De meetlat beweegt mee

Waarom niemand precies weet wat een AI-systeem kan, en waarom dat sinds dit jaar een bedrijfsrisico is

2026

In één oogopslag

De best gedocumenteerde maat voor AI-vooruitgang laat zien dat de lengte van taken die een model aankan sinds 2024 ongeveer elke drie maanden verdubbelt. Toen de organisatie achter die maat in januari 2026 haar takenreeks en testomgeving verving, verschoven de scores van dezelfde modellen met 57% naar beneden en 55% naar boven. De enige gerandomiseerde veldproef ooit met echte ontwikkelaars vond een vertraging van 19%, en de vervolgproef liep stuk omdat deelnemers weigerden nog zonder het gereedschap te werken. Van 445 onderzochte benchmarkartikelen rapporteert 16% een onzekerheidsmarge bij het cijfer. De Europese wetgever liep tegen dezelfde muur en stelde de regels voor hoogrisico-AI uit tot december 2027, omdat niemand op tijd had opgeschreven hoe je conformiteit meet.

Dit rapport gaat niet over de vraag of AI werkt. Het gaat over de vraag of u het kunt vaststellen, en wat het kost wanneer dat niet lukt.

16%

van 445 onderzochte AI-benchmarks rapporteert enige onzekerheidsmaat bij het gepubliceerde cijfer. Bean e.a., NeurIPS 2025

Vooraf

Elk gesprek over AI in een onderneming eindigt vroeg of laat bij een getal. Een percentage op een ranglijst, een tijdsbesparing per week, een slaagkans op een testreeks. Dat getal beslist over een investering, een aanwerving, een contractclausule of een aanbesteding. Dit rapport gaat over de herkomst van die getallen.

Er is gekeken naar de primaire evaluatiebronnen die tussen maart 2025 en juli 2026 zijn gepubliceerd: de meetprotocollen van de onderzoeksorganisatie die de bekendste maat voor autonome AI-vaardigheid onderhoudt, het systematische literatuuronderzoek naar de geldigheid van benchmarks, de enige gerandomiseerde proef met beroepsontwikkelaars, en de Europese wetgevingsteksten die van meten een rechtsplicht maken.

Wat die bronnen gemeen hebben is opvallend. Ze zijn geschreven door de mensen die het meten zelf doen, en ze zijn ongewoon openhartig over hoe wankel hun eigen cijfers zijn. Die openhartigheid haalt zelden de samenvatting. Dit rapport zet ze bij elkaar en beschrijft wat ze betekenen voor iemand die op basis van zo'n cijfer iets moet beslissen.

Venture Campus

01 Van demonstratie naar bewijslast

Tot voor kort was de vraag rond een AI-systeem of het werkte. Iemand liet iets zien, het publiek was onder de indruk of niet, en daarmee was de zaak afgehandeld. Die periode is voorbij, en de datum is bekend. Verordening (EU) 2024/1689, de AI-verordening, maakt van een AI-systeem een gereguleerd product met een conformiteitsbeoordeling, technische documentatie en een verplichting om aan te tonen wat het doet.

Vanaf 2 augustus 2026 gelden de transparantieverplichtingen van artikel 50 in de hele Unie. Wie met een AI-systeem communiceert, moet dat kunnen weten, en wie inhoud genereert, moet die als zodanig markeren. Die datum is niet verschoven. Voor legacysystemen die op dat moment al op de markt zijn, geldt artikel 50, lid 2 vanaf 2 december 2026. De nieuwe verboden gelden vanaf dezelfde datum.

Het gevolg reikt verder dan de wetgeving. Een aanbesteding vraagt om prestatiecijfers. Een verzekeraar vraagt om een foutmarge. Een klant vraagt hoe vaak het mis gaat. Een rechter, zoals verderop blijkt bij de behandeling van de regelgeving, vraagt om een norm waaraan hij kan toetsen. Al die vragen komen op hetzelfde neer: bewijs het.

En daar begint het probleem. Want de bewijsapparatuur is niet meegegroeid met de technologie. De cijfers waarmee de sector zichzelf beschrijft zijn grotendeels afkomstig uit testopstellingen die niemand heeft gebouwd met de gedachte dat er ooit een contract, een boete of een rechtszaak aan zou hangen. Dit rapport gaat over die apparatuur: wat ze meet, wat ze niet meet, en wat er gebeurt wanneer u erop steunt.

Bron: Verordening (EU) 2024/1689 (AI-verordening), artikel 50 · Raad van de Europese Unie, persbericht over de definitieve goedkeuring van de digitale omnibus inzake AI, 29 juni 2026

02 De beste meetlat die er is

De meest serieuze poging om AI-vaardigheid over de tijd te meten komt van METR, een Amerikaanse onderzoeksstichting zonder winstoogmerk. Haar maat heet de tijdshorizon, en de vondst is elegant: meet niet hoeveel vragen een model juist beantwoordt, maar hoe lang een taak mag duren voordat het model faalt. Voor elke taak in de reeks wordt gemeten hoeveel tijd een menselijke professional eraan besteedt. De horizon van een model is de taaklengte waarbij het de helft van de keren slaagt.

Die maat groeit exponentieel, en dat doet ze al zes jaar. In de oorspronkelijke publicatie van maart 2025 verdubbelde de horizon van het beste model ongeveer elke 196 dagen, iets meer dan zeven maanden, over de periode 2019 tot 2025. Na de herziening van januari 2026 blijft die zeven maanden overeind voor de hele reeks, maar de deelperiodes tekenen een steiler beeld: 130,8 dagen sinds 2023, en 88,6 dagen sinds 2024. Dat laatste getal is minder dan drie maanden.

> De bekendste maat voor AI-vooruitgang meet hoe lang een taak mag duren voordat het model de helft van de keren faalt.

Lees dat cijfer nauwkeurig, want de precisie zit in het woord dat gemakkelijk wegvalt. De horizon is gedefinieerd bij een slaagkans van vijftig procent. Het is de taaklengte waarbij het model even vaak faalt als slaagt. Bij een taak van vier minuten haalt de huidige generatie bijna honderd procent; bij taken die een mens meer dan vier uur kosten, zakt de slaagkans onder de tien procent. Wat een onderneming nodig heeft ligt bijna nooit bij vijftig procent. Boekhouding heeft 99,9% nodig. Een medisch verslag ook.

METR is over die beperking bijzonder helder. De taken zijn ontworpen om op zichzelf te staan en volledig gespecificeerd te zijn, zodat ze eerlijk zijn voor mens en machine. Echt werk is dat niet: het leunt op eerdere gesprekken, op stilzwijgende kennis, op vertrouwdheid met een bestaande codebasis. Een taak van twee uur in de reeks komt volgens METR eerder overeen met wat een nieuwe medewerker of een ingehuurde freelancer in twee uur klaarkrijgt dan met wat een ervaren collega in dezelfde tijd doet. En de reeks bestaat overwegend uit programmeerwerk, machinaal leren en beveiligingstaken.

Bron: METR, Measuring AI Ability to Complete Long Tasks, 19 maart 2025 · METR, Time Horizon 1.1, 29 januari 2026 · Kwa e.a., arXiv:2503.14499

03 Wat er gebeurde toen de meter zelf veranderde

Op 29 januari 2026 publiceerde METR een nieuwe versie van haar schattingen. Twee dingen waren veranderd. De takenreeks ging van 170 naar 228 taken: 73 toegevoegd, 15 geschrapt, 53 aangepast omdat de omschrijving verwarrend was, de scorefunctie fouten bevatte of de taak te gemakkelijk viel uit te buiten. En de testomgeving verhuisde van Vivaria, dat METR zelf had gebouwd, naar Inspect, het opensourceraamwerk van het Britse AI Security Institute.

De modellen waren niet veranderd. De cijfers wel. De geschatte horizon van GPT-4 uit november 2023 zakte met 57%. Die van een oudere GPT-4 zakte met 35%. GPT-5 steeg met 55%. Claude Opus 4.5 steeg met 11%. Alleen al de verhuizing van testomgeving deed de horizon van GPT-4o dalen van 9,2 naar 6,0 minuten, en bij twee modellen was het verschil tussen beide omgevingen statistisch significant: dezelfde vraag, andere aansturing, aantoonbaar ander resultaat.

Nog een detail uit dezelfde publicatie, en het is het belangrijkste van het hele rapport. De reeks bevat 31 taken die naar schatting acht uur of meer kosten. Van die 31 is er voor vijf een menselijke basistijd daadwerkelijk gemeten. De andere 26 zijn schattingen. De trendlijn die de sector aanhaalt als bewijs van een naderende omwenteling steunt aan de bovenkant, precies waar het interessant wordt, op ingeschatte in plaats van gemeten menselijke tijd.

Niets hiervan is een aanklacht tegen METR. Het tegendeel: geen enkele leverancier publiceert dit soort zelfkritiek, en het is alleen bekend omdat METR het zelf heeft opgeschreven. De les is methodologisch. De gemeten vaardigheid van een model is geen eigenschap van het model. Het is een eigenschap van het model, de aansturing eromheen, de takenreeks, de scorefunctie en de menselijke referentietijd samen. Verander er één, en het cijfer verandert met tientallen procenten, zonder dat er ook maar iets aan het model is gebeurd.

5 van 31

van METR's langste taken (acht uur of meer) heeft een gemeten menselijke basistijd. De overige 26 zijn schattingen. Time Horizon 1.1, 29 januari 2026

Bron: METR, Time Horizon 1.1, 29 januari 2026, inclusief de vergelijkingstabellen TH1 tegenover TH1.1 en Vivaria tegenover Inspect

04 Wat de literatuur over benchmarks zegt

Als de zorgvuldigste meting in het veld zo gevoelig is voor haar eigen opstelling, wat zegt dat over de rest? Op die vraag bestaat sinds eind 2025 een antwoord. Een team van 29 vakreviewers, met onderzoekers van onder meer Oxford, Yale en Stanford, schermde 46.114 artikelen uit zes grote conferenties in machinaal leren en taaltechnologie, hield er 445 over die een benchmark introduceren, en beoordeelde die één voor één op constructvaliditeit: meet dit ding wat het beweert te meten?

Vrijwel elk artikel vertoont een zwakte op minstens één punt. Zestien procent gebruikt een onzekerheidsschatting of een statistische toets om resultaten te vergelijken. Zevenentwintig procent gebruikt een steekproef die is samengesteld op basis van beschikbaarheid. Kernbegrippen als veiligheid en robuustheid worden vaak niet gedefinieerd of niet geoperationaliseerd, waardoor de conclusies erover niet houdbaar zijn. Het werk verscheen op NeurIPS 2025 en eindigt met acht aanbevelingen en een controlelijst, wat op zich al zegt hoe basaal het probleem is.

> Zestien van de honderd benchmarkartikelen zetten een onzekerheidsmarge bij hun eigen cijfer.

Daaronder ligt een tweede laag: de benchmarks zelf bevatten fouten. Een herannotatie van MMLU, een van de meest aangehaalde testreeksen ter wereld, vond dat 6,49% van de vragen een fout bevat. Een audit van SWE-bench, de standaardreeks voor programmeertaken, rapporteerde 32,7% oplossingslekkage: de oplossing was op de een of andere manier vindbaar in de opgave. Het Gemeenschappelijk Centrum voor Onderzoek van de Europese Commissie kwam in 2025 tot negen systemische problemen in AI-benchmarking, waaronder verkeerd geplaatste prikkels en de mogelijkheid om de test te bespelen.

Het gevolg is niet dat benchmarks waardeloos zijn. Het gevolg is dat een verschil van enkele procenten op een ranglijst geen betekenis heeft die u kunt gebruiken, omdat er in de meeste gevallen geen foutmarge bij staat en niemand heeft aangetoond dat de test meet wat de naam belooft. Twee modellen met een identieke topscore kunnen radicaal verschillende sterktes en zwaktes hebben, en een gemiddelde verbergt dat per definitie.

Bron: Bean e.a., Measuring what Matters: Construct Validity in Large Language Model Benchmarks, NeurIPS 2025 Datasets and Benchmarks Track, arXiv:2511.04703 · Gema e.a., herannotatie van MMLU · Aleithan e.a., audit van SWE-bench · Eriksson e.a., meta-review in opdracht van het Gemeenschappelijk Centrum voor Onderzoek van de Europese Commissie, 2025

05 De veldproef die stukliep

Benchmarks meten wat een model kan in een laboratorium. De vraag die een onderneming stelt is anders: wordt mijn ploeg er sneller van? Daarop bestaat precies één gerandomiseerde proef in echte omstandigheden. METR rekruteerde in 2025 zestien ervaren ontwikkelaars van volwassen opensourceprojecten, die gemiddeld al vijf jaar aan hun eigen codebasis werkten. Ze leverden zelf de taken aan uit hun eigen achterstand: 246 stuks, elk willekeurig toegewezen aan wel of geen AI. Toegestaan waren Cursor Pro en Claude 3.5 en 3.7 Sonnet, de toenmalige voorhoede.

Vooraf voorspelden de deelnemers dat AI hen 24% sneller zou maken. Achteraf, na afloop, schatten ze dat het hen 20% sneller had gemaakt. Gemeten waren ze 19% trager. Het betrouwbaarheidsinterval liep van 2% tot 39% vertraging. De onderzoekers schreven onomwonden dat ze een versnelling hadden verwacht. Het opvallendste is niet de vertraging maar de kloof: mensen die de vertraging zelf hadden ondergaan, geloofden na afloop nog steeds het omgekeerde van wat de klok had opgetekend.

Toen kwam de tweede ronde, en die is leerzamer dan de eerste. METR startte in augustus 2025 opnieuw, met 57 ontwikkelaars over 143 codebases en meer dan 800 taken. Op 24 februari 2026 publiceerde de organisatie dat ze haar eigen resultaat als een slechte benadering beschouwde en het proefontwerp zou wijzigen. De reden: de controlegroep bestaat niet meer. Een groeiend deel van de ontwikkelaars wil niet meedoen omdat ze de helft van hun werk niet zonder AI willen doen. Tussen 30% en 50% gaf aan dat ze bepaalde taken bewust niet indienden, uit angst dat die aan de zonder-AI-arm zouden worden toegewezen. Eén deelnemer voltooide geen enkele taak in die arm.

De ruwe cijfers wijzen intussen de andere kant op. Voor de deelnemers uit de eerste studie schat METR nu een versnelling van 18%, met een interval van 38% versnelling tot 9% vertraging. Voor de nieuw geworven groep is dat 4%, met een interval van 15% versnelling tot 9% vertraging. Beide intervallen snijden door nul. En dat is de kern: het effect is vermoedelijk positief geworden, en het is niet meer meetbaar met de methode die het ooit negatief mat, omdat mensen zich niet meer laten randomiseren. Een deelnemer omschreef werken zonder het gereedschap als te voet door de stad gaan nadat je aan de taxi gewend bent geraakt.

Bron: Becker, Rush, Barnes en Rein, Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, METR, 10 juli 2025, arXiv:2507.09089 · Becker, Rush, Cunningham, Rein en Mahamud, We are Changing our Developer Productivity Experiment Design, METR, 24 februari 2026

06 De wetgever loopt tegen dezelfde muur

De AI-verordening steunt op een Europese constructie die bij machines en speelgoed al decennia werkt: de wet formuleert essentiële eisen, en geharmoniseerde normen vertalen die naar toetsbare specificaties. Wie de norm volgt, geniet een vermoeden van conformiteit. Voor AI kreeg CEN-CENELEC daarvoor het technisch comité JTC 21, opgericht op 1 juni 2021, intussen meer dan 300 experts uit meer dan twintig landen, met normalisatieverzoek M/593 en amendement M/613 als opdracht.

Die normen zijn er niet gekomen. De oorspronkelijke deadline van 30 april 2025 werd door de Commissie formeel verschoven naar 31 augustus 2025 en ook die werd gemist. prEN 18286, de norm voor het kwaliteitsmanagementsysteem en het scharnierstuk waaraan de andere hangen, ging pas op 30 oktober 2025 de publieke enquête in, acht maanden te laat. In oktober 2025 grepen de technische besturen van CEN en CENELEC naar een uitzonderingsprocedure: bij een positieve enquêtestemming mag rechtstreeks worden gepubliceerd, zonder aparte formele stemming, en een kleine redactiegroep zou zes van de traagste ontwerpen afwerken. Doel: beschikbaar tegen het vierde kwartaal van 2026.

De reden waarom het zo lang duurt, is precies het onderwerp van dit rapport. Een deel van wat de AI-verordening vraagt, bestaat nog niet als stand van de techniek. De regelgever vraagt om normalisatie op terreinen waar geen norm is om te normaliseren, en een consensusproces kan geen consensus organiseren over een meting die niemand kan uitvoeren. Zonder norm zou vanaf augustus 2026 een juridisch vacuüm zijn ontstaan: verplichtingen zonder een manier om te weten of je eraan voldoet.

De Commissie deed op 19 november 2025 een voorstel dat de toepassing van de hoogrisicoregels koppelde aan de beschikbaarheid van ondersteunende middelen, met vaste einddata als vangnet. Het Parlement en de Raad bereikten op 7 mei 2026 een voorlopig akkoord, het Parlement keurde het op 16 juni 2026 goed en de Raad gaf op 29 juni 2026 het definitieve licht op groen. De uitkomst staat hieronder. Ze is het bekijken waard, want het is een wetgever die vaststelt dat hij niet kan reguleren wat hij nog niet kan meten, en daar zestien maanden voor uittrekt.

Datum	Wat begint te gelden
2 augustus 2026	Transparantieverplichtingen artikel 50 (ongewijzigd)
2 december 2026	Artikel 50, lid 2 voor bestaande systemen; nieuwe verboden praktijken
2 augustus 2027	Ten minste één nationale AI-testomgeving per lidstaat
2 december 2027	Hoogrisicoverplichtingen voor zelfstandige systemen (bijlage III), voorheen 2 augustus 2026
2 augustus 2028	Hoogrisicoverplichtingen voor AI ingebed in gereguleerde producten (bijlage I)

Toepassingsdata van de AI-verordening na de digitale omnibus, goedgekeurd op 29 juni 2026

Bron: Europese Commissie, voorstel digitale omnibus inzake AI, 19 november 2025 · Raad van de EU en Europees Parlement, voorlopig akkoord, 7 mei 2026; Parlement 16 juni 2026; Raad, definitieve goedkeuring, 29 juni 2026 · CEN-CENELEC, Update on CEN and CENELEC's Decision to Accelerate the Development of Standards for Artificial Intelligence, 23 oktober 2025 · CEN-CLC/JTC 21, normalisatieverzoek M/593 en amendement M/613

07 Wat het verandert voor bedrijven

De eerste gevolgtrekking is onaangenaam en onvermijdelijk: uw bewijs is niet te koop. Een score op een publieke ranglijst is een uitspraak over de takenreeks, de aansturing en de scorefunctie van degene die de test uitvoerde. Wat dat cijfer voorspelt over uw taken, uw gegevens en uw betrouwbaarheidsdrempel, heeft niemand vastgesteld. De enige evaluatie die iets zegt over uw toepassing, is er een op uw eigen werk. Dat is geen randactiviteit meer maar een kostenpost met een eigen budgetlijn.

De tweede is dat de drempel de investering bepaalt, niet het model. Het verschil tussen een horizon bij vijftig procent slaagkans en een die u werkelijk kunt gebruiken, is het hele project. Voor

een taak waar een fout goedkoop is en zichtbaar, is vijftig procent bruikbaar. Voor een taak waar een fout duur is en stil, is negenennegentig procent het minimum, en de weg daarheen loopt niet via een groter model maar via controle, verificatie en het ontwerp van het proces eromheen. Dat werk is de eigenlijke innovatie, en het is precies het werk dat bij een demonstratie niet te zien is.

> Een score op een publieke ranglijst is een uitspraak over de testopstelling van de leverancier, niet over uw werk.

De derde is dat zelfrapportage geen bewijs is. De METR-proef leverde daarvoor de scherpste illustratie die de literatuur kent: mensen die 19% trager waren, geloofden achteraf dat ze 20% sneller waren geweest. Wie zijn AI-investering verantwoordt met een tevredenheidsenquête onder de gebruikers, meet het gevoel, niet het effect. Dat gevoel is systematisch en fors in de verkeerde richting misgaan, bij deskundigen, in een gecontroleerde opzet, met de klok erbij.

De vierde is een agenda-afspraak. De zestien maanden uitstel voor hoogrisicoverplichtingen zijn geen pauze, ze zijn een verplaatste deadline, en de transparantieverplichtingen van 2 augustus 2026 zijn niet verplaatst. Wie tussen nu en december 2027 een systeem bouwt dat onder bijlage III valt, bouwt tegen normen die op dit moment nog in ontwerp zijn en die volgens de planning pas eind 2026 beschikbaar komen. Het ontwerp van uw evaluatie is dus zelf een gok op de definitieve tekst van een norm.

Bron: METR, arXiv:2507.09089, 10 juli 2025 · Verordening (EU) 2024/1689, artikel 40 (vermoeden van conformiteit) en artikel 50 · METR, Time Horizon 1.1, 29 januari 2026, over de 50%- en 80%-horizon

08 Wat er nog niet klopt

Het eerlijkste dat over dit onderwerp te zeggen valt, is dat de richting van het productiviteitseffect nog steeds niet met zekerheid vaststaat. De enige gerandomiseerde proef vond een vertraging, en haar eigen auteurs achten dat cijfer inmiddels achterhaald. De vervolgproef is niet interpreteerbaar. Wat overblijft is zelfrapportage, en dat is precies het instrument dat in dezelfde studie faliekant misging. METR's enquête van 11 mei 2026 onder 349 technische medewerkers vindt een mediaan zelfgerapporteerde waardetoename van 1,4 tot 2 keer, en zet er zelf bij dat er redenen zijn om aan die orde van grootte te twijfelen.

De trendlijn verdient dezelfde terughoudendheid. De versnelling naar 88,6 dagen verdubbelingstijd is gemeten over één jaar. De herziene reeks bevat veertien modellen waar de vorige er drieëndertig telde. De betrouwbaarheidsintervallen zijn smaller geworden maar blijven breed: de bovengrens voor het beste model ligt nog altijd op ruim het dubbele van de puntschatting. METR schrijft zelf dat het aantal taken dat de nieuwste generatie niet aankan, klein is geworden, en dat de meetlat aan de bovenkant dus dreigt te verzadigen. Een extrapolatie op zo'n basis is een hypothese, geen prognose.

Ook de observatiecijfers die als alternatief worden aangedragen, meten iets anders dan wat ze lijken te meten. Dat ongeveer 4% van de commits op GitHub door een agent zou zijn geschreven, is een adoptiecijfer. Het zegt niets over de kwaliteit, de hoeveelheid herwerk of de netto tijdswinst. Enkele deelnemers aan de METR-proef merkten op dat ze met agenten andere taken aanpakten

dan zonder, en dat de kwaliteit van het eindresultaat verschilde tussen beide armen. Als de taak verandert, meet een tijdsvergelijking niets meer.

Tot slot een spanning die de sector nog niet heeft uitgepraat. De hoogrisicoregels zijn uitgesteld omdat de normen ontbraken. Vanaf december 2027 zijn diezelfde normen de meting, met een vermoeden van conformiteit als rechtsgevolg. Ze worden nu afgewerkt onder een uitzonderingsprocedure die de formele stemming overslaat, en binnen JTC 21 is daar openlijk bezwaar tegen gemaakt, met een verzoek om de maatregel te herzien. Een norm die via een verkorte weg tot stand komt en die vervolgens beslist of een product legaal op de markt mag, is een uitnodiging tot procederen. Wie in 2027 zijn conformiteit op die normen bouwt, bouwt op een fundament waarover de discussie nog loopt.

Bron: METR, Measuring the Self-Reported Impact of Early-2026 AI on Technical Worker Productivity, 11 mei 2026 · METR, Time Horizon 1.1, 29 januari 2026 · METR, We are Changing our Developer Productivity Experiment Design, 24 februari 2026 · Cantero Gamito, From Consensus to Exceptionality: What the EU's AI Standards Crisis Reveals About Delegated Technical Governance, 28 november 2025

Over Venture Campus

Venture Campus is een Belgisch adviesbureau voor innovatiefinanciering. Het bureau bereidt aanvragen voor, schrijft ze en dient ze samen met de klant in bij de Vlaamse, federale en Europese instrumenten.

De beoordelingscriteria zijn sectorneutraal. Alle sectoren, producten en diensten komen in aanmerking. De klant brengt de technologie, het domein en het plan. Venture Campus brengt de regeling, haar criteria en het perspectief van de beoordelaar, toegepast voor de indiening.

Voor het werk dat dit rapport beschrijft, het opbouwen van kennis en het wegnemen van technische onzekerheid, is het O&O-instrument van VLAIO doorgaans het aangewezen kanaal: 25 tot 45% van de aanvaarde kosten voor ontwikkelingsprojecten, 25 tot 60% voor onderzoeksprojecten, met een maximum van 3 miljoen euro. VLAIO financiert kennisopbouw en technisch risico, geen producten. De percentages gelden op de datum op de cover.

Op de meeste opdrachten geldt no cure, no pay: u betaalt wanneer de steun is toegekend. Dat geldt niet voor Strategische Transformatiesteun en het beschrijft de fiscale trajecten niet. De analyse vooraf is gratis, net als de subsidiescan met zes vragen. Is er geen realistisch geval, dan zegt het bureau dat.

Venture Campus BV | Schaarbeekstraat 20E bus 11 | 9120 Beveren, Belgie | +32 333 11 333 | venturecampus.com

BRONNEN

Verordening (EU) 2024/1689 (AI-verordening), artikelen 40 en 50 · Raad van de Europese Unie, persbericht over de definitieve goedkeuring van de digitale omnibus inzake AI, 29 juni 2026 · Raad van de EU en Europees Parlement, voorlopig akkoord over de digitale omnibus inzake AI, 7 mei 2026; goedkeuring Europees Parlement 16 juni 2026 · Europese Commissie, voorstel voor een verordening tot vereenvoudiging van de uitvoering van geharmoniseerde regels inzake AI (digitale omnibus inzake AI), 19 november 2025 · CEN-CENELEC, Update on CEN and CENELEC's Decision to Accelerate the Development of Standards for Artificial Intelligence, 23 oktober 2025 · CEN-CLC/JTC 21, normalisatieverzoek M/593 en amendement M/613; ontwerpnorm prEN 18286, publieke enquête vanaf 30 oktober 2025 · METR, Measuring AI Ability to Complete Long Tasks, 19 maart 2025 (Kwa e.a., arXiv:2503.14499) · METR, Time Horizon 1.1, 29 januari 2026 · METR, Measuring the Impact of Early-2025 AI on Experienced Open-Source Developer Productivity, 10 juli 2025 (Becker, Rush, Barnes en Rein, arXiv:2507.09089) · METR, We are Changing our Developer Productivity Experiment Design, 24 februari 2026 · METR, Measuring the Self-Reported Impact of Early-2026 AI on Technical Worker Productivity, 11 mei 2026 · Bean e.a., Measuring what Matters: Construct Validity in Large Language Model Benchmarks, NeurIPS 2025 Datasets and Benchmarks Track, arXiv:2511.04703 · Gema e.a., herannotatie van MMLU; Aleithan e.a., audit van SWE-bench · Eriksson e.a., meta-review AI-benchmarking, Gemeenschappelijk Centrum voor Onderzoek van de Europese Commissie, 2025 · Cantero Gamito, From Consensus to Exceptionality: What the EU's AI Standards Crisis Reveals About Delegated Technical Governance, 28 november 2025