Van buzz naar basis: AI voor beleggers

Tijdens gesprekken met investeerders hoor ik steeds vaker de term 'AI-washing'. Maar wat is AI vandaag, en wat is haar rol in het beleggingsteam van morgen?

Door *Reza Kahali*

Als student bouwde ik Bayesiaanse netwerken om de volatiliteit van aandelen te schatten. Bij GSK ontwikkelden we 'digital-twin' modellen om machines te simuleren voor de productie van astma-medicatie. Bij NN IP (nu GSAM) werkten we weer met totaal andere modellen, om beleggingsideeën te genereren. Dit alles werd destijds al onder de noemer AI geschaard.

Maar sinds GPT-3 associëren we AI vooral met generatieve taalmodellen die data produceren. Dat is echt anders dan bijvoorbeeld het model van de Belastingdienst dat fraude detecteert. Dat is een discriminatief model dat geen data genereert maar classificeert.

En AI-analisten of -agenten? Dat zijn diezelfde taal-modellen waarmee je een autonome agent nabootst. Denk aan agent Smith uit The Matrix. Nabootsen doe je door een recursief proces op gang te brengen via vraag-en-antwoord-interacties. Interacties tussen het taal-model en bijvoorbeeld het internet, andere taalmodellen of planningsalgoritmes.

Ik wil het hebben over het gebruik van deze techno-logie, niet om processen of administratie efficiënter te maken, maar binnen het beleggingsproces – waar de kansen én de hindernissen het grootst zijn. Ik bespreek mijn meest prominente uitdagingen. Onderwerpen als cognitieve beleggings-biases van AI binnen het beleggingsproces laat ik achterwege.

Om die uitdagingen te begrijpen, moeten we beginnen bij de oorsprong: machinevertaling en spraak-synthese. In 2016 deed ik econometrisch onderzoek voor macro-simulaties. Machinevertaling ontwik-kelde zich snel en was een enorme inspiratiebron. Of je het nu over reeksen consu-mentenprijzen hebt of over woordreeksen: beide zijn tijdreeksen.

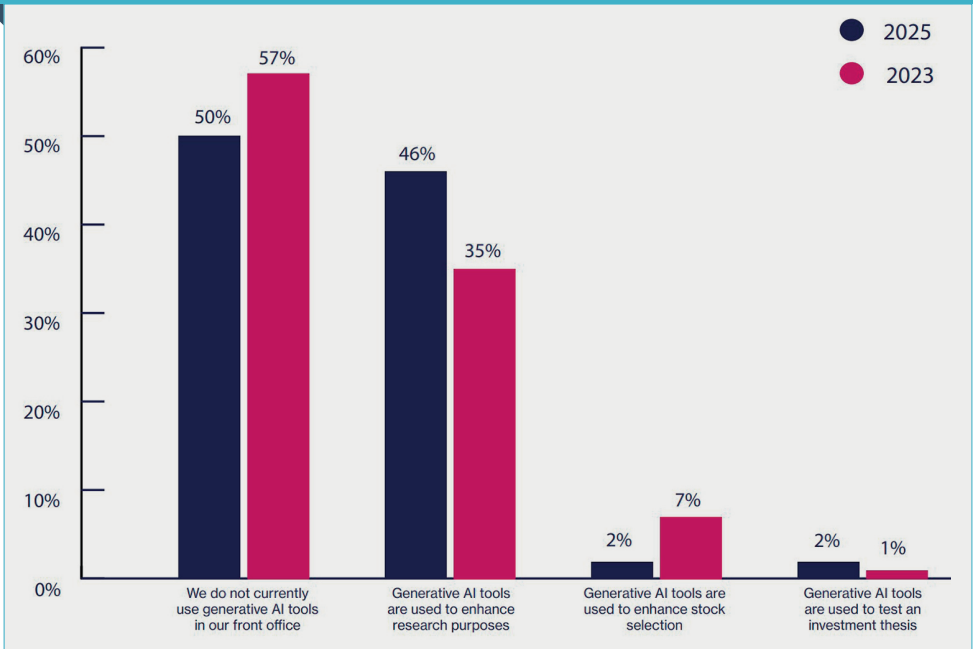
Een paar jaar eerder (2014) beschreven Sutskever (een van de oprichters van OpenAI) en anderen een Long Short-Term Memory (LSTM)-netwerk dat een zin encodeert in een vaste vector, waarna een tweede LSTM een nieuwe zin decodeert. De onderzoekers merkten echter dat één vaste vector met alle tekst en recursie een knelpunt was. Ze introduceerden het concept van 'softe aandacht' om dit op te lossen. Met aandacht kan het model tijdens het decoderen relevante delen van de tekst activeren. Hierdoor werd de rekenkracht efficiënter.

In 2016 wees een senior onderzoeker bij Ortec Finance mij op het paper 'WaveNet'. 'Hier moet je echt naar kijken voor tijds-reeksen', zei hij. Ik was verbluft. WaveNet (2016) liet zien dat convolutionele netwerken ook tijdreeksen kunnen modelleren en genereren zonder recursie. In hun geval was de reeks rauwe audio.

En als klap op de vuurpijl introduceerden Vaswani en anderen in 2017 het Transformer-model met zelf-attentie. Er waren geen recursie of convoluties meer, waardoor het meer paralleliseerbaar is en zeer grote modellen mogelijk werden. Sindsdien vormen Transformers de kern van bijna alle moderne AI-systemen.

De grote uitdaging is dat deze fundamenten uitgaan van een statische wereld.

FIGUUR 1: GEBRUIK VAN AI-TOOLS BINNEN VERMOGENSBEHEER



Bron: AIMA, 'Charting the course: Lessons from AI leaders in alternative investment', p. 31

Taal verandert wel, maar heel langzaam (lees bijvoorbeeld eens literatuur uit de 19de eeuw). Markten en de economie zijn dynamischer.

Het eerste struikelblok is een bekende: overfitting. Een model dat te goed leert op historische data en slechter op nieuwe data presteert. Het fine-tunen van grote modellen leverde fantastische resultaten in de tests op. Maar bij nieuwe data generaliseerden de AI-analisten niet goed. De literatuur en ervaring leren dat een mix van kleinere, taakgerichte modellen met zorgvuldig contextmanagement stabieler en voorspelbaarder is. Veel beter dan één monsterlijk model.

Het tweede probleem is inherent aan generatieve modellen: hallucinaties. Taalmodellen verzinnen 'feiten' die logisch klinken, maar onwaar zijn. Zie een groot taalmodel als een compressor van de wereld: miljarden parameters comprimeren en coderen patronen. Bij gebruik wordt deze kennis weer gedecomprimeerd. Het model heeft echter niet altijd genoeg 'bits'. Dan vult het de leemte met het meest waarschijnlijke patroon – dat soms onjuist blijkt te zijn.

Voorbeeld: comprimeer een foto van een meisje met een moedervlek. Bij compressie verdwijnt het detail van de moedervlek. Reconstrueer je de foto, dan vult het model die plek met haar huidskleur. Het resultaat is een subtiele 'hallucinatie', een meisje zonder moedervlek. Dit komt doordat het model niet geoptimaliseerd is om de waarheid te vertellen, maar om het meest waarschijnlijke scenario te schetsen.

In mijn ervaring helpt het enorm om je AI-analist gezaghebbende data als vangrail mee te geven. Bijvoorbeeld feiten uit officiële databases en officiële statistieken. Maar ook technieken zoals 'Chain-of-Verification' of 'DoLa' kunnen helpen. Vertrouwde statistische methodes blijven echter nuttig. Denk aan stabiliteitstesten met herhaalde runs en Monte Carlo-simulaties.

Het derde probleem is datacontaminatie. Denk bijvoorbeeld aan een specialistisch macro-AI-analist die advies moet geven tijdens de COVID-19-beurscrash. Deze modellen hebben in de voortraining vaak al onderzoeken uit 2021–2022 gezien. In de simulatie geeft de analist perfecte voorspellingen. Niet vanuit 'analyse', maar vanuit 'gelekte' kennis.

Ruis of fictieve variaties toevoegen helpt, zodat het originele patroon niet meer letterlijk in de data zit. Daarnaast is Domeinoverdracht handig: vergelijkbare vragen stellen in een nieuwe context. Het model kan dan niet simpelweg het geijkte antwoord ophalen. In de praktijk kun je een neuraal netwerk niet eenvoudig specifieke kennis ex-ante laten vergeten.

Er spelen natuurlijk meer zaken binnen een professioneel beleggingsteam. Denk aan algoritme-aversie. Dit viel mij als jonge 'quant' al op. Mensen vertrouwen vaak meer op hun eigen oordeel dan op dat van een algoritme, ook al is het algoritme beter. Ze kiezen liever voor eigen risico en eer dan voor een consequent inschatbaar resultaat. Mede daarom is de adoptie daar dan ook laag (zie Figuur 1). De kansen zijn echter groot.

Stel bijvoorbeeld dat je financieringsstress in de Amerikaanse repomarkt wilt bepalen. Je kwantitatieve model kan het verschil analyseren tussen de overnight-reporente en het middelpunt van de federal-funds-doelbandbreedte, ofwel de repo-spread. Vervolgens beoordeelt het statistisch model of dit verschil abnormaal is. Daarnaast leest je AI-analist academisch onderzoek en het beleidsplan van de centrale bank. Zo schat hij in of het beleid naar verwachting meer of minder liquiditeit oplevert. Daarmee kan de AI-analist de blinde vlekken in je kwantitatieve modellen aanvullen, iets wat een ervaren beleggingsstrateeg normaliter deed. Je statistisch model kleurt bijvoorbeeld felrood, maar je AI-analyst geeft aan dat het beleid dit heeft gecorrigeerd.

Mijn overtuiging is dan ook dat de toekomst van beleggen ligt in samenwerking tussen mens en AI. De AI-analisten en modellen verwerken razendsnel data en voeren strategieën uit. Daaromheen staan menselijke professionals voor creatief denkwerk, strategie uitzetten en out-of-the-box-ideeën.

Deze modellen kun je nog niet succesvol de opdracht geven om bijvoorbeeld een geheel nieuwe beleggingsfilosofie te ontwikkelen of radicaal andere invalshoeken te verzinnen. Daar zijn wij mensen (voorlopig) beter in. ■

Reza Kahali

Fund Manager en CEO, argan.ai

IN HET KORT

De fundamenten van AI gaan uit van een statische wereld. Dat werkt goed bij taal, die slechts langzaam verandert, maar markten en economie zijn juist dynamisch.

Moderne AI biedt enorme kansen, maar kampt met de bekende hoofdpijndossiers: 1. Overfitting 2. Hallucinaties en 3. Datacontaminatie.

Oplossingen liggen onder meer in kleine gespecialiseerde AI-agenten, schone en gezaghebbende databronnen, verificatieketens, stochastische runs of domeinoverdracht.

De toekomst ligt in de samenwerking tussen mens (creativiteit) en een combinatie van AI en klassieke modellen.

FOTO: ARCHIEF ARGAN