

Taalmodellen van onschatbare waarde voor forensisch onderzoek

CHATGPT: DIGITALE SPEURNEUS VOOR RECHERCHEURS



ARTIFICIAL INTELLIGENCE (AI) WORDT EEN GAMECHANGER VOOR EEN VEILIGER SAMENLEVING. DE INZET ERVAN ZAL HET FORENSISCH ONDERZOEKSVELD INGRIJPEND VERANDEREN. CHATGPT ZORGT NU AL VOOR EEN VERSNELLING VAN DE INZET VAN AI, ZIET HANS HENSELER. DE IDEALE DIGITALE SPEURNEUS IS HET NIET, MAAR EEN SLIMME ASSISTENT KAN HET ZEKER WEL WORDEN.

door Hans Henseler beeld Shutterstock

AI-TECHNOLOGIE WORDT NU AL GEBRUIKT IN BIJVOORBEELD GEZICHTS- EN SPRAAKHERKENNING. Uit de chatdienst Encrochat werden bijvoorbeeld 25 miljoen berichten onderschept die met de inzet van deep learning technieken gefilterd konden worden op misdrijven als ontvoeringen, liquidaties en zo werden voorkomen. Met het beschikbaar komen van ChatGPT eind november is de inzet van AI in een stroomversnelling gekomen. ChatGPT is een groot taalmodel dat ontwikkeld is door OpenAI en is de afgelopen maanden enorm in de belangstelling gekomen vanwege zijn unieke vermogen om in een natuurlijke dialoog vragen te beantwoorden over een breed

scala aan onderwerpen die het gezien heeft tijdens het leerproces.

SLIMME STUDENT

ChatGPT blijkt een slimme student en kan wellicht ook rechercheurs helpen bij het efficiënter en effectiever onderzoeken van zaken met digitaal bewijs die het nog nooit eerder heeft gezien. Bijvoorbeeld door het vertalen van onderzoeksvragen in natuurlijke taal naar gestructureerde zoekopdrachten, waardoor rechercheurs sneller het juiste bewijs kunnen vinden zonder dat ze een geavanceerde zoektaal hoeven te leren (zie kader: Helpen bij het schrijven van zoekvragen). Ook kan ChatGPT digitale sporen zoals e-mails, chatberichten en

browsersgeschiedenis doorlezen en op verzoek samenvatten, waardoor rechercheurs snel kunnen zien wie, wat, waar en of wanneer iets is gebeurd (zie kader: Samenvatten van informatie). ChatGPT kan ook verbanden tussen data analyseren, zoals herhaaldelijk genoemde e-mailadressen of telefoonnummers, waardoor rechercheurs interactief snel belangrijke personen en relevante onderwerpen kunnen identificeren. Ook andere onderzoekers in het E-Discovery veld zijn van mening dat ChatGPT en aanverwante technieken hier een belangrijke rol in kunnen spelen. In een recente publicatie 'What will E-Discovery Lawyers do after ChatGPT' in LegalTech News and Law schetsen de

HULP BIJ HET FORMULEREN VAN ZOEKVRAGEN

Hansken is een open platform voor opsporings- en veiligheidsdiensten voor het zoeken naar en analyseren van digitale sporen die afkomstig zijn uit in beslaggenomen digitale apparatuur zoals telefoons en computers. Hansken bestaat al meer dan tien jaar. Het is ontwikkeld door het Nederlands Forensisch Instituut en is gebaseerd op o.a. Hadoop en Elasticsearch waardoor het in staat is om petabytes aan digitale data te verwerken en op te slaan. Hansken verwerkt digitale sporen aan de hand van een sporenmodel. Met behulp van de Hansken Query Language (HQL) kunnen deze sporen razendsnel in de Elasticsearch database opgezocht worden. HQL is een krachtige taal die vergelijkbaar is met andere query talen zoals SQL maar die volledig gericht is op het Hansken-sporenmodel. Dankzij de eigenschappen van ChatGPT is het niet moeilijk om met een paar voorbeelden en uitleg van de HQL syntax, ChatGPT vragen in gewone taal te laten omzetten in HQL syntax. Bijvoorbeeld de vraag "Vind email sporen met bijlagen die verzonden zijn tussen 1 juli en 28 juli 2022 in HQL" wordt door ChatGPT moeiteloos vertaald als "email.hasAttachment:true email.sentOn>='2022-07-01' email.sentOn<='2022-07-28'". Om dat te kunnen doen heeft ChatGPT de Hansken HQL-handleiding gelezen en een aantal definities in het sporenmodel, waaronder het type e-mail.

ChatGPT helpt rechercheurs bij het onderzoeken van digitaal bewijs

SAMENVATTEN VAN INFORMATIE

Telefoons en computers bevatten veel sporen van communicatie, denk aan chatberichten zoals WhatsApp, Telegram, sms, maar ook e-mails. Bij aanvang van een onderzoek is weinig bekend, zoals de naam van een verdachte en een vermoeden van een strafbaar feit. Aan de hand van contacten in een telefoon en laptop kan een onderzoeker inzicht krijgen in het netwerk van de verdachte, bepaalde gebeurtenissen, locaties gekoppeld aan datum en tijd. Het uitpluizen van al die informatie is tijdrovend en niet alle sporen leiden naar relevante informatie. ChatGPT kan transcripts van chats samenvatten en ordenen. Vervolgens kan ChatGPT ook vragen beantwoorden. Uit wat korte experimenten met de huidige versie van ChatGPT blijkt de lengte van dit soort samenvattingen beperkt. Maar het illustreert wel de kracht van een slimme assistent die in natuurlijke taal bevestigd kan worden over welke personen er voorkomen in de communicatie en waar er al zoal over wordt gecommuniceerd. Het zal zeker geen bewijs zijn, maar een onderzoeker kan met zo'n assistent wel sneller de informatie verwerken en op het spoor komen dat uiteindelijk tot het bewijs kan leiden.

E-DISCOVERY SYMPOSIUM

Geïnteresseerd in AI toepassingen in digitaal forensisch onderzoek en E-Discovery?

Kom dan naar het jaarlijkse E-Discovery Symposium dat Hogeschool Leiden organiseert op dinsdagmiddag 14 maart 2023. Deelname is gratis maar aanmelden is verplicht. Voor meer informatie, zie: <https://www.hsleiden.nl/digital-forensics/agenda/symposium/e-discovery-2023>

auteurs een aantal experimenten waarbij ze ChatGPT vragen om een complexe boolean query op te stellen aan de hand van een vrij eenvoudige vraag. Inhoudelijk kan ChatGPT geen vragen beantwoorden over informatie die het niet eerder gezien heeft. Maar de bekende Enron-zaak, die uitgebreid besproken wordt op internet en die in E-Discoveryonderwijs ook veelvuldig wordt gebruikt, is wel bekend bij ChatGPT. Op de vraag om voorbeelden te geven hoe Enron de accounting standaarden van de Federale Overheid van de VS heeft overtreden, geeft ChatGPT moeiteloos (en volgens de auteurs indrukwekkend) antwoord.

NIEUWE ZAKEN PROBLEMATISCH

De auteurs vermoeden dat ChatGPT deze voorbeelden heeft verzameld uit Wikipedia en talloze andere publicaties die uitputtend de ondergang van Enron analyseren. Zulke analyses bestaan natuurlijk niet bij nieuwe onderzoeken. De vraag is of ChatGPT ook documenten kan analyseren in nieuwe onderzoeken. Op basis van een beperkte test blijkt

ChatGPT gegeven een specifieke vraag in staat om te bepalen of een e-mail uit de Enron-set wel of niet relevant is en kan bovendien uitleggen waarom dat zo is. Maar voor nieuwe zaken werkt dit niet, want het onderliggende GPT-3-model heeft daar geen informatie over gelezen. Zulke nieuwe informatie kan aan GPT-3 geleerd worden en OpenAI biedt de mogelijkheid om modellen in de cloud te finetunen (zie kader: Doorleren met GPT: Prompt design versus finetuning). Maar dat is kostbaar en advocaten zijn terughoudend met het uploaden van gevoelige documenten van hun cliënten. Die situatie verandert waarschijnlijk als er het komende jaar alternatieven voor GPT3 komen die wel op eigen hardware (die bovendien steeds krachtiger wordt) getraind kunnen worden. ChatGPT is een gefinetuned versie van GPT-3.5. Die finetuning is gebaseerd op Reinforcement Learning from Human Feedback (RLHF). Dat wil zeggen dat ChatGPT met behulp van menselijke reviewers geleerd heeft om geen onwenselijke taal (gewelddadig, sexistisch, racistisch enz.) te gebruiken in conversaties. Ook heeft ChatGPT geleerd wat

De antwoorden zijn indrukwekkend, volgens de auteurs

nuttige antwoorden zijn zodat conversaties vlot verlopen. De inzet van mensen om duizenden prompts te beoordelen is een kostbare zaak en lijkt een van de belangrijke succesfactoren te zijn van ChatGPT. Maar voor een digitale speurneus die conversaties tussen criminelen moet helpen analyseren is het natuurlijk wel een handicap als een rechercheur als antwoord krijgt "I'm sorry, I cannot generate inappropriate or aggressive language. The conversation you have provided contains profanity and offensive language, and as an AI language model, I am programmed to refrain from engaging in that type of behavior".

ONSCHATBARE WAARDE

ChatGPT is dus nog niet de ideale speurneus die kan helpen bij het doorzoeken van digitaal bewijs maar het is wel duidelijk dat grote taalmodellen van onschatbare waarde kunnen zijn. Wat rechercheurs nodig hebben is SleuthGPT waarbij een sleuth een Engelstalige

benaming is voor detective of speurneus, iemand die gespecialiseerd is in het oplossen van misdaden. Op de prompt "Given that ChatGPT has been trained on conversations on a very broad range of topics, how should a SleuthGPT be trained?", antwoordt ChatGPT met enige vertraging en op verzoek in het Nederlands:

"Een SleuthGPT-model kan worden getraind door het originele GPT-3-model te finetunen op een dataset specifiek gericht op onderzoek en detectiewerk. Deze dataset kan voorbeelden bevatten van detectiewerk en onderzoeken, inclusief fictieve en echte gevallen, politierapporten, rechtszittingstranscripten en andere relevante materialen. Het finetuningproces zou het model de taal, technieken en logica die worden gebruikt bij het speuren, laten leren en meer informatieve en accurate antwoorden kunnen genereren bij vragen over onderzoeksprocessen en technieken. Het doel van de finetuning zou zijn om een model te creëren dat speurders kan ondersteunen en assisteren in hun werk, door relevante informatie te verstrekken, verbanden te leggen en hypotheses te genereren."

Zover is het nog niet, maar zoals we inmiddels gewend zijn van ChatGPT klinkt het antwoord zeer overtuigend. 🗣️

REACTIES EN BIJDAGEN

Voor reacties en nieuwe bijdragen van IT-experts:
Tanja de Vrede
020-2467230
t.d.vrede@agconnect.nl

DOORLEREN MET GPT: PROMPT-DESIGN VERSUS FINETUNING

Een groot taalmodel als GPT is in staat om vragen te beantwoorden die aansluiten bij de tekst die gebruikt is tijdens de training. Maar GPT kan ook doorleren. Door middel van een prompt met enkele voorbeelden begrijpt GPT in veel gevallen intuïtief welke taak er wordt bedoeld en kan zelf nieuwe prompts aanvullen. Dit wordt ook wel few-shot learning genoemd en is heel handig om met een beperkt aantal voorbeelden het model een antwoord te laten geven. Maar de capaciteit hiervan is beperkt. OpenAI biedt daarom ook de mogelijkheid om een model te finetunen met specifieke data. Ze geven daarbij aan dat de kwaliteit van de resultaten beter is dan bij prompt-design, dat je met meer data kunt trainen dan mogelijk is bij prompts, dat je uiteindelijk kosten bespaart doordat een gefinetuned model met kortere prompts kan werken en dat de responstijd verbetert.