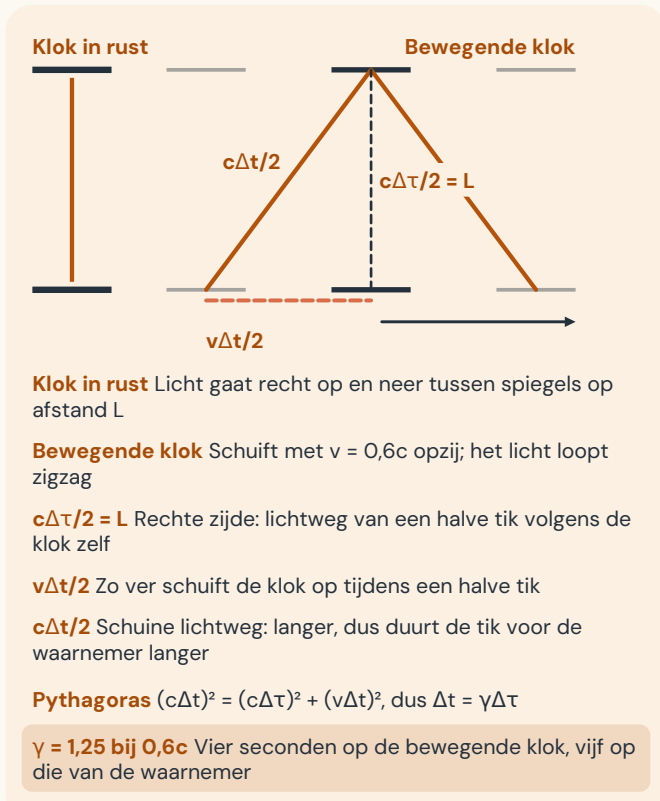


FIG. 1 Model: ideale lichtklok, op schaal getekend voor $v = 0,6c$. Ter vergelijking: bij $0,1c$ is $\gamma \approx 1,005$, bij $0,8c \approx 1,67$.

01 Twee klokken, verschillende trajecten

Speciale relativiteit koppelt ruimte- en tijdcoördinaten aan relatieve beweging. De schaal is de Lorentzfactor $\gamma = 1/\sqrt{1 - v^2/c^2}$. Bij $v = 0,6c$ is $\gamma = 1,25$. Wie 'echt' beweegt, is geen zinvolle vraag; wat telt, is welke gebeurtenissen je vergelijkt. Ontmoeten twee klokken elkaar na verschillende reizen, dan kun je hun verstreken tijd direct naast elkaar leggen – en dan blijkt die verschillend.

02 De geometrie van één tik

Een lichtklok tikt als een lichtflits tussen twee spiegels op afstand L heen en weer gaat. In haar eigen ruststelsel duurt een halve tik $\Delta\tau/2 = L/c$. Een waarnemer die de klok met snelheid v ziet bewegen, ziet het licht schuin lopen: de lichtweg $c\Delta t/2$ is de schuine zijde van een rechthoekige driehoek met zijden $c\Delta\tau/2$ en $v\Delta t/2$. Pythagoras geeft $(c\Delta t)^2 = (c\Delta\tau)^2 + (v\Delta t)^2$, dus $\Delta t = \gamma\Delta\tau$. Een GPS-satelliet beweegt met $3,87 \text{ km/s}$: $v/c \approx 1,3 \times 10^{-5}$ en $\gamma - 1 \approx 8,3 \times 10^{-11}$. Piepklein – maar over een dag is dat $7,2$ microseconden.

ACHTERGROND

Gemeten: klokken rond de wereld

In oktober 1971 namen Joseph Hafele en Richard Keating vier cesiumklokken mee op lijnvluchten rond de aarde, eerst oostwaarts en daarna westwaarts.

Oostwaarts

De vliegende klokken verloren (59 ± 10) ns ten opzichte van de klokken op de grond.

Westwaarts

Ze wonnen (273 ± 7) ns: snelheid en hoogte werkten nu dezelfde kant op.

Vervolg op pagina 31

Een GPS-satelliet ondergaat dezelfde twee effecten, elke dag opnieuw.

Bron: 48



■ RUIMTETIJD · GPS

Een klokcorrectie om de aarde

Hoger tikt sneller, sneller tikt trager. Netto lopen GPS-klokken zo'n 38 microseconden per dag voor – en het systeem is daar vanaf de fabriek op ingericht.

Een afstand uit een tijdsignaal

Een ontvanger meet hoe laat signalen van minstens vier satellieten vertrokken en berekent daaruit zijn positie én zijn eigen klokfout. Licht legt in één nanoseconde ongeveer dertig centimeter af, dus elke nanoseconde klokfout telt. Atmosfeer, baanfouten en de stand van de satellieten tellen ook mee, maar de klokken zijn het fundament.

Twee effecten, één rekening

De satellieten draaien op zo'n 20.200 km hoogte met 3,87 km/s. Berekend met de relativiteitstheorie: hoger in het zwaarteveld tikt een klok sneller, +45,8 microseconden per dag. De snelheid laat haar trager tikken: -7,2 microseconden per dag. Netto loopt een GPS-klok ongeveer 38,6 microseconden per dag voor op een klok op de grond. Daarom krijgt elke satellietklok vóór de lancering een frequentie van 10,22999999543 MHz in plaats van 10,23 MHz. De ontvanger corrigeert daarnaast een klein, wisselend restant door de enigszins elliptische baan: hooguit 23 nanoseconden bij een excentriciteit van 0,01.

Gemeten in 1977

De eerste cesiumklok in een baan, aan boord van de testsatelliet NTS-2, draaide eerst zo'n twintig dagen zonder correctie. Gemeten: hij liep $442,5 \times 10^{-12}$ sneller dan klokken op de grond; de algemene relativiteitstheorie voorspelde voor die baan $446,5 \times 10^{-12}$. Een toets van ongeveer 1%. Vermenigvuldig de netto 38,6 microseconden met c en je krijgt ruim 11 kilometer lichtweg per dag: de schaal van de fout als niemand zou corrigeren. Een geleerd navigatiemodel dat goede posities voorspelt, heeft relativiteit dus niet zelf ontdekt – het systeem had de fysica al ingebouwd. ♦

“Licht legt in één nanoseconde ongeveer dertig centimeter af, dus elke nanoseconde klokfout telt.”



FIG. 2 Artistieke impressie van een GPS III-satelliet (U.S. Air Force, publiek domein). Aan boord tikken atoomklokken.

DATAPUNT

+45,8 μ s/dag

zwaartekracht: hoger in het zwaarteveld tikt een klok sneller (berekend)

DATAPUNT

-7,2 μ s/dag

snelheid: bij 3,87 km/s tikt een klok trager (berekend)

DATAPUNT

$\approx$ +38,6 μ s/dag

netto voorsprong van een GPS-klok op een klok op de grond (berekend, Ashby 2003)

METEN IS AL MODELWERK

Vraag bij een dataset hoe de “grondwaarheid” ontstond. Ook een referentiewaarde kan instrumenten, correcties en aannames bevatten.



AXI Op 9.545 km van het middelpunt van de aarde heffen beide effecten elkaar precies op. Zou jij daar je klok parkeren?

Vier meetkaartjes, één sluitend spoor

Logica, eenheden en volgorde worden puzzelmateriaal. Antwoorden met bewijs staan op pagina 48.

1 Een lineaire sensor geeft $z=ax+b$. Foutloze ijkpunten zijn $(x,z)=(0,2)$ en $(3,8)$. Welke echte x hoort bij $z=12$?

Mijn antwoord

3 Begin met $P=0,04 \text{ m}^2$. Operatie M halveert P ; operatie B voegt $0,01 \text{ m}^2$ toe. Je mag beide precies één keer uitvoeren. Welke volgorde geeft de kleinere eindvariantie?

Mijn antwoord

2 Een rekenbegroting telt varianten $0,01 \text{ m}^2$ en $0,04 \text{ m}^2$. Een redacteur schrijft $\sqrt{0,01} + \sqrt{0,04} = 0,30 \text{ m}$ als gecombineerde standaardonzekerheid. Wat is onder onafhankelijkheid het juiste resultaat?

Mijn antwoord

Schrijf de tussentreden op

De optelsommen zijn gekozen om zonder computer oplosbaar te zijn. Bij het middelste

raadsel betekent onafhankelijk uitsluitend dat de opgegeven ruiscomponenten geen covariantie hebben.

Ontwerp een eerlijk sensorlogboek

Maak een onderzoeksblad voor een fictieve railrobot. Je kunt de opdracht volledig op papier uitvoeren met zelfgekozen, duidelijk gelabelde data.

01 Leg de afspraak vooraf vast

Kies één constante snelheid en één afwijkende sliprit. Noteer x^- , P^- , z , R en een aparte referentie per stap. Laat een lege plek voor een gedeelde timingfout: die aanname hoort zichtbaar te blijven. Je doel is niet de mooiste foutband, maar een reconstructeerbare correctie.

02 Schrijf een tegenproef

Bedenk een rit waarbij een perfect aangeleerde oude beweging niet klopt: het karretje stopt eerder of glijdt verder. Welke innovatie verwacht je? Welke instelling zou je na die waarneming veranderen? Houd daarna een nieuwe rit apart om de verandering te beoordelen. Een tegenvoorbeeld is hier een nuttige onderzoeksuitkomst.

Mijn meetcontract

Vul in: aangenomen beweging; onafhankelijke referentie; mogelijke gedeelde fout; verwachte innovatie; apart gehouden eindrit.

Mijn antwoord

Vier sommen, één eenhedencheck

Nieuwe getallen, dezelfde fysica. Schrijf eerst de relatie op, dan de eenheden en pas daarna het getal. Zo vind je een fout terug in de juiste stap.

Eenheden als tweede taal

Kilogram maal meter per seconde kwadraat is newton; een joule is $\text{kg} \cdot \text{m}^2/\text{s}^2$. Een eenhedencheck bewijst niet dat je model klopt, maar maakt een onmogelijke combinatie meteen zichtbaar. Zoek bij elk getal welke grootte het voorstelt: massa, snelheid, kracht, tijd of kans. Schrijf $\approx$ bij een afgerond antwoord. Een exact gedefinieerde constante zoals c blijft exact, ook als je met drie cijfers rekent.

- 1 Een karretje van 0,5 kg staat stil en krijgt 3 s lang een resulterende kracht van 2 N. Hoe snel gaat het daarna?**

- (A) 4 m/s
(B) 6 m/s
(C) 12 m/s

Reken hier

- 2 Een test vindt 95% van de zieken en geeft bij 2% van de gezonden vals alarm. Eén op de 200 mensen is ziek. Hoe groot is de kans op ziekte na een positieve test?**

- (A) Ongeveer 19%
(B) Ongeveer 49%
(C) Ongeveer 95%

Reken hier

- 3 Wat is de Landauergrens voor het wissen van 1 kilobyte (8.000 bits) bij 77 K, de temperatuur van vloeibare stikstof? Gebruik $k = 1,38 \times 10^{-23} \text{ J/K}$ en $\ln 2 \approx 0,693$.**

- (A) $7,4 \times 10^{-22} \text{ J}$
(B) $5,9 \times 10^{-18} \text{ J}$
(C) $2,3 \times 10^{-17} \text{ J}$

Reken hier

- 4 Een satelliet in een lage baan beweegt met 7,5 km/s. Hoeveel loopt zijn klok per dag achter door alleen de snelheid? Gebruik $\Delta t \approx (v^2/2c^2) \times 86.400 \text{ s}$ en $c = 3,0 \times 10^8 \text{ m/s}$.**

- (A) 7,2 μs
(B) 2,7 μs
(C) 27 μs

Reken hier

Jouw antwoorden

1 2 3 4

+ ANTWOORDEN OP PAGINA 47

De uitwerkingen tonen elke tussenstap. Een juist getal met een verkeerde eenheid is nog geen juist antwoord.

■ ONDERZOEKSNOTITIE · ZES STAPPEN

Van interessante vraag naar kleine toets

Kies één relatie uit dit nummer en onderzoek haar zelf: een voorspelling, één verandering, een eerlijke conclusie.

1

Maak de vraag kleiner

“Begrijpt AI de natuur?” is te groot. “Daalt het verlies bij deze vijf leerstappen?” is toetsbaar. Kies invoer, uitkomst en maatstaf. De leerstapsimulatie op pagina 15 is een goed begin: de formule staat erbij en elke uitkomst is na te rekenen. Bij een fysische proef volstaat vaak een eenvoudige opzet; onderzoek begint met afbakening, niet met dure materialen.

2

Bewaar beslissingen, niet alleen successen

Schrijf je voorspelling vooraf op en bewaar ook onverwachte uitkomsten. Pas je je verklaring aan, noteer dan welke waarneming daartoe leidde. Een curve zonder instellingen is weinig waard. Zo kun je later rekenfout, verkeerde interpretatie en onjuiste aanname uit elkaar houden.

3

Een conclusie op maat

Schrijf: “Binnen deze instellingen zagen we X; dat past bij Y; situatie Z is niet onderzocht.” Zelfs een start bij $x = 0$ die op 0 blijft, is een klein controleerbaar resultaat. Kies daarna één vervolgstap die echt nieuwe informatie oplevert. De grootte van een claim moet passen bij de toets.

Zo doe je het

- 1 Schrijf één vraag op en de grootte die je wilt beoordelen.
- 2 Leg de aannames, de systeemgrens en de gebruikte formule vast.
- 3 Noteer je voorspelling voordat je meet of rekent.
- 4 Voer een kleine vergelijking uit waarin je één instelling verandert.
- 5 Bewaar uitkomsten, eenheden en afwijkingen van je verwachting.
- 6 Formuleer de conclusie, de beperking en één zinvolle volgende toets.



AXI'S LABBOEK

Eerst voorspellen, dan kijken. Een onverwacht resultaat is een nieuwe vraag, geen reden om de oude notitie te wissen.



AXI Mijn mooiste notitie ooit: ‘Voorspelling fout. Waarom?’ Daar begon elk goed experiment.



Onzekerheid tellen in bits

Een taalmodel verdeelt kans over mogelijke vervolgen. Shannon-entropie beschrijft onzekerheid van die verdeling, niet hoeveel rommel in een computerkast ligt.

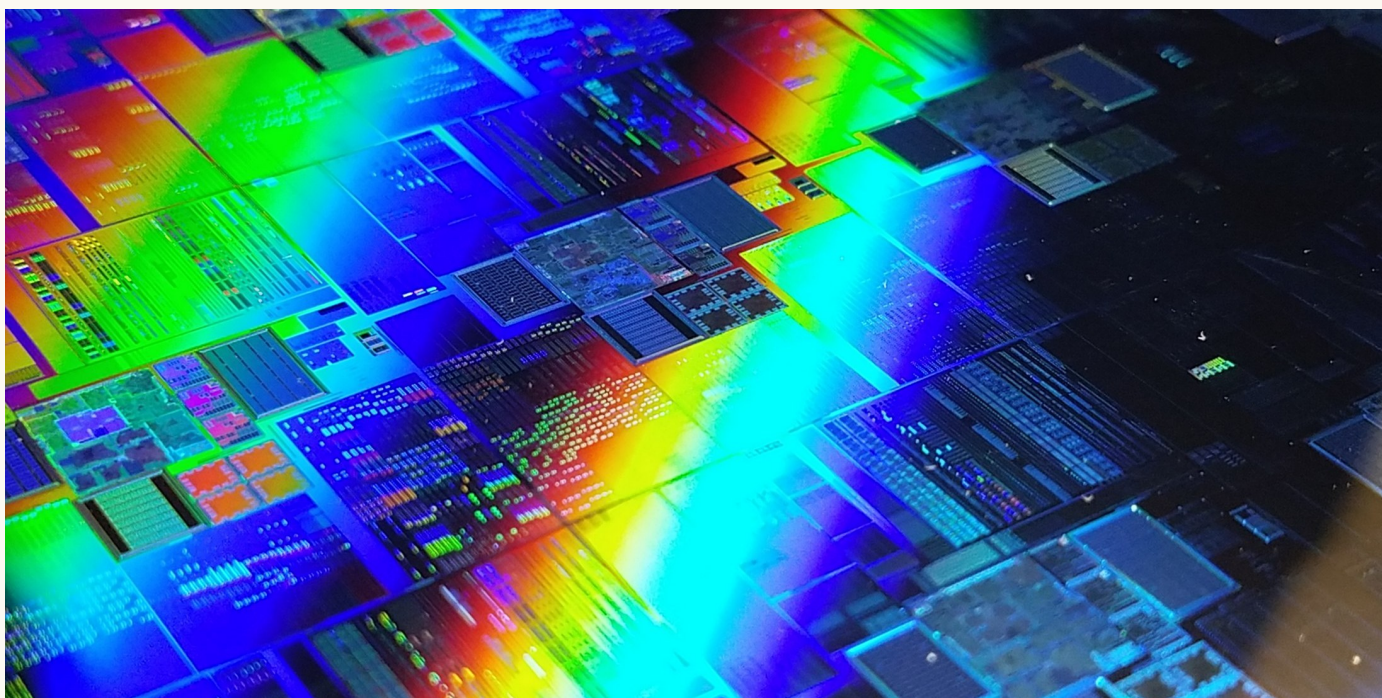


FIG. 1 Een siliciumwafer • documentaire foto, geen beeld van modelonzekerheid of van individuele gewiste bits.

Een verdeling krijgt een getal

Shannon-entropie is $H = -\sum p_i \log_2 p_i$, in bits. Twee even waarschijnlijke mogelijkheden hebben $H=1$. Als één mogelijkheid met zekerheid bekend is, is $H=0$. Honderd onafhankelijke eerlijke bits hebben $H=100$ en 2^{100} mogelijke reeksen. Deze telling gaat over een beschreven bronverdeling. Meer tekens, meer bestanden of ingewikkelder afbeeldingen betekent niet automatisch meer entropie: de kansverdeling en haar aannamen bepalen de rekening.

Een taalvoorspelling is geen waarheidstoets

Een model kan vrijwel alle kans op één volgend token leggen en daardoor een lage voorspellingsentropie hebben. Dat maakt de uitkomst niet automatisch feitelijk juist. Het getal beschrijft modeltwijfel binnen die verdeling.

Trainingsdoelen kunnen modelkansen met de waargenomen tekst vergelijken, terwijl een onafhankelijke evaluatie moet onderzoeken of de volledige toepassing doet wat nodig is. Zeker klinken en goed presteren zijn verschillende onderzoeksvragen.

Thermodynamica vraagt een andere grens

Fysische entropie heeft eenheid J/K; Shannon-entropie heeft eenheid bit. De tweede hoofdwet gaat over het geïsoleerde geheel. Een koelkast mag zijn binnenkant afkoelen zolang de omgeving meer entropie wint. Landauers wisproces op pagina 37 verbindt informatie met een fysieke omgeving onder concrete voorwaarden. Je kunt de logaritme herkennen zonder de twee grootheden of een modeltoken en een gewiste bit gelijk te stellen.



■ LANDAUER · DE PRIJS VAN ÉÉN BIT

De ondergrens van vergeten

Wis een onbekende bit en je betaalt minstens $kT \ln 2$ aan warmte: $2,9 \times 10^{-21}$ joule bij kamertemperatuur. Experimenten kwamen er vlak bij. Chips zitten er nog ver boven.

Wat de formule zegt

Rolf Landauer (IBM) betoogde in 1961 dat logisch onomkeerbaar wissen warmte kost. Zet je een bit die met gelijke kans 0 of 1 is op een vaste 0, dan verdwijnt één bit informatie en moet de entropie van de omgeving minstens $k \ln 2$ stijgen. Bij temperatuur T hoort daar een warmteafgifte van minstens $kT \ln 2$. Met de constante van Boltzmann $k = 1,380649 \times 10^{-23}$ J/K en $T = 300$ K is dat $2,87 \times 10^{-21}$ joule: 2,9 zeptojoule.

Gemeten: de grens is bereikbaar

In 2012 sloot een team in Lyon één colloïdaal bolletje op in een dubbele potentiaalput: links is 0, rechts is 1. Hoe trager ze wisten, hoe dichter de gemiddelde warmte bij $kT \ln 2$ kwam (Bérut e.a., Nature). In 2016 maten onderzoekers in Berkeley bij het resetten van nanomagnetische bits $(4,2 \pm 0,9)$ zeptojoule: binnen twee standaardafwijkingen van de grens van 2,87 zeptojoule (Hong e.a., Science Advances).

Hoe ver zit een chip van de grens?

Een overzicht van Ho, Erdil en Besiroglu (2023) zet de kosten in moderne processors op een rij. Het schakelen van één CMOS-poort (de standaardschakeling in vrijwel alle chips) kost 100 attojoule tot 100 femtojoule: 35 duizend tot 35 miljoen keer $kT \ln 2$. Eén bit uit het werkgeheugen (DRAM) ophalen kost rond 5 picojoule, zo'n 1,7 miljard keer de grens. Verplaatsen en schakelen domineren de rekening, niet het wissen zelf. ♦

DATAPUNT

$$2,87 \times 10^{-21} \text{ J}$$

Landauergrens per gewiste bit bij 300 K (berekend)

DATAPUNT

$$(4,2 \pm 0,9) \text{ zJ}$$

gemeten per reset van een nanomagnetische bit (2016)

GRENS, GEEN ENERGIEMETER

Tokens zijn geen gewiste bits en een chip is geen ideaal wisexperiment. Wie het verbruik van AI wil kennen, meet aan het stopcontact.



AXI Vergeten kost dus altijd iets. Is dat eindelijk een goed excuus om nooit meer iets te wissen?

SCHATTING

Een gigabyte vergeten

Hoeveel warmte kost het minimaal om 1 gigabyte te wissen — en hoeveel kost het om die bits te verplaatsen?

Reken mee

1 GB is 8×10^9 bits. Maal $2,87 \times 10^{-21}$ J geeft $2,3 \times 10^{-11}$ joule: zo weinig dat geen thermometer

het zou merken. Dezelfde 8×10^9 bits één keer uit DRAM lezen kost bij 5 pJ per bit ongeveer 0,04 joule, zo'n 1,7 miljard keer meer. Wie rekenen zuiniger wil maken, wint dus vooral bij het verplaatsen en schakelen van bits.

Bronnen: 37, 39

Quiz: vier checks vóór de demonstratie

Een evaluatie krijgt inhoud zodra je kunt aanwijzen welke uitkomst door welke proef zou worden tegengesproken.

Een aannamencontract

Voor deze vragen zijn alle data fictief. De casus op pagina 44 rapporteert fouten per rit; de strip op pagina 18 toont één mogelijke afhankelijkheid. Oplossingen staan op pagina 49.

- 1 Vier absolute fouten zijn 0,10; 0,20; 0,00; 0,10 m. Wat is hun gemiddelde?**

- (A) 0,05 m
(B) 0,10 m
(C) 0,40 m

Waarom?

- 2 Van twintig gerapporteerde intervallen bevatten zestien de referentiepositie. Welke geobserveerde dekking is dat?**

- (A) 80%
(B) 95%
(C) 16%

Waarom?

- 3 Je stelt een filter bij nadat je de sliptest zag. Waarop hoort de nieuwe eindbeoordeling te gebeuren?**

- (A) Dezelfde sliptest blijft volledig onaangeroerd bewijs.
(B) Alleen op trainingsritten.
(C) Op apart gehouden nieuwe evaluatieritten.

Waarom?

- 4 Twee thermometers geven hetzelfde resultaat maar werden tegen dezelfde verkeerd ingestelde referentie geijkt. Wat bewijst overeenstemming?**

- (A) Hun fouten zijn zeker nul.
(B) Alleen dat hun waarden hier overeenkomen.
(C) Dat hun ijking onafhankelijk was.

Waarom?

Jouw antwoorden

1 2 3 4

De kopie die zekerheid leek te verdubbelen

Twee kleine gedachtewerelden vragen hoeveel informatie er werkelijk bijkomt. Hun voorwaarden zijn expliciet, zodat je de conclusie kunt veranderen wanneer een voorwaarde verandert.

De gespiegelde sensor

Een echte sensor maakt één meting z . Een programma kopieert z naar twee identieke schermen en noemt die sensoren A en B. Alle getallen op beide schermen zijn altijd gelijk.

Levert het tweede scherm een tweede onafhankelijke waarneming?

Denk aan:

- Waar ontstaat de ruis?
- Wat verandert als een tweede fysiek instrument afzonderlijk meet?

Mijn voorspelling en waarom

De foutbalk zonder wereld

Filter A meldt $2,24 \pm 0,09$ m. Filter B meldt $2,24 \pm 0,01$ m. Er zijn nog geen referentiemetingen; beide intervallen zijn door hun makers opgegeven.

Welke is aantoonbaar betrouwbaarder?

Denk aan:

- Welke waarneming zou een te smal interval onthullen?
- Moet je dekking en breedte apart bewaren?

Mijn voorspelling en waarom

CASUS · NOBELPRIJS SCHEIKUNDE 2024

AlphaFold: wat 0,96 ångström wél en niet zegt

In de blinde toets CASP14 lag de typische afwijking van AlphaFold 2 onder de breedte van één atoom. In 2024 leverde dat Demis Hassabis en John Jumper de Nobelprijs voor de Scheikunde op.

VERMELD DE VERSIE

Deze cijfers gelden voor AlphaFold 2 in CASP14. Latere systemen en andere eiwittaken vragen hun eigen evaluatie.



AXI Mediaan 0,96 ångström. Prachtig — maar is het stuk dat jij nodig hebt donkerblauw of oranje?

ACHTERGROND

Twee domeinen, twee soorten vertrouwen

Stel dat een voorspeld eiwit uit twee domeinen bestaat. Elk domein heeft een hoge pLDDT, maar de PAE tussen de domeinen is groot. Dan zijn de vormen overtuigend, hun onderlinge stand niet. Een vraag over één domein kan het model dus goed beantwoorden; een vraag over het contact tussen de twee vraagt ander bewijs.

Bronnen: 27, 28

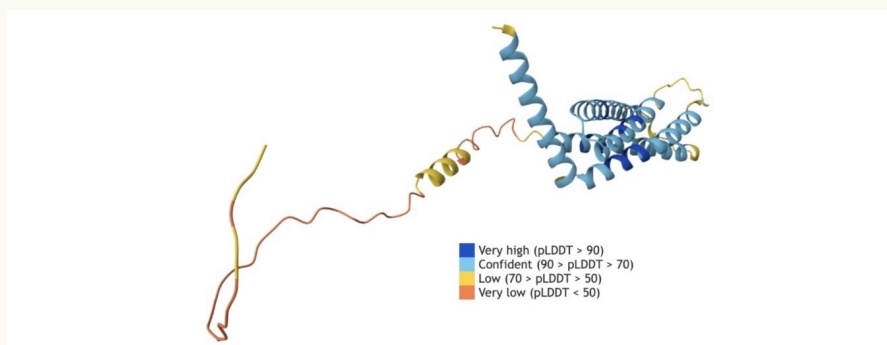


FIG. 1 AlphaFold-voorspelling van het menselijke eiwit TMEM253, gekleurd naar pLDDT: donkerblauw boven 90, lichtblauw 70–90, geel 50–70, oranje onder 50. De gevouwen kern is betrouwbaar, de lange staart niet. Beeld: Liu02511, CCO.

Welke vraag werd beantwoord?

CASP (Critical Assessment of protein Structure Prediction) is een tweejaarlijkse blinde toets: teams voorspellen structuren die al experimenteel bepaald zijn, maar nog niet openbaar. Gemeten in CASP14 (2020): de mediane afwijking van de ruggengraat was 0,96 ångström, tegen 2,8 Å voor de op één na beste methode. Een koolstofatoom is ongeveer 1,4 Å breed. De mediane GDT-score — ruwweg het percentage aminozuren dicht bij de juiste plek — was 92,4 op 100. AlphaFold 2 combineert uitlijningen van verwante eiwitsequenties met een netwerk dat geometrie afdwingt. De andere helft van de Nobelprijs ging naar David Baker, die met computers nieuwe eiwitten ontwerpt.

Het model laat zijn vertrouwen lokaal zien

Bij elk aminozuur geeft AlphaFold een pLDDT-score van 0 tot 100 (predicted local distance difference test): boven 90 zeer betrouwbaar, onder 50 nauwelijks. De predicted aligned error (PAE) zegt hoe zeker de onderlinge stand van delen is. Wat de voorspelling niet zegt: hoe het eiwit beweegt, wat het doet of hoe het bindt. Lage zekerheid kan wijzen op een van nature ongevouwen stuk, maar dat vraagt een eigen toets. AlphaFold 3 (2024) breidde het werk uit naar complexen met DNA, RNA en kleine moleculen; deze casus gaat over versie 2. ♦

FYSICA IN HET MODEL · DE NATUURWET ALS EIS

Een natuurwet in de verliesfunctie

Een vergelijking begrenst de oplossing, een foutmaat begrenst de conclusie en een nieuwe meting toetst het model. Begin bij $u' + u = 0$.

Residu nul is niet genoeg

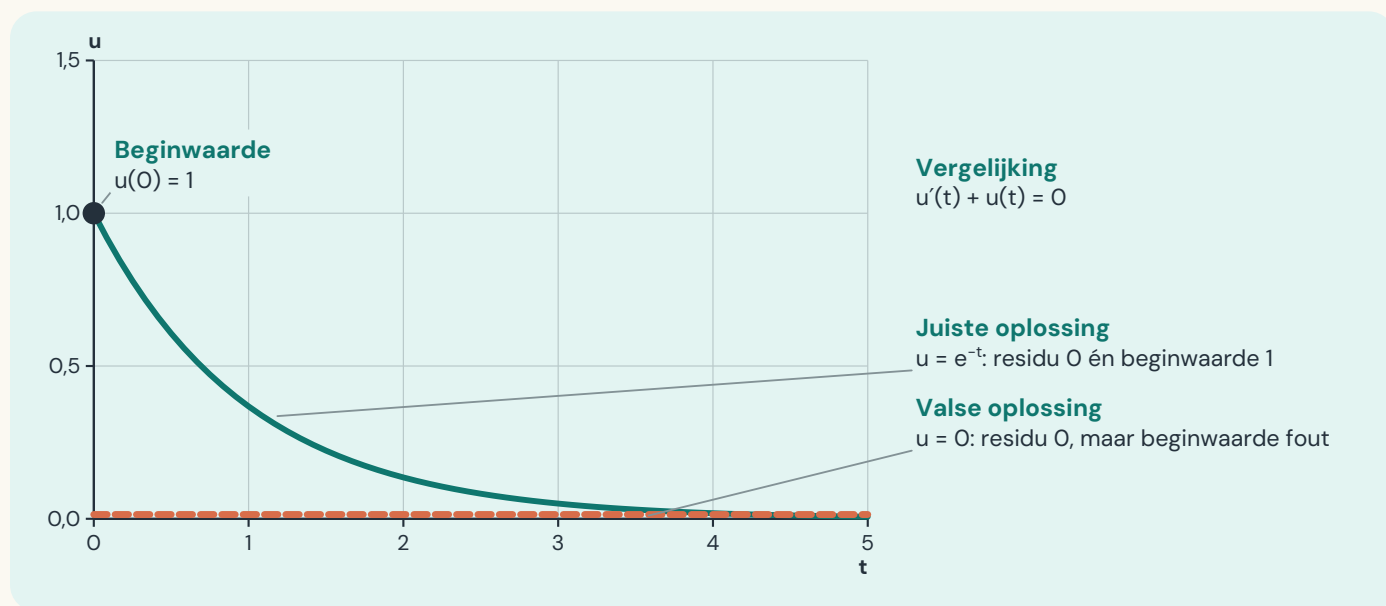


FIG. 1 Model: dimensieloos oefenprobleem met exacte referentie, getekend voor $0 \leq t \leq 5$.

Drie manieren om fysica mee te nemen

Je kunt een netwerk meetuitkomsten laten voorspellen zonder bewegingsvergelijking. Je kunt een bestaande numerieke oplosser aanvullen met een geleerd onderdeel, bijvoorbeeld voor turbulentie. Of je gebruikt een bekende vergelijking als eis tijdens het trainen. De aanpakken verschillen in hoeveel kennis ze vooraf meekrijgen en wat ze garanderen. Zeg dus precies wat het model doet: een materiaalconstante schatten, een heel veld voorspellen of een oplosser corrigeren? Pas dan is een prestatievergelijking eerlijk.

Een residu als extra meetlat

Een physics-informed neural network (PINN, Raissi e.a. 2017) telt in zijn verlies drie dingen op: de afwijking van metingen, de afwijking van

begin- en randvoorwaarden, en het residu – hoeveel de vergelijking op gekozen punten niet klopt. Automatische differentiatie levert de afgeleiden van het netwerk. Omdat de termen tegen elkaar worden afgewogen, is de natuurwet een zachte eis: het netwerk mag ruilen. En een klein residu op gekozen punten garandeert nog niets over de punten daartussen.

Kennis zichtbaar houden

Neem het oefenprobleem $u'(t) + u(t) = 0$ met $u(0) = 1$, dimensieloos. De exacte oplossing is $u(t) = e^{-t}$. Een PINN kan datafout, beginvoorwaarde en residu $r(t) = u'(t) + u(t)$ in één verlies combineren. Alleen r klein maken is onvoldoende: ook $u(t) = 0$ heeft residu nul, maar de verkeerde beginwaarde. Alleen $u(0) = 1$ eisen laat oneindig veel verkeerde curves toe. Pas samen leggen ze de oplossing vast.

FYSICA IN HET MODEL · FOUTMATEN

De fout verandert met de schaal

Dezelfde oplossing e^{-t} krijgt een offset van 0,1. Dat brengt ons van een trainingsvoorwaarde naar de foutmaat die voor een toepassing werkelijk telt.

Overall 0,1 eraast, niet overall even erg

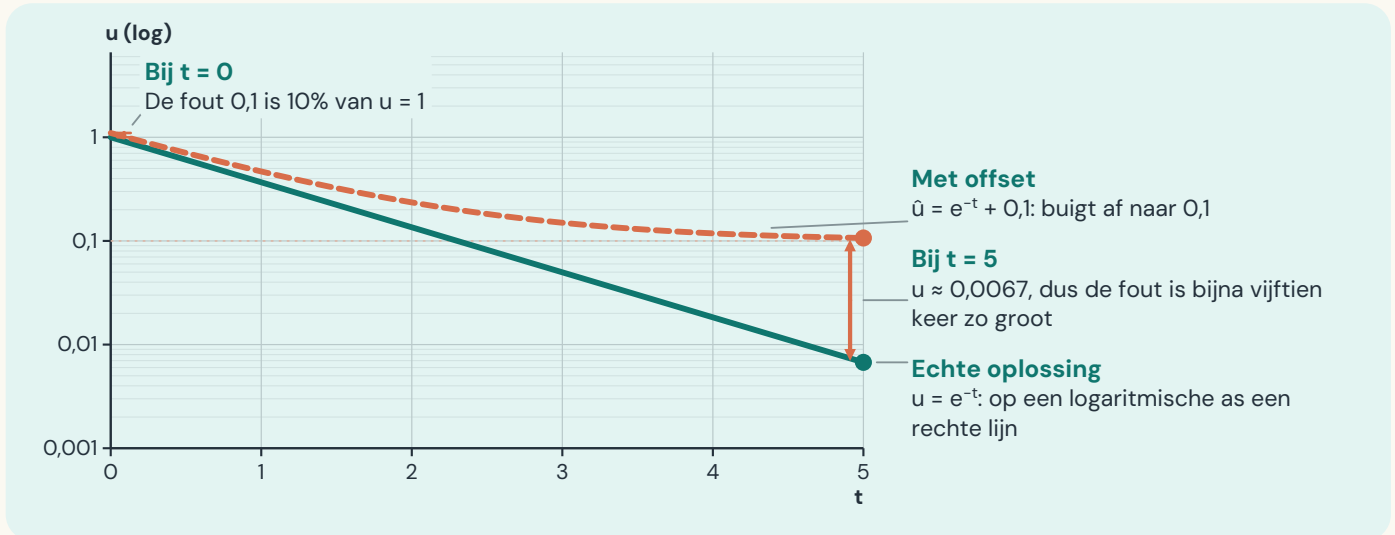


FIG. 2 Model: dimensieloos oefenprobleem, getekend voor $0 \leq t \leq 5$. Verticale as logaritmisch: elke streep is een factor 10.

SNELHEID HEEFT EEN NOEMER

Rapporteer welke oplosser, foutgrens, hardware en taak zijn vergeleken — en of de trainingstijd is meegeteld.



AXI Duizend keer sneller — inclusief de uren training ervoor, of pas daarna?

“Die winst geldt per voorspelling, na de training: voor veel verwante gevallen loont dat, voor één nieuwe situatie zelden.”

Wanneer een zachte eis een lastig landschap maakt?

Krishnapriyan en collega's (2021) lieten zien dat PINN's al bij iets lastigere convectie-, reactie- en diffusieproblemen kunnen vastlopen. Niet omdat het netwerk de oplossing niet kan voorstellen, maar omdat de training haar niet vindt: het residu maakt het probleem slecht geconditioneerd, zodat kleine veranderingen in de gewichten grote sprongen in het verlies geven. Gemeten: met een leercurriculum — eerst een makkelijke versie van de vergelijking, dan steeds lastiger — of door de tijd in stukken te leren, werd de fout tot 10 à 100 keer kleiner.

Een operator leert een hele familie?

Een neural operator leert een afbeelding tussen functies: van een beginveld naar het veld op een later tijdstip. De Fourier Neural Operator (Li e.a., 2020) doet dat in het frequentiedomein en werd getest op drie partiële differentiaalvergelijkingen (PDE's): Burgers, Darcy-stroming en Navier–Stokes. Gemeten: tot duizend keer sneller dan traditionele oplosers. Die winst geldt per voorspelling, na de training: voor veel verwante gevallen loont dat, voor één nieuwe situatie zelden.

Van demo naar onderzoeksresultaat?

Vergelijk voor hetzelfde oefenprobleem $u(t) = e^{-t}$ met $\hat{u}(t) = e^{-t} + 0,1$. De absolute fout is overall 0,1, het residu $\hat{u}' + \hat{u}$ ook. Maar bij $t = 5$ is $u \approx 0,00674$: de offset is daar bijna vijftien keer de echte waarde. Eén gemiddelde foutmaat verbergt waar het misgaat. Leg foutmaat, interval en grenzen vast vóór je evalueert.

BEWIJS · CODE EN MEETKETEN

Code, meetketen en nieuwe toets

De derde stap kijkt buiten de training. Voert de code haar methode goed uit, past het model bij de waarneming, en welke nieuwe meting onderscheidt verklaringen?



FIG. 3 De maan, gefotografeerd door de Galileo-sonde. De kleuren zijn versterkt om verschillen in samenstelling te tonen.

Vier controles

Verificatie Voert de code de gekozen methode correct uit?

Validatie Past het model bij relevante waarnemingen?

Meetketen Hoe kwamen de referentiewaarden tot stand?

Nieuwe toets Welke observatie onderscheidt de verklaringen?

FIG. 4 Een goede berekening en een goed passend model zijn aparte onderzoeksvragen.

Correct rekenen is iets anders dan passend modelleren

Verificatie vraagt of code de gekozen vergelijkingen en algoritmen correct uitvoert. Validatie vraagt of die vergelijkingen bij de werkelijkheid passen, voor deze taak. Een foutloze code kan dus keurig het verkeerde model

oplossen. Vergelijk eerst met analytische oplossingen of betrouwbare numerieke referenties, daarna met waarnemingen en hun onzekerheid. Twee curves die op elkaar lijken, zijn nog geen nauwkeurigheidscclaim met een foutmarge.

Een meetketen heeft zelf aannames

Sensorwaarden ontstaan via instrumenten, kalibratie en signaalbewerking: data kunnen al toegepaste fysica bevatten. Een model dat ze reproduceert, ontdekt daarmee niet vanzelf een natuurwet. Wie dat wél wil, moet het model erop bouwen. Gemeten: AI Feynman (Udrescu en Tegmark, 2020) zocht uit data naar formules en vond alle 100 geteste vergelijkingen uit de Feynman Lectures terug, tegen 71 voor eerdere openbare software. Ook beelden hebben een meetketen: de Galileo-opname van de maan heeft versterkte kleuren – echt waargenomen, zichtbaar bewerkt. ♦

EEN NIEUW EXPERIMENT

Kies de meting die verschil maakt

Een derde herhaling is niet altijd de beste volgende vraag. Zoek de plek waar verklaringen uiteenlopen – en leg die vooraf vast.

Vier of zeven?

Rekenvoorbeeld: twee modellen voorspellen op $x = 0$ hetzelfde en op $x = 1$ bijna hetzelfde, maar op $x = 2$ respectievelijk 4 en 7. Een meting daar, met onzekerheid $\pm 0,2$, beslist meteen; nog een herhaling bij $x = 0$ leert je niets nieuws. Schrijf de meetplek en de beslisregel op vóór je meet. Anders wordt een onverwachte uitkomst ongemerkt een nieuwe selectie achteraf.

Bron: 25

VECTOR • MODELDOSSIER • PEILDATUM 08-10-2026

OpenAI in het meetlab: GPT-6 Astra en de foutband

Peildatum 8 oktober 2026. Een actueel OpenAI-model kan een meetverslag verwerken; onze aangenomen filtercasus laat zien welke uitspraken een aparte toets blijven vragen.

Een echte modelversie, een begrensde taak

De officiële pagina noemt gpt-6-astra met tekst- en beeldinvoer, tekstuitvoer, een contextvenster van 1.050.000 tokens en reasoning-effort van low tot max. Audio staat daar niet als ondersteunde invoer. Een onderzoeker kan dus een grafiek plus logtekst aanleveren, maar een sensordataformaat of foutbalk krijgt niet door die capaciteit vanzelf een juiste betekenis. Dit zijn productspecificaties op de peildatum; onderstaande waarden zijn ons fictieve beoordelingsvoorbeeld, geen GPT-6-prestaties.

De vier controleposities

De referenties zijn 1,00; 2,00; 3,00; 4,00 m. Filter A geeft 1,10; 1,80; 3,00; 4,10 m. De absolute fouten zijn 0,10; 0,20; 0,00; 0,10 m: gemiddeld 0,10 m. Filter B geeft 1,15; 2,15; 3,15; 4,15 m en heeft gemiddeld 0,15 m. Op deze vier posities wint A die ene foutmaat.

De aparte intervalproef

Stel daarnaast twintig andere controleposities voor. A's bedoelde 95%-intervallen bevatten de referentie zestien keer; B's negentien keer. De waargenomen dekkingen zijn 80% en 95%. Die kleine aantallen bewijzen niet dat B overall gekalibreerd is. Ze tonen wel dat "A heeft minder gemiddelde fout" de intervalvraag niet beantwoordt.

Wat moet het verslag bewaren?

Noteer de vierpuntsfout en de twintigpuntsdekking als aparte proeven met hun aantallen. Publiceer niet één samengestelde winnaar zonder vooraf gekozen doel. Verander je A's band op grond van deze twintig gevallen, dan heeft die verzameling een ontwikkelrol gekregen. Reserveer nieuwe ritten voor de uiteindelijke toets, inclusief slip en gedeelde timingvariatie.

BEWIJS · FOUTBALKEN

Een foutbalk met een betekenis

Een AI-model kan een strakke foutbalk tekenen en een onbekende meetoffset missen. Veertig herhalingen halveren de statistische onzekerheid tegenover tien; kalibratie vraagt een eigen toets.

Meer metingen, dezelfde offset

Ware waarde 0: wat een perfect instrument zou geven

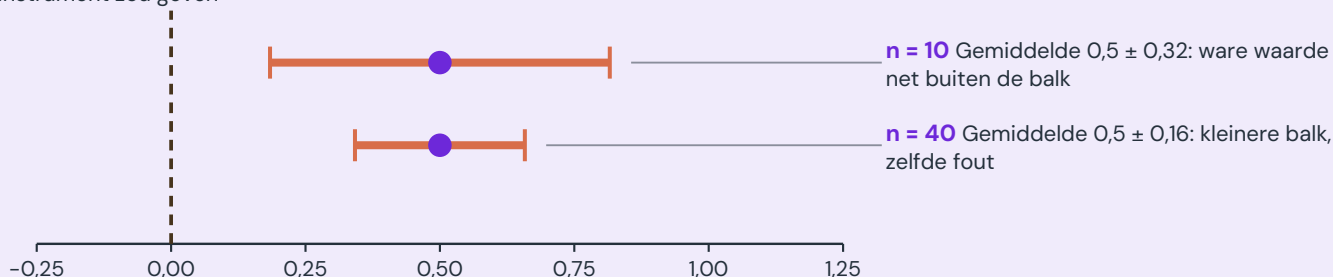


FIG. 1 Rekenvoorbeeld: spreiding 1 per meting, offset 0,5; foutbalk = $1/\sqrt{n}$.

Variatie in meting en gaten in kennis

Metingen variëren door ruis en wisselende omstandigheden: aleatorische onzekerheid. Modellen missen soms gegevens of beschrijven het systeem te grof: epistemische onzekerheid, die met meer gegevens of een beter model kan afnemen. De grens is niet absoluut. Wat vandaag ruis lijkt, kan morgen met een extra sensor verklaarbaar zijn. Gebruik de termen om te bepalen wat helpt: meer herhalen, of beter begrijpen.

Wat beschrijft de band?

Een interval kan een parameter begrenzen, een volgende waarneming voorspellen of numerieke tolerantie tonen. Dat zijn drie verschillende beloftes. Een smalle band rond een gemiddelde betekent niet dat de volgende meting weinig afwijkt. NIST onderscheidt onzekerheid die je statistisch uit herhalingen schat (type A) van onzekerheid die je met andere informatie beoordeelt, zoals een kalibratiecertificaat (type B).

Maak de fout zichtbaar vóór de conclusie

Rekenvoorbeeld: een instrument met een spreiding van 1 eenheid per meting en een

onbekende offset van 0,5. Bij onafhankelijke ruis krimpt de onzekerheid van het gemiddelde met $1/\sqrt{n}$: 0,32 bij tien metingen, 0,16 bij veertig. De offset blijft 0,5. Bij veertig metingen ligt de ware waarde dus ruim drie foutbalken naast het gemiddelde – met een strakke, misleidende band. Alleen kalibratie tegen een referentie vindt zo'n offset. ♦

DATAPUNT

$1/\sqrt{n}$

krimpt van gemiddelde-onzekerheid bij onafhankelijke meetruis

+ VRAAG BIJ IEDERE FOUTBALK

Wat is onzeker: het gemiddelde, de volgende waarneming, de parameter, de modelkeuze of de numerieke berekening?



AXI Veertig keer gemeten, foutbalk gehalveerd, offset onaangeroerd. Welke van de twee zit in jouw data?

Zes papers, zes verschillende vragen

Je hoeft niet elke afleiding bij de eerste lezing te volgen. Begin bij de onderzoeksclaim, lees de evaluatie en gebruik de methodebeschrijving om te begrijpen welke informatie het resultaat draagt.

Attention en kalibratie

Vaswani en collega's, NeurIPS 2017 · arXiv:1706.03762. Bekijk de vertaaltaken en baselines, en volg scores, gewichten en values in de attentionformule. Welke onderdelen herken je nog in het systeem dat jou interesseert? Guo en collega's, ICML 2017 · arXiv:1706.04599. Begin bij het verschil tussen goed classificeren en betrouwbaar twijfelen, en bekijk de betrouwbaarheidsdiagrammen. Welke nieuwe test is nodig als jouw toepassing afwijkt van de validatieset?

Meetgrens en fysische voorwaarden

Hensen en collega's, Nature 2015 · doi:10.1038/nature15759. Lees hoe de Delftse groep detectie- en lokaliteitsmazen tegelijk sloot, met de Nobelachtergrond van 2022 ernaast. Waarom volstonden 245 pogingen? Raissi, Perdikaris en Karniadakis, 2017 · arXiv:1711.10561. Volg hoe residuen en waarnemingen in het verlies terechtkomen. Noteer welke informatie vooraf gegeven is en welke uit data wordt geschat.

Trainingsfalen en eiwitstructuur

Krishnapriyan en collega's, NeurIPS 2021 · arXiv:2109.01050. Lees deze naast de vorige. Let op het verschil tussen een model dat een oplossing kan voorstellen en een training die haar

vindt. Jumper en collega's, Nature 2021 · doi:10.1038/s41586-021-03819-2. Begin bij CASP14 en de nauwkeurigheidsmaten; lees de EMBL-EBI-uitleg van pLDDT en PAE erbij. Welke vraag over beweging beantwoordt de structuur niet? ◆

+ JE LEESNOTITIE

Schrijf per bron: centrale vraag, gebruikte informatie, hoofdresultaat, sterkste toets en belangrijkste resterende beperking.

■ ALLE ANTWOORDEN • QUIZZEN, REKENWERK EN DENKPROEVEN

De redenering achter de keuze

Een antwoord is zo sterk als de aannames waarop het expliciet rust. Vergelijk niet alleen de letter, maar ook de formule, de eenheid en de reikwijdte van de conclusie.

Quiz • pagina 8

- 1 B.** Het trainingsverlies gaat alleen over de trainingsgegevens en de gekozen maat. De validatiefout kan zelfs stijgen: het klassieke teken van overfitting.
- 2 C.** Backpropagation rekent met de kettingregel de gradiënt uit. Een optimalisatiemethode zoals gradient descent gebruikt die daarna voor de update. De gradiënt wijst lokaal bergaf, niet gegarandeerd naar het globale minimum.
- 3 A.** Van alle “99%-zekere” voorspellingen klopt maar ongeveer 95%: de scores zijn te hoog. Guo en collega’s vonden precies dit bij diepe netwerken; temperature scaling helpt vaak.
- 4 A.** Het residu wordt alleen op gekozen punten berekend en de termen worden tegen elkaar afgewogen. Controleer residu en randvoorwaarden apart, ook tussen en buiten de trainingpunten.
- 5 B.** Een mediaan is de middelste waarde: de helft van de domeinen lag dicht bij de experimentele structuur, de andere helft verder weg. Dynamiek en binding zijn andere vragen.
- 6 C.** Een score betekent iets binnen taak, gegevens, procedure en maatstaf. Een tweede benchmark helpt alleen als die wél op jouw toepassing lijkt; parameters tellen zegt niets over geschiktheid.

Quiz • pagina 20

- 1 C.** Constante snelheid in een rechte lijn betekent versnelling nul, dus volgens $\Sigma F = ma$ ook resulterende kracht nul. Zwaartekracht en

normaalkracht heffen elkaar op; voor doorgaan is geen kracht nodig.

- 2 A.** $E_k = \frac{1}{2}mv^2$. De snelheid wordt drie keer zo groot, dus de energie $3^2 = 9$ keer zo groot.
- 3 B.** Lokale modellen met vooraf vastgelegde uitkomsten halen hooguit $S = 2$; de quantumtheorie tot $2\sqrt{2}$. Elk station ziet afzonderlijk willekeurige uitkomsten, dus seinen kan niet.
- 4 A.** De snelheid kost $-7,2 \mu\text{s}$ per dag, maar de hogere positie in het zwaarteveld, op 20.200 km hoogte, levert $+45,8 \mu\text{s}$. Netto lopen de klokken ongeveer $38,6 \mu\text{s}$ per dag voor.
- 5 B.** De tweede hoofdwet gaat over het geïsoleerde geheel. De koelkast gebruikt arbeid om warmte naar de keuken te pompen; daar stijgt de entropie meer dan ze binnenin daalt.
- 6 B.** Landauers grens geldt voor logisch onomkeerbaar wissen in een warmtebad. Omkeerbare bewerkingen vallen er in principe niet onder, en het verbruik van een taalmodel wordt door heel andere processen bepaald.

Rekenwerk • pagina 34

- 1 C.** $a = F/m = 2 \text{ N} / 0,5 \text{ kg} = 4 \text{ m/s}^2$. Na 3 s vanuit rust: $v = at = 4 \text{ m/s}^2 \times 3 \text{ s} = 12 \text{ m/s}$. Wie 4 m/s kiest, verwacht versnelling met snelheid.
- 2 A.** Per 100.000 mensen: 500 zieken, waarvan 475 positief; 99.500 gezonden, waarvan 1.990 vals positief. Na een positieve test: $475/(475 + 1.990) = 475/2.465 \approx 0,193$, dus ongeveer 19%.

- 3 B.** Per bit: $kT \ln 2 = 1,38 \times 10^{-23} \times 77 \times 0,693 \approx 7,4 \times 10^{-22} \text{ J}$. Maal 8.000 bits: $\approx 5,9 \times 10^{-18} \text{ J}$. $7,4 \times 10^{-22} \text{ J}$ is één bit; $2,3 \times 10^{-17} \text{ J}$ hoort bij 300 K.
- 4 C.** $v^2/2c^2 = (7.500 \text{ m/s})^2 / (2 \times 9,0 \times 10^{16} \text{ m}^2/\text{s}^2) = 3,125 \times 10^{-10}$. Maal 86.400 s $\approx 2,7 \times 10^{-5} \text{ s} = 27 \mu\text{s}$. $7,2 \mu\text{s}$ is de waarde voor een GPS-satelliet, die trager beweegt.

Denkproeven • pagina 14

- 1. Demon:** neem gas, poort, geheugen en omgeving samen. Met een eindig geheugen vraagt hergebruik dat bits worden gewist of afgevoerd, en ideaal klassiek wissen kost minstens $kT \ln 2$ per bit aan warmte. Onbeperkt nieuw geheugen is een verbruikte hulpbron, geen gratis cyclus.
- 2. Kopieerapparaat:** als $|0\rangle$ naar $|00\rangle$ en $|1\rangle$ naar $|11\rangle$ gaat, maakt lineariteit van een superpositie een verstrengelde combinatie van $|00\rangle$ en $|11\rangle$, geen twee onafhankelijke kopieën. Teleportatie bewaart geen origineel en vraagt gedeelde verstrengeling plus twee klassieke bits.
- 3. Twee curves:** de productterm $x(x-1)(x-2)$ is nul bij $x = 0, 1$ en 2 . Bij $x = 3$ is zij 6: A geeft 9, B geeft 15. Perfecte trainingsfit beslist dus niet; een nieuwe meting of expliciete voorkennis moet kiezen.

Simulaties • pagina's 15, 28

Leerstap: $x' = (1 - 2\eta) \cdot x$. Elke start $x_0 \neq 0$ convergeert voor $0 < \eta < 1$; bij $\eta = 0,5$ in één stap, tussen 0,5 en 1 met wisselend teken. Bij $\eta = 1$ blijft de grootte gelijk, daarboven groeit zij. Golven: de somamplitude is $2A \cdot |\cos(\phi/2)|$. Bij gelijke A en frequentie geeft $\phi = 0$ amplitude $2A$, $\phi = \pi/2$ amplitude $\sqrt{2} \cdot A$ en $\phi = \pi$ amplitude 0.

Nieuwe puzzels: het volledige bewijs

De voorbeelden zijn aangenomen miniwerelden. Iedere oplossing gebruikt precies de aangegeven voorwaarden.

Raadsels • pagina 7

1. $w = 2/3$. $2 = w + 4 - 4w = 4 - 3w$, dus $3w = 2$. Het toegestane interval bevat $2/3$; de lineaire vergelijking heeft maar één oplossing.

2. **Nee:** bijvoorbeeld $(x,b) = (7,0)$ en $(5,2)$. Beide paren geven 7. Sterker: voor ieder x past $b = 7-x$. Eén som zonder aanvullende informatie identificeert de echte positie niet.

3. $t = 2$ s. Bij $t=0$ voorspellen beide 0. Bij $t=2$ voorspelt A 2 en B 4. Alleen de tweede keuze kan onder de gegeven foutloze regels uniek onderscheiden.

Raadsels • pagina 32

1. $x = 5$. Het eerste punt geeft $b=2$; het tweede $8=3a+2$, dus $a=2$. Dan $12=2x+2$, zodat $x=5$. De twee

verschillende ijkpunten leggen de lijn uniek vast.

2. $\sqrt{0,05} \text{ m} \approx 0,2236 \text{ m}$. Tel eerst de varianten: $0,01+0,04=0,05 \text{ m}^2$. Neem daarna de wortel. $0,30 \text{ m}$ zou de som van standaardafwijkingen zijn, niet deze onafhankelijke combinatie.

3. **B gevolgd door M:** $0,025 \text{ m}^2$. $B \rightarrow M$ geeft $(0,04+0,01)/2 = 0,025$. $M \rightarrow B$ geeft $0,04/2+0,01 = 0,030$. Het verschil is $0,005 \text{ m}^2$. Dit zijn opgegeven operaties, geen universele Kalmanregel.

Gedachtenproeven • pagina 39

De gespiegelde sensor: Nee. De kopie voegt geen informatie over de positie toe: beide schermen zijn functies van dezelfde meting.

Een onafhankelijke tweede meting kan wel nieuwe informatie leveren, maar ook die aanname moet worden gecontroleerd. Gelijke outputs van twee echte sensoren bewijzen op zichzelf niet dat ze kopieën zijn; hier is de kopie expliciet gegeven. De foutbalk zonder wereld: De kleinere opgegeven band is nog geen bewijs van betrouwbaarheid. Toets veel nieuwe ritten tegen een geschikte referentie. Vergelijk dekking bij de bedoelde populatie en de breedte; alleen breedte belooft overzekerheid. Ook goede gemiddelde dekking kan slechte dekking in specifieke omstandigheden verbergen. De proef moet relevante ritgroepen apart bekijken.

■ VECTOR • BEWIJS BIJ DE NIEUWE OPGAVEN

Nieuwe quizzen: berekening en reden

De voorbeelden zijn aangenomen miniwerelden. Iedere oplossing gebruikt precies de aangegeven voorwaarden.

Quiz • pagina 6

1. B. $K = 0,04 / (0,04 + 0,01) = 0,8$. De aangenomen sensorvariantie is kleiner; zie pagina 12.
2. A. De innovatie is 0,30 m; $2,00 + 0,8 \times 0,30 = 2,24$ m. Zie pagina 12.
3. C. $K = 0,04 / 0,13 = 4/13 \approx 0,308$. De voorspelling krijgt $1 - K = 9/13$. Zie pagina 12.
4. A. Gemeenschappelijke fout kan correlatie of bias geven; meer signalen zijn dan geen onafhankelijke herhaling. Zie pagina 13.

Quiz • pagina 26

1. C. $\sqrt{(0,03^2 + 0,04^2)} = \sqrt{0,0025} = 0,05$ m; zie pagina 27.
2. A. Neem de wortel van de variantie: $\sqrt{0,008} \approx 0,08944$ m. Zie pagina 13.
3. B. Onafhankelijke varianten tellen op: $0,008 + 0,002 = 0,010$ m². Zie pagina 13.
4. C. Middelen kan onafhankelijke ruis verminderen; dezelfde systematische verschuiving blijft. Zie pagina 27.

Quiz • pagina 38

1. B. De som is 0,40 m; delen door vier geeft 0,10 m. Zie pagina 44.
2. A. $16/20 = 0,80$. Dat is de dekking in deze kleine verzameling, geen garantie voor andere ritten. Zie pagina 44.
3. C. Bijstellen gebruikt informatie uit de bekeken test; reserveer een nieuwe eindtoets. Zie pagina 44.
4. B. Overeenstemming kan een gedeelde bias bevatten; zoek een onafhankelijk gecontroleerde referentie. Zie pagina 18.

Colofon

Bronnen en beeldmakers.

Verantwoording

Oorspronkelijke uitleg en rekenvoorbeelden. AXI en de strip zijn fictie. Bronnen en beeldcredits

onderscheiden onderzoek, foto's, structuurmodellen en AI-illustraties.

Bronnen

1 Turing (1950) — Computing Machinery and Intelligence · [doi:10.1093/mind/LIX.236.433](https://doi.org/10.1093/mind/LIX.236.433) 2 Rumelhart, Hinton & Williams (1986) — Learning representations by back-propagating errors · [doi:10.1038/323533a0](https://doi.org/10.1038/323533a0) 3 Krizhevsky, Sutskever & Hinton (2012) — ImageNet Classification with Deep Convolutional Neural Networks · papers.nips.cc 4 Vaswani et al. (2017) — Attention Is All You Need · [arXiv:1706.03762](https://arxiv.org/abs/1706.03762) 5 Elhage et al. (2021) — A Mathematical Framework for Transformer Circuits · transformer-circuits.pub 6 Lin, Hilton & Evans (2021) — TruthfulQA · [arXiv:2109.07958](https://arxiv.org/abs/2109.07958) 7 Liang et al. (2022) — Holistic Evaluation of Language Models · [arXiv:2211.09110](https://arxiv.org/abs/2211.09110) 8 NIST (2023) — AI Risk Management Framework 1.0 · nvlpubs.nist.gov 9 NIST (2024) — Generative AI Profile · nvlpubs.nist.gov 10 MIT OpenCourseWare — Conditioning and Bayes' Rule · ocw.mit.edu 11 Guo et al. (2017) — On Calibration of Modern Neural Networks · [arXiv:1706.04599](https://arxiv.org/abs/1706.04599) 12 OpenStax — Newton's Second Law · openstax.org 13 OpenStax — Kinetic Energy · openstax.org 14 Feynman Lectures, I. 4 — Conservation of Energy · feynmanlectures.caltech.edu 15 OpenStax — Wave Functions · openstax.org 16 Nobelpriz Fysica 2022 (Aspect, Clauser & Zeilinger) — Wetenschappelijke achtergrond · nobelprize.org 17 Feynman Lectures, I. 15 — Special Relativity · feynmanlectures.caltech.edu 18 ESA Navipedia — Relativistic Clock Correction · gssc.esa.int 19 OpenStax — Entropy · openstax.org 20 NIST — Fundamental Physical Constants · physics.nist.gov 21 Hong, Lambson, Dhuey & Bokor (2016) — Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits · [doi:10.1126/sciadv.1501492](https://doi.org/10.1126/sciadv.1501492) 22 Raissi, Perdikaris & Karniadakis (2017) — Physics Informed Deep Learning (Part I) · [arXiv:1711.10561](https://arxiv.org/abs/1711.10561) 23 Krishnapriyan et al. (2021) — Characterizing possible failure modes in physics-informed neural networks · [arXiv:2109.01050](https://arxiv.org/abs/2109.01050) 24 Li et al. (2020) — Fourier Neural Operator · [arXiv:2010.08895](https://arxiv.org/abs/2010.08895) 25 NIST Technical Note 1297 — Measurement Uncertainty · nist.gov 26 Jumper et al. (2021) — Highly accurate protein structure prediction · nature.com 27 EMBL-EBI — AlphaFold 2 confidence scores · ebi.ac.uk 28 EMBL-EBI / DeepMind — AlphaFold Database FAQ · alphafold.ebi.ac.uk 29 Feynman Lectures, I. 48 — Beats · feynmanlectures.caltech.edu 30 Paszke et al. (2019) — PyTorch · [arXiv:1912.01703](https://arxiv.org/abs/1912.01703) 31 BIPM — The International System of Units ·

[bipm.org](#) **32** Udrescu & Tegmark (2020) — AI Feynman: a physics-inspired method for symbolic regression · [doi:10.1126/sciadv.aay2631](#) **33** NASA Science — Neutron Stars Are Weird! · [science.nasa.gov](#) **34** Bennett (2003) — Notes on Landauer's principle, reversible computation, and Maxwell's Demon · [arXiv:physics/0210005](#) **35** IBM Quantum Learning — Quantum teleportation · [quantum.cloud.ibm.com](#) **36** Ashby (2003) — Relativity in the Global Positioning System · [doi:10.12942/lrr-2003-1](#) **37** Landauer (1961) — Irreversibility and Heat Generation in the Computing Process · [doi:10.1147/rd.53.0183](#) **38** Bérut et al. (2012) — Experimental verification of Landauer's principle linking information and thermodynamics · [doi:10.1038/nature10872](#) **39** Ho, Erdil & Besiroglu (2023) — Limits to the Energy Efficiency of CMOS Microprocessors · [arXiv:2312.08595](#) **40** Bell (1964) — On the Einstein Podolsky Rosen Paradox · [doi:10.1103/PhysicsPhysiqueFizika.1.195](#) **41** Clauser, Horne, Shimony & Holt (1969) — Proposed Experiment to Test Local Hidden-Variable Theories · [doi:10.1103/PhysRevLett.23.880](#) **42** Hensen et al. (2015) — Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres · [doi:10.1038/nature15759](#) **43** Shannon (1948) — A Mathematical Theory of Communication · [doi:10.1002/j.1538-7305.1948.tb01338.x](#) **44** Hopfield (1982) — Neural networks and physical systems with emergent collective computational abilities · [doi:10.1073/pnas.79.8.2554](#) **45** Nobelprijs Natuurkunde 2024 — Persbericht (Hopfield & Hinton) · [nobelprize.org](#) **46** Nobelprijs Scheikunde 2024 — Persbericht (Baker; Hassabis & Jumper) · [nobelprize.org](#) **47** IceCube Collaboration (2023) — Observation of high-energy neutrinos from the Galactic plane · [doi:10.1126/science.adc9818](#) **48** Hafele & Keating (1972) — Around-the-World Atomic Clocks: Observed Relativistic Time Gains · [doi:10.1126/science.177.4044.168](#) **49** DeepMind (2020) — AlphaFold: a solution to a 50-year-old grand challenge in biology · [deepmind.google](#) **50** Brown et al. (2020) — Language Models are Few-Shot Learners · [arXiv:2005.14165](#) **51** Tromp — Number of legal Go positions · [tromp.github.io](#) **52** Hüllermeier & Waegeman (2021) — Aleatoric and epistemic uncertainty in machine learning · [doi:10.1007/s10994-021-05946-3](#) **53** Griewank (2012) — Who Invented the Reverse Mode of Differentiation? · [maths.tcd.ie](#) **54** EMBL-EBI — AlphaFold Protein Structure Database · [alphafold.ebi.ac.uk](#) **55** Abramson et al. (2024) — Accurate structure prediction of biomolecular interactions

with AlphaFold 3 • [doi:10.1038/s41586-024-07487-w](https://doi.org/10.1038/s41586-024-07487-w) **56** Sanchez-Gonzalez e.a. (2020) • Learning to Simulate Complex Physics with Graph Networks • [proceedings.mlr.press 57](https://proceedings.mlr.press/57/Battaglia_e.a.(2018).) Battaglia e.a. (2018) • Relational Inductive Biases • [arXiv:1806.01261](https://arxiv.org/abs/1806.01261) **58** Kalman (1960) • A New Approach to Linear Filtering and Prediction Problems • people.math.harvard.edu **59** Revach e.a. (2021/2022) • KalmanNet • [arXiv:2107.10043](https://arxiv.org/abs/2107.10043) **60** NIST • Technical Note 1297, gecombineerde standaardonzekerheid • nist.gov **61** OpenAI • GPT-6 Astro, officiële modeldocumentatie • developers.openai.com

Foto's

Via Wikimedia Commons, tenzij anders vermeld.

p. 2, 40 · TMEM253 AlphaFold Predicted Structure · LiuQ2511 · [CCO](#) · Computerweergave van een AlphaFold-voorspelling (geen foto), gekleurd naar pLDDT | **p. 2, 36** · Si wafer · 2x910 · [CC BY-SA 4.0](#) | **p. 2, 29** · The IceCube Neutrino Observatory · Karen Andeen en Matthias Plum voor de IceCube Collaboration · [CC BY 4.0](#) · Schema (geen foto) met Engelstalige labels | **p. 2, 23** · Russell Falls 2 · Noodle snacks (JJ Harrison) · [CC BY-SA 3.0](#) · Russell Falls, Mount Field National Park, Tasmanïë (2009); Gedeeld onder dezelfde licentie | **p. 31** · GPS Block IIIA · U.S. Air Force · [Publiek domein](#) · Artistieke impressie (geen foto); Bijgesneden tot 3/2 rond de satelliet | **p. 43** · Full moon · NASA/JPL/USGS · [Publiek domein](#)

Illustraties

p. 1, 4, 10, 15, 17, 21, 29, 31, 35, 37, 40, 42, 45 · AXI,
de waarnemer · Originele AI-illustratie voor
VECTOR | **p. 1** · Vector O1 cover new · Originele
AI-illustratie voor VECTOR | **p. 2, 10, 11** ·
Informatieroutes en gewichten · Originele
AI-illustratie voor VECTOR | **p. 2, 3** · AXI en de
perfecte voorspeller · Originele AI-illustratie voor
VECTOR | **p. 2, 18** · Vector O1 axi comic 3 ·
Originele AI-illustratie voor VECTOR | **p. 12** ·
Vector O1 kalman lab · Originele AI-illustratie voor
VECTOR | **p. 24, 25** · Twee meetstations ·
Originele AI-illustratie voor VECTOR | **p. 50** ·
VECTOR O2 — omslagillustratie · Originele
AI-illustratie voor VECTOR

Foto's zijn verkleind en kunnen voor de opmaak zijn bijgesneden; ze houden hun aangegeven licentie. Teksten, diagrammen, puzzels en vormgeving: VECTOR. Klik op een brontitel of domein voor het volledige bronadres.



IN DE VOLGENDE EDITIE · Vanaf 8 oktober 2026

Hoe taalmodellen werken

- Van tekst naar tokens.
- Hoe attention context verwerkt.
- Van een antwoord naar een toets.