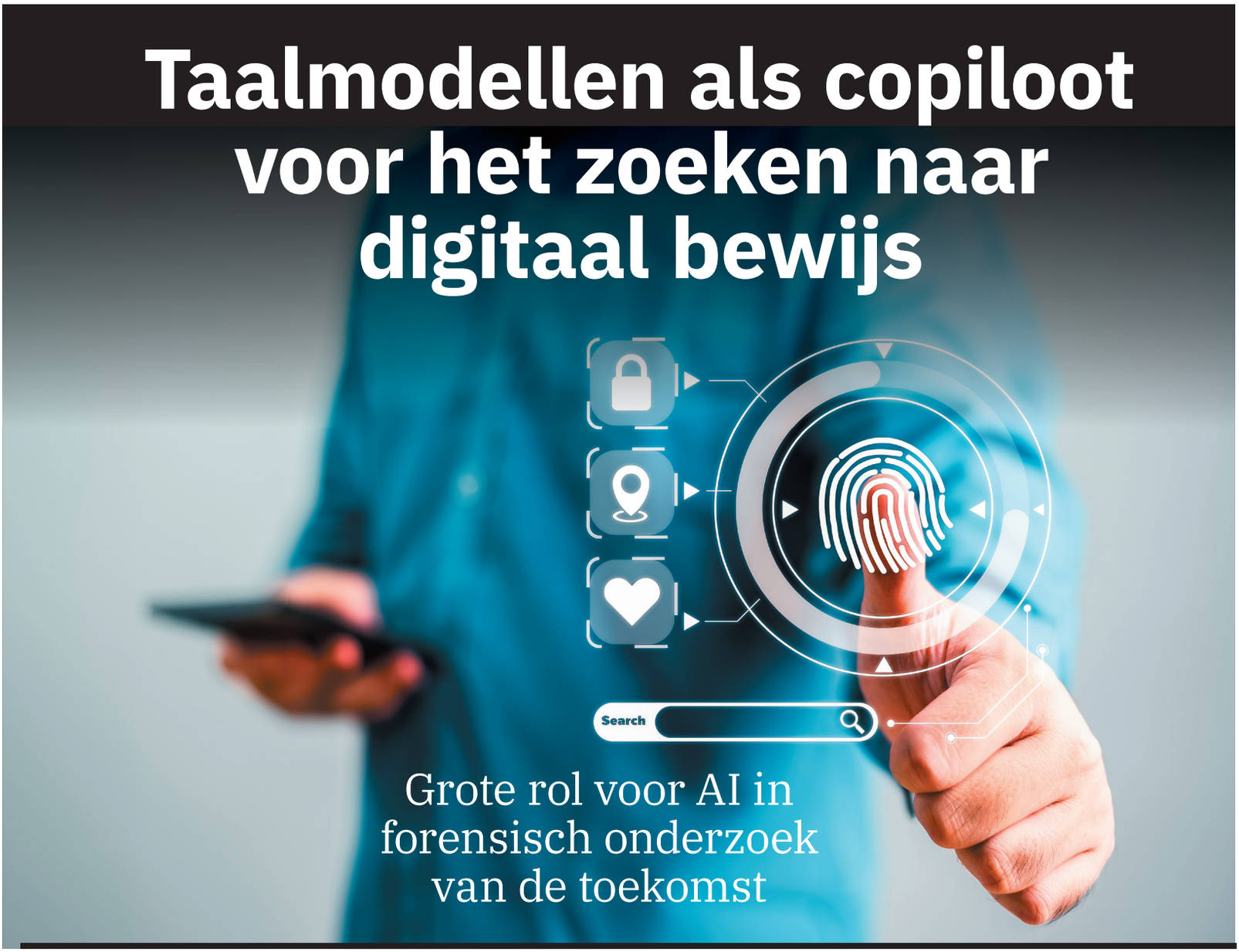
Taalmodellen als copiloot voor het zoeken naar digitaal bewijs



Grote rol voor AI in forensisch onderzoek van de toekomst

Is ChatGPT een digitale speurneus voor rechercheurs? Nog steeds niet, maar we zijn er bijna, ziet Hans Henseler. Vorig jaar concludeerde hij dat AI een slimme assistent voor rechercheurs kon worden. Dat is gedeeltelijk gelukt, de uitdaging is nog om dit on-premises te doen en niet afhankelijk te zijn van een groot taalmodel in de BigTech cloud.

BEGIN 2023 STOND DE WERELD VERSTELD VAN CHATGPT 3, een deep neural network met 170 miljard parameters. Nu, een jaar later, hebben we niet alleen veel betere modellen zoals ChatGPT 4 Turbo, maar blijkt ook dat we met die grote modellen synthetische data kunnen genereren waarmee we kleine modellen met 7 miljard of zelfs minder parameters kunnen trainen die

net zo goed of zelfs beter zijn dan de versie van ChatGPT die we vorig jaar gebruikten (zie kader: Klein maar fijn). Kleine modellen zijn sneller en vereisen veel minder hardware. Daarmee is een belangrijk bezwaar tegen ChatGPT verholpen, namelijk dat de gebruiker zijn gegevens moet delen met de cloud van OpenAI.

Als digitale speurneus had ChatGPT 3 nog een paar andere nadelen. Het model geeft soms onjuiste of verzonnen antwoorden (ook wel hallucineren genoemd, wat overigens voor creatieve opdrachten een hele fijne eigenschap is), geeft geen bronverwijzingen en kan alleen vragen beantwoorden over informatie tot september 2021. Inmiddels is ChatGPT 4 al bij tot april 2023, maar het blijft een probleem dat de kennis van ChatGPT statisch is. Met Retrieval Augmented Generation (RAG) is een taalmodel niet langer beperkt door de data waarmee het oorspronkelijk is getraind en krijgen gebruikers bij de antwoorden ook bronverwijzingen.

Custom GPTs kunnen, net als ChatGPT4, Bing Chat en Google Gemini, zoeken naar informatie op het internet en daar hun antwoorden op baseren. Daarnaast kunnen custom GPTs ook informatie raadplegen die alleen toegankelijk is via een API. Denk bijvoorbeeld aan geldwisselkoersen, wedstrijduitslagen, coördinaten van plaatsen, een interne database met producthandleidingen, et cetera. Er zijn zelfs custom GPTs die kunnen helpen bij het maken van presentaties, tekenen van diagrammen of ontwerpen van een logo. Sinds november zijn er maar liefst 3 miljoen custom GPTs gebouwd, waarvan er 159.000 in de GPT Store te vinden zijn.

Taalmodellen gebruiken

Taalmodellen als assistent lijken niet meer weg te denken. Microsoft heeft de term copiloot in 2021 geïntroduceerd met Github, copilot voor softwareontwikkelaars. In 2023 heeft Microsoft een hele reeks aan copilots gelanceerd met Bing Chat, Windows Copilot en Office 365 Copilot. Ook Google en Apple hebben vergelijkbare assistenten geïntroduceerd of aangekondigd. Kortom, iedereen die met een computer werkt, kan gebruik gaan maken van taalmodellen. Een copiloot die rechercheurs helpt bij het ontcijferen digitale sporen op zoek naar bewijs lijkt haalbaar, maar moet nog gebouwd worden.

ChatGPT geeft soms onjuiste of verzonnen antwoorden

Het belang en nut van taalmodellen staat niet ter discussie, maar de Nederlandse overheid ziet risico's in het kader van de bescherming van persoonsgegevens en mogelijke schending van het auteursrecht. In het GPT-NL-project werken TNO, SURF en het NFI samen aan een eigen taalmodel voor Nederland dat getraind zal worden op open data (zie kader: Symposium E-Discovery).

Onderzoekers in het Datalab van het Ministerie van Justitie en Veiligheid experimenteren met RAG en ChatGPT in de trusted cloud in toepassingen voor verschillende justitieonderdelen. De bedoeling is dat uiteindelijk die toepassingen met GPT-NL of een klein lokaal taalmodel kunnen werken, maar voorlopig werken ze met een gecontracteerde Enterprise oplossing van Microsoft.

Digitale forensische tool

Naast juridische complicaties zijn er ook voldoende technische uitdagingen. Met een custom GPT kan een copiloot gemaakt worden die vragen over een digitaal forensische tool kan beantwoorden. Denk aan het schrijven van zoekvragen en bij het ontwikkelen van software om gegevens te analyseren en rapporteren. De uitdaging is om de kwaliteit van zo'n custom GPT te evenaren met een klein taalmodel. Een andere uitdaging is om het taalmodel verbinding te laten maken met een digitaal

Klein maar fijn

Phi-2, ontwikkeld door Microsoft Research, is een klein taalmodel met 2,7 miljard parameters dat opmerkelijke prestaties levert op diverse benchmarks in vergelijking met grotere modellen. Het model kan uitstekend redeneren en taal begrijpen en presteert beter dan sommige modellen die tot 25 keer groter zijn. Phi-2 volgt in een reeks van kleine taalmodellen die begon in de zomer bij Llama2-7b van Meta, Orca2 van Microsoft , Mistral en Mixtral van Mistral, Gemini Nano 2 van Google en Ferret van Apple. Deze modellen kunnen gebruikers van smartphones, smart glasses en vr-brillen op een vloeiende manier helpen met spraak en beeld. Een klein model kan ook in betrekkelijk korte tijd gefinetuned worden. Dat scheelt in de kosten en het model begrijpt sneller wat er bedoeld wordt, geeft sneller antwoord en de kwaliteit van de antwoorden is beter. In combinatie met RAG kun je zo een slimme assistent bouwen die niet onder hoeft te doen voor een groot model als ChatGPT en op sommige terreinen zelfs veel meer weet.

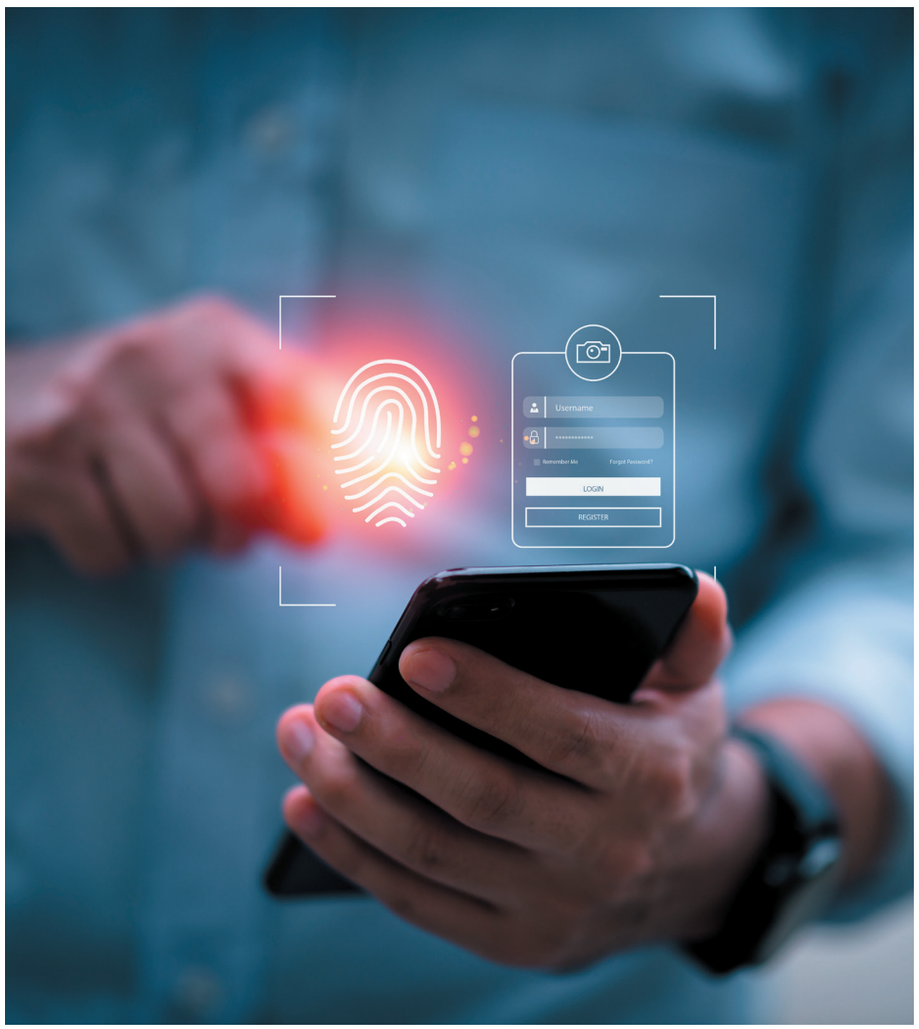
Symposium E-Discovery

Het veertiende E-Discovery Symposium van Hogeschool Leiden wordt 2 april gehouden. Het thema is: praktische toepassingen van taalmodellen in digitaal forensisch onderzoek. Deelname aan het symposium is gratis, registratie is verplicht.

forensische tool. Dan kan de gebruiker namelijk met de copiloot een interactieve dialoog voeren over de digitale sporen die zijn verzameld in het onderzoek. De copiloot kan dan een rechercheur helpen met bijvoorbeeld sporen samenvatten, rapporteren, verbanden leggen of het sentiment van chats analyseren.

Retrieval-Augmented Generation en custom GPTs

De beperkingen van grote taalmodellen als gevolg van verouderde kennis en hallucinaties kunnen worden opgelost met behulp van Retrieval-Augmented Generation (RAG). RAG combineert retrieval (het zoeken naar relevante informatie in een externe database) en generatie (het creëren van antwoorden gebaseerd op die informatie) met taalmodellen. Dit proces helpt hallucinaties voorkomen, doordat het model antwoorden baseert op actuele, geverifieerde informatie. Vergelijk het met een onderzoeker die informatie op het internet raadpleegt om het meest relevante en actuele antwoord te geven.
Met de opensourcesoftware LangChain is het betrekkelijk eenvoudig om een RAG-applicatie te bouwen die werkt met een bestaand taalmodel. Nieuwe start-ups zoals CustomGPT.ai, PDF.ai, AskYourPDF.com, Perplexity.ai en ChatDoc.com bieden RAG als een service aan. OpenAI introduceerde in november vorig jaar RAG in de vorm van



beeld: Shutterstock

Kan een klein taalmodel de kwaliteit van een custom GPT evenaren, dat is de vraag

custom GPTs. Een custom GPT is een combinatie van ChatGPT met een custom prompt, aangevuld met een aantal documenten die ChatGPT raadpleegt door middel van RAG. Publieke custom GPTs zijn te vinden via de GPT-store die door OpenAI in januari is gelanceerd.

Hans Henseler
is lector digital forensics en E-Discovery Hogeschool Leiden en senior wetenschappelijk medewerker NFI.