

Werkwijze in het kort

VIJF TECHNISCHE WAARBORGEN ROND ONZE AI-AGENT – OM INTERN TE DELEN MET WIE HIEROVER MEE MOET BESLISSEN

pAinapple laat een AI-agent, met een eigen gebruikersaccount en eigen rechten, zelfstandig werk opleveren binnen een bedrijf. Voor wie dat vanuit compliance, IT of risicobeheer moet beoordelen, is de kernvraag steeds hetzelfde: wat weerhoudt zo'n agent ervan iets kapot te maken, of bij data te komen waar hij niet hoort? Onderstaand zijn de vijf waarborgen die dat afdwingen — geen belofte, maar de werkwijze waarmee wij zelf, vandaag, ook werken.

1 Elke wijziging eerst voorgesteld, nooit direct doorgevoerd

Een agent past nooit rechtstreeks iets aan in een productieomgeving. Iedere wijziging wordt eerst vastgelegd als een apart, leesbaar voorstel — met een duidelijk verschil ten opzichte van de vorige staat — voordat het actief wordt.

2 Altijd in een eigen, afgeschermd werk-kopie

Een agent werkt in zijn eigen geïsoleerde kopie van de omgeving, nooit in dezelfde kopie waar op dat moment een collega mogelijk zelf iets aan het uitproberen is. Twee taken kunnen elkaar zo niet in de weg zitten of overschrijven.

3 Een onafhankelijke controle voordat iets live gaat

Voordat een voorstel wordt doorgevoerd, beoordeelt een tweede, onafhankelijke stap het daadwerkelijke resultaat — niet enkel de eigen samenvatting van de uitvoerende agent over wat er zou zijn gedaan.

4 Snel kunnen terugdraaien, in plaats van alles vooraf dichttimmeren

De nadruk ligt niet op ieder denkbaar scenario vooraf afvangen, maar op razendsnel en volledig kunnen terugdraaien zodra iets niet klopt. Elke wijziging heeft daarom een schoon, direct bruikbaar terugdraaipunt.

5 Mensen beslissen over wat onomkeerbaar is

Een vaste, expliciete lijst van handelingen voert een agent nooit zelfstandig uit — het aanmaken van accounts, het intrekken of wijzigen van toegangsgegevens, wijzigingen aan domeinen, het verwijderen van back-ups, direct klantcontact. Die landen altijd eerst bij een mens, ongeacht hoe routinematig de rest van het werk is.

Dit zijn geen aanvullende maatregelen bovenop het werk van de agent — het is de manier waarop dat werk zelf is ingericht, bij elke wijziging, elke dag.