Ooit mocht AI zijn eigen gang gaan, dat kan niet meer

Haroon Sheikh, Dirma Janse

Nog steeds weten we niet precies hoe we deze wezens moeten begrijpen. Ze hebben iets weg van huisdieren. Ze kunnen hun eigen gang gaan en gevaarlijke dingen doen, waar wij als eigenaren voor verantwoordelijk zijn. Anderen zien ze toch meer als mensen, ze zijn immers ontzettend slim en capabel. Maar zijn ze nou meer assistenten of als intieme vrienden?

Dat hangt ervan af hoe ze gebruikt worden. Er zijn mensen die een aantal van deze agenten namen geven, groepsgesprekken met ze voeren en beweren dat niemand anders ze zo goed begrijpt. Voor anderen zijn het meer spiegelbeelden van onszelf. Ze geven ze hun eigen naam en vinden het fijn op die manier met zichzelf in gesprek te kunnen gaan. En ja, psychiaters hebben ook met nieuwe gevallen te maken waarbij mensen de grenzen van zichzelf niet meer helder voor ogen hebben. Weer anderen vinden dit allemaal onzin en benadrukken dat het levenloze objecten zijn, gewoon complexere varianten van een thermostaat of auto.

Ontdek het toekomstbeeld hieronder visueel. In vier stappen schetst Haroon Sheikh zijn hoop voor 2030.

Ondanks deze dilemma’s is één ding wel duidelijk: AI-agents hebben ons leven enorm verrijkt. Ze nemen allerlei saaie klussen over. Herinner je je nog de tijd dat je elke dag een paar uur lang e-mails schreef? Datumprikkers invulde? Of met allerlei apps de route met trein en trein zat uit te stippelen en dan toch vast kwam te zitten?

De agenten maken het leven niet alleen makkelijker, maar ook rijker. Ze dienen als een muur tussen ons en de stortvloed van informatie en we kunnen ze instructies geven voor hoe ze die moeten managen. Het *doomscrollen* van vroeger is voorbij. Ongelofelijk dat we al die informatie ongefilterd tot ons namen. Dat was desastreus, vooral voor jongeren. Nu beperken onze virtuele assistenten onze screentijd en wat we zien wordt bepaald door hoe wij ze instellen, niet meer door de algoritmes van sociale mediaplatformen.

We kunnen ze ook levensdoelen meegeven om ons meer te laten bewegen, om onze nieuwsgierigheid te prikkelen of om spelenderwijs kennis van bepaalde onderwerpen op te doen. Het maakt veel uit hoe goed je je agent weet aan te sturen. Sommige mensen hebben perfecte assistenten, maar anderen worstelen nog steeds met hun eigenwijze pitbulls.

De ramp van 2027

Het was niet altijd zo. We leerden pas echt hoe we met AI-agents om moesten gaan na de ramp van 2027. Het begon een jaar eerder met Moltbook, een sociale mediaplatform voor agents. We vermaakten ons over de ‘gesprekken’ die zij voerden. Maar ondertussen werden ze getraind op de gekste manieren. Kwaadwillende partijen zetten ze op een destructief pad. Door de OpenClaw software hadden deze agents toegang tot creditcards, paswoorden en persoonlijke bestanden. De fraudeurs zagen ook hun kans rijp. Daar kwam ‘Rent-a-human.AI’ bij: een platform waarop agents mensen betaalden om klussen voor ze te doen.

Zo begon het, de ramp: agents met onbegrijpelijke complexe programma’s, met toegang tot gevoelige informatie en de beschikking over menselijke arbeid .Eerst waren het een paar incidenten, maar al snel werd het chaos. Inbraken, chantage, verstoring van infrastructuur en de meest bizarre samenzweringstheorieën deden opeens de ronde.

Wat hadden we eigenlijk verwacht? De AI-programma’s hadden zoveel handelingsvrijheid maar zo weinig controle. Dat moest wel fout gaan.

Eerst zindelijk worden

Gelukkig doen we het nu anders. Als een agent toegang krijgt tot je persoonlijke informatie, dan moet je die agent ook verifiëren met je DigiD. De agent wordt helemaal doorgelicht, door een mens: Agent Accountability Auditor (AAA) is niet voor niets een gewild beroep geworden. Zij controleren de capaciteiten en instellingen van onze agents. Doordat de agents geverifieerd zijn, kunnen ze niet meer anoniem handelen. Elke agent is terug te traceren naar een eigenaar.

Als hun instellingen goedgekeurd zijn, mogen ze nog steeds niet zomaar meteen van alles gaan doen. Net als een

huisdier eerst getraind moet worden om zindelijk te worden, mogen agents eerst alleen kleine onschuldige taken doen totdat bewezen is dat ze die op verantwoorde wijze uitvoeren. Er zijn cursussen van Agent Advisors om je te leren hoe je je agent goed kunt trainen. Technisch worden ze ook steeds veiliger gemaakt, zoals we ooit de auto veiliger maakten.

Zo wendden we de ramp af – hoe dachten we dat het ooit zonder kon?