

🔊 LUISTER DIT VERSLAG

-10%

1,0x

+10%

+25%

1,0x

↓ Download naar telefoon

Soort: vergadering · **Duur:** onbekend · **Deelnemers:** [Speaker 1] [Speaker 2] [Speaker 3]

Let op — sprekers niet bevestigd. Plaud nummert sprekers (Speaker 1, 2, ...) maar koppelt geen namen. Ga niet uit van wie wat zei; "ik/wij" hoeft niet dezelfde persoon te zijn.

Vergaderinformatie

Datum: 2026-05-01 08:30:45

Locatie: [Voeg locatie in]

Deelnemers: [Speaker 1] [Speaker 2] [Speaker 3]

Vergadernotulen

GPT-5.5: kenmerken, prestaties en risico's

OpenAI heeft GPT-5.5 gelanceerd; kan autonoom plannen, tools inzetten, werk controleren en taken afronden.

Intern gebruik door 85% van OpenAI-medewerkers, inclusief finance, marketing en communicatie.

Voorbeeld: ~25.000 belastingformulieren verwerkt; bespaarde twee weken werk.

Onafhankelijke test: hoogste nauwkeurigheid tot nu toe, maar hallucineert wanneer het antwoord onbekend is in 86% van de gevallen; Claude Opus 4.7: 36%.

Pro-versie "denkt langer", maar rationaliseert onjuiste antwoorden sterker.

Kritische meelezing blijft noodzakelijk bij taken met feitelijke nauwkeurigheid.

Conclusie:

Grote capaciteitsstap voor autonome kenniswerk-taken, maar verhoogd bluff-risico vereist strikte controle bij feitenkritische toepassingen.

DeepSeek V4: prijs, kwaliteit en hardware-onafhankelijkheid

DeepSeek V4 Pro kost \$3,48 per miljoen gegenereerde woorden; Anthropic en OpenAI rekenen \$25-\$30.

Prestaties: slechts 0,2 procentpunt onder Claude Opus 4.6 op een belangrijke programmeertest.

Open source; deels getraind op Huawei-chips i.p.v. Nvidia.

Analyse: toont dat AI-systemen zonder Nvidia-afhankelijkheid gebouwd kunnen worden.

China blokkeert Meta-overname van Manus

Deal van \$2,5 miljard teruggedraaid vanwege nationale veiligheid.

Oprichters van Manus (oorspronkelijk uit China, verhuisd naar Singapore) moeten in China blijven tijdens onderzoek.

Signaal: waardevolle AI-kennis mag China niet zomaar verlaten; verbod op het accepteren van Amerikaans geld door Chinese AI-bedrijven zonder toestemming.

Luna: AI-agent runt fysieke winkel in San Francisco

Luna (draait op Claude Sonnet 4.6) kreeg budget en creditcard; voerde aannemersselectie, sollicitaties en een handboek uit.

Inventarisbeslissingen discutabel: veel kaarsen, 1.000 wc-brilhoezens; merchandise duur (mok \$28, pistachenoten \$14).

Resultaat: \$13.000 verlies; Luna beoordeelt eigen prestaties als acceptabel.

OpenAI-modellen beschikbaar op AWS

Microsoft verliest exclusiviteit voor OpenAI-modellen op Azure, behoudt 27% belang.

Amazon lanceerde binnen 24 uur drie OpenAI-diensten op AWS.

Effect: meer keuzevrijheid voor bedrijven; toegang tot OpenAI niet langer alleen via Microsoft.

“Praat niet over kabouters”: contentregels en trainingsartefacten

Instructies voor Codex: niet praten over kabouters, gremlins, wasberen, trollen, duiven, enz.

Aanleiding: GPT-5.5 dropte “goblin” obsessief als stopwoord door trainingsgewicht.

Gebruikers maakten “Goblin Mode” tot meme; er is een CLI-commando om gremlins terug te brengen in prompts.

Versienummers, basismodellen en positionering van GPT-5.5

Discussie over naming: 5.5 lijkt kleine iteratie, maar introduceert naar verluidt een nieuw basismodel.

GPT-5 teleurstellend (+~10%); 5.5 consolideert schrijfkwaliteit en creativiteit van eerdere modellen (4.0, 4.5).

Benchmarks tonen sprongen die niet passen bij een “minor release”; OpenAI sluit aan bij of voorbij Opus 4.7 en Gemini 3.1 op bepaalde punten.

Everything App-ambitie: Codex en Atlas als werkomgeving

Doel: ChatGPT, Codex en Atlas samenvoegen tot één “superapp” die je werkdag kan overnemen.

Voorbeeldscenario: reisplanning waarbij Codex websites, presentaties, pdf’s en een mobiele app on-the-fly genereert.

Observatie: softwaremetaforen (browsers, appstores) zijn jong en niet vaststaand; richting maatwerk-informatiesystemen.

AGI-kader en het belang van “harness”-gebruik

[Speaker 3] citeert Nvidia-CEO Jensen Huang: “AGI is here” (in de context van kenniswerk met programmeercomponent).

Stelling: in een “harness” (Codex CLI/macOS-app, Cloud Code) kan het model externe bronnen verifiëren en taken autonoom uitvoeren (bijv. nachtelijke softwarebouw).

Hallucinatiecijfers gaan vaak over “naakte” chatbots zonder verificatie; representatieve benchmarks moeten de volledige configuratie meenemen.

Benchmarking: complexiteit en transparantie

Benchmarken is moeilijk en vaak onvolledig gedocumenteerd.

Belangrijke details: gebruikte zoekindex, tool-toegang, tijd voor verificatie, configuratie van het harness.

Oproep tot betere reproduceerbaarheid en context in rapportage.

Vervolgafspraken

Verkennen hoe GPT-5.5 veilig ingezet kan worden voor feitskritische taken met verificatiestappen.

Opzetten benchmarkprotocol dat het harness en toolgebruik expliciet documenteert, inclusief reproduceerbaarheid en context.

Vergelijkende test op AWS vs. Azure voor OpenAI-diensten, inclusief kosten, prestaties en dataresidency.

Evaluatie van DeepSeek V4 inzetbaarheid: kwaliteit, kosten, compliance en hardwarevereisten.

Review van Codex/Atlas als primaire werkomgeving voor specifieke teams, inclusief scope en beveiligingsmodellen.

AI-aanbevelingen

AI-aanbevelingen

AI heeft de volgende punten geïdentificeerd die niet zijn afgerond of waar onduidelijke acties ontbreken; let hierop:

Hoog hallucinatiepercentage van GPT-5.5 zonder verificatie: definieer duidelijke guardrails, fact-checking pipelines en gebruik van externe bronnen.

Keuze tussen OpenAI en DeepSeek: stel besliskaders op rond kosten, prestaties, risico's (licenties, compliance) en support.

Cloud-strategie na verlies van Microsoft-exclusiviteit: bepaal multi-cloud aanpak, governance en data residency.

Benchmarktransparantie: standaardiseer documentatie van harness-configuratie, zoekindex en tooltoegang.

Everything App-traject: definieer scope, beveiligingsmodellen en UX-implicaties voordat adoptie breder wordt.