Triad

/EDITIE 1

Triad • Betrouwbaarheidsrapport • Editie 1

# Het AI-agent betrouwbaarheidsrapport voor vastgoedbeheer

/EDITIE 1

# Het AI-agent betrouwbaarheidsrapport voor vastgoedbeheer

Je team is de hele dag bezig met berichten lezen die geen antwoord nodig hebben, dezelfde updates in dezelfde systemen typen en de agenda's van monteurs puzzelen. Wij bouwen AI-medewerkers die dat werk overnemen: software die in je inbox, je ticketsysteem en je planning zit, leest en beslist als een ingewerkte collega, en het geval aan een mens overdraagt zodra ze twijfelt. Voordat we er één bij een vastgoedbeheerder neerzetten, hebben we gemeten. Dit rapport is die meting, cijfer voor cijfer, met daarbij hoe elk cijfer is geteld.

> BASIS VAN DE METING
>
> Alle resultaten zijn testresultaten, gemeten door Triad in haar eigen test- en validatieomgevingen, geanonimiseerd en geaggregeerd; geen enkel cijfer in dit rapport heeft betrekking op een specifieke klant, huurder, pand of leverancier.

/INHOUD

# In deze editie

/01 Een eigenaar zou een leverancier niet op zijn woord hoeven geloven

/02 Niemand zou betaald moeten worden om een bericht te lezen dat geen antwoord nodig heeft

/03 Wat er niet meer gebeurt

/04 Niets gaat live wat niet 19 van de 20 keer goed was

/05 De AI krijgt geen sleutels die jij haar niet hebt gegeven

/06 Een huurder zou geen 11 minuten moeten wachten op wat in 2 minuten kan

/07 Waar je ons aan mag houden


BIJLAGE


/A Hoe we hebben geteld

/B Waar de externe cijfers vandaan komen

/HOOFDSTUK 01

# Een eigenaar zou een leverancier niet op zijn woord hoeven geloven

AI kan een gedeelde inbox lezen, klachten sorteren en monteurs inplannen. De moeilijkere vraag voor elke eigenaar is eenvoudiger: kun je erop vertrouwen? De twijfel is reëel en meetbaar. In het brancheonderzoek van AppFolio uit 2026 gaf 78% van de vastgoedbeheerders aan dat zij nog niet kunnen vertrouwen op de AI-functies in hun bestaande software. Toch verwachten de bedrijven die AI wél hebben ingevoerd dat hun portefeuilles dit jaar met 31% groeien, bijna het drievoudige van de groei die bedrijven zonder AI verwachten.[1]

Die kloof, tussen twijfel en potentie, is waarom we dit rapport hebben geschreven. De meeste leveranciers vertellen je dat hun AI nauwkeurig is. Wij laten liever de telling zien: hoeveel gevallen, over welke periode, door wie gecontroleerd, en waar het misging. Elk cijfer hieronder heeft een steekproefgrootte en een betrouwbaarheidsinterval in de bijlage. Waar een getal een projectie is, staat dat erbij.

/HOOFDSTUK 02

# Niemand zou betaald moeten worden om een bericht te lezen dat geen antwoord nodig heeft

Eerst het probleem, in gewone taal. In een vastgoed- of facilitaire organisatie vreten drie soorten handwerk stilletjes de loonsom op:

| | |
|---|---|
| /01 | Lezen. Het meeste wat in een gedeelde inbox binnenkomt, vraagt geen enkele actie, maar iemand moet elk bericht lezen om dat te weten. |
| /02 | Overtypen. Elke melding wordt een ticket, een notitie, een update aan de huurder, een update aan de eigenaar, een werkbon voor een leverancier. Dezelfde feiten, vier keer getypt. |
| /03 | Puzzelen. Welke monteur, welke dag, welke route? Plannen is een dagelijkse legpuzzel, opgelost door wie er het minste tijd voor heeft. |

Niets hiervan is vakwerk. Alles hiervan is betaald werk. Dit is het werk dat onze AI-medewerkers weghalen, en dit is wat er gebeurde toen ze dat deden.

/HOOFDSTUK 03

# Wat er niet meer gebeurt

## /3.1 Niemand leest de inbox meer

In onze vastgoedbeheeromgeving hebben we elk bericht dat het systeem in vier dagen afhandelde, één voor één laten teruglezen door een mens. Het resultaat: 95% had helemaal geen menselijk antwoord nodig. Op dat zomertempo is een jaar van deze inbox 13.350 gelezen berichten, waarvan er maar 623 een mens nodig hebben. Bij één minuut per bericht is dat ruwweg 210 uur leeswerk per jaar dat niet meer plaatsvindt. De rest heeft het systeem zelf gelezen, gesorteerd en doorgezet: facturen, leverancierthreads, ruis. En het oordeelde juist. In een apart venster van 876 berichten negeerde het ongeveer 9 van de 10 rommelberichten terecht, terwijl het de ene echte noodmelding eruit pikte en als urgent markeerde. En omdat het software is, wordt de inbox om 3 uur 's nachts precies zo gelezen als om 3 uur 's middags: geen dienst, geen achterstand op maandagochtend.

FIGUUR 1 · LEESWERK WEGGEHAALD

13.350
→ 623

berichten per jaar, op het gemeten tempo: van 13.350 gelezen berichten hebben er maar 623 een menselijk antwoord nodig. De overige 95% is leeswerk dat simpelweg ophoudt te bestaan. (Vierdaagse telling van 150 berichten, geprojecteerd naar een jaar op dat tempo; de uren gaan uit van één minuut per bericht.)

## /3.2 Niemand plant de monteurs meer

In onze servicedienstomgeving plant de planningsengine de monteurs in: ze kent elke agenda, berekent reistijden en schrijft bij elke klus op waarom ze koos wie ze koos. Drie momenten uit het planningslog, één zomer:

/01 De noodmelding. Een urgente melding kwam binnen via het noodformulier. De engine had dezelfde ochtend een monteur op locatie en schoof een routineklus opzij om ruimte te maken. Niemand heeft een agenda aangeraakt.

/02 De 2½ uur die niemand heeft gereden. Drie monteurs waren beschikbaar: 186, 56 en 34 minuten rijden. De engine stuurde die op 34 minuten: 22 minuten dichterbij dan de op één na beste keuze, 2½ uur dichterbij dan de slechtste. Vermenigvuldig die keuze met elke klus, elke week: dat is brandstof, uren en humeur waar je niet meer voor betaalt.

/03 De gestolen sleutels. Een huurster die haar eigen voordeur niet meer op slot kreeg. De engine koos de monteur die een uur dichterbij was en vier dagen eerder kon dan het alternatief, en bundelde daarna een latere klus op hetzelfde adres, zodat één rit drie tickets oploste.

## /3.3 Niemand typt hetzelfde feit nog vier keer over

Als een melding binnenkomt, sorteert het systeem die in de juiste categorie (verwarming, lekkage, apparatuur), en die indeling bepaalt wie betaalt en wie wordt gebeld. We hebben dit op de harde manier getest: dezelfde gevallen, keer op keer aan het systeem gevoerd, over elke categorie die we afhandelen. Het gaf elke keer hetzelfde, juiste antwoord: 444 runs over alle 15 categorieën, nul afwijkingen. Een mens met een slechte dinsdag doet dat niet. Vanuit die ene beslissing volgen het ticket, het antwoord aan de huurder in de taal van de huurder, de update aan de eigenaar en de werkbon automatisch.

## /3.4 Een klantenservicelijn zou geen nachtdienst nodig moeten hebben

CASE: EEN COMPLETE WHATSAPP-SERVICELIJN, GERUND DOOR AI

In onze klantenserviceomgeving draait een complete WhatsApp-servicelijn op onze AI-medewerker: ze antwoordt dag en nacht, sorteert elke klacht, opent het ticket, en als een klant compensatie verdient, geeft ze zelf de kortingscode uit. Compensatie die geld kost, afgehandeld door software, binnen grenzen die ze niet kan overschrijden.

In één periode van twee weken handelde ze zo'n 215 gesprekken en taken per dag af. Van ruwweg 3.000 runs zijn er 3 gestopt of blijven hangen (99,9% runstabiliteit), en elk daarvan is opgemerkt door onze eigen monitoring, niet door een klant. Of de antwoorden zelf klopten, is apart gemeten, en daar komt het geld om de hoek: in twaalf dagen kregen 36 klanten precies één code, niemand kreeg er twee, en in 120 bewust lastige gevallen is geen enkele echte klant onterecht geweigerd.

Nog één cijfer dat een ondernemer zal waarderen: toen er een beter AI-model beschikbaar kwam, zijn we overgestapt en hebben we de draaikosten van dit systeem met een factor twintig verlaagd, van ongeveer 11 cent per gesprek naar ruim onder een cent, gemeten over 4.939 gesprekken, met de kwaliteit opnieuw getest en op peil. Je automatisering wordt goedkoper naarmate ze rijpt, niet duurder.

/HOOFDSTUK 04

# Niets gaat live wat niet 19 van de 20 keer goed was

Elke bewering hierboven houdt stand door hoe we werken, niet door geluk.
Drie gewoontes, in gewone taal:

/01 Niets gaat live onder 19 van de 20. Voordat een wijziging live gaat, draaien we haar twintig keer tegen dezelfde gevallen. Geeft ze niet minstens 19 van de 20 keer het juiste antwoord, dan gaat ze niet live. Punt.

/02 Honderden automatische controles, bij elke wijziging. Als we iets aanpassen, test het systeem zichzelf opnieuw tegen zijn eigen geschiedenis. Na één recente wijziging hebben we 588 eerdere gevallen opnieuw gedraaid: geen enkele uitkomst veranderde die niet had mogen veranderen.

/03 Een mens controleert de uitkomsten. Periodiek leest iemand die los staat van de bouw de uitkomsten één voor één terug, zoals bij de vierdaagse inboxcontrole hierboven, om te bevestigen dat het systeem deed wat het moest doen.

Van de drie controlelagen die we hebben gemeten, scoorde de zwakste 99,38% en het gepoolde resultaat over 1.638 runs 99,69%: hoe je het ook bekijkt een foutpercentage onder de 1%, met de volledige verantwoording in de bijlage voor wie onze rekensom wil nakijken. Dat is het punt van dit rapport: je hoeft ons nergens op ons woord te geloven.

/HOOFDSTUK 05

# De AI krijgt geen sleutels die jij haar niet hebt gegeven

De snelste manier om geld te verliezen met AI is haar sleutels geven die ze niet zou moeten hebben. Daarom komen onze systemen met harde grenzen, en één regel boven alles:

> Jij bepaalt wat ze mag versturen.
> Niet wij, en niet de AI.

Of de AI zelfstandig berichten mag versturen of werkbonnen mag aanmaken, is een keuze die jij per proces maakt, nooit een standaardinstelling. Standaard kijkt een mens naar alles voordat het de deur uitgaat. Waar je bewust voor autonomie kiest (zeg, routinebevestigingen aan huurders), gebeurt het versturen door vaste, voorspelbare software die een goedgekeurd plan volgt, met een automatische inhoudscontrole, nooit door een AI die vrij improviseert.

| | |
|---|---|
| /01 | Een onbekende leverancier krijgt nooit een werkbon; jij krijgt in plaats daarvan een melding. |
| /02 | Kosten boven het mandaat dat jij instelt, leiden eerst tot een goedkeuringsverzoek. |
| /03 | Als het systeem het niet zeker weet, stopt het en geeft het het geval met een samenvatting aan een mens; het gokt niet. |

Dit zijn geen beloftes; het is gemeten gedrag. De compensatiegrenzen in de servicelijn-case hierboven zijn die grenzen in werking, met geld op het spel.

/HOOFDSTUK 06

# Een huurder zou geen 11 minuten moeten wachten op wat in 2 minuten kan

Onafhankelijk onderzoek binnen de klantenservice in bredere zin laat zien wat medewerkers dagelijks ervaren: gepubliceerde branchecijfers melden dat AI verzoeken gemiddeld in 1,9 minuten afhandelt tegenover 11,4 minuten bij menselijke afhandeling, met een foutklachtpercentage van ongeveer een derde procent.[2] Andere processen dan die van ons, niet rechtstreeks vergelijkbaar, maar de richting is onmiskenbaar. De bedrijven die als eerste bewegen, krijgen het cumulatieve voordeel: elke maand van geautomatiseerde intake, sortering en planning is een maand loonkosten die in plaats daarvan aan werk gaat dat er wél toe doet.

> Het werk dat jouw team niet zou moeten doen, is bij ons al verdwenen.

/HOOFDSTUK 07

# Waar je ons aan mag houden

Als je met Triad werkt, zijn dit geen ambities. Het zijn de regels waaronder de cijfers hierboven tot stand zijn gekomen, en ze gelden vanaf dag één voor jouw implementatie:

/01 Niets gaat live onder 19 van de 20. Elk proces wordt twintig keer tegen jouw eigen gevallen gedraaid voordat het je inbox, je huurders of je leveranciers raakt.

/02 Elk cijfer komt met zijn telling. Je krijgt dezelfde verantwoording die je net hebt gelezen: steekproefgrootte, periode, missers. Maandelijks, voor je eigen organisatie.

/03 Een mens controleert de uitkomsten. Iemand die los staat van de bouw leest volgens een vast schema echte uitkomsten terug, en de missers worden aan jou gerapporteerd, niet weggestopt.

/04 Jij bepaalt wat ze mag versturen. Autonomie wordt per proces aangezet, door jou, nooit standaard. Onbekende leveranciers, kosten boven je mandaat en twijfelgevallen gaan altijd naar een mens.

Dit rapport is het eerste in een reeks; elke volgende editie voegt nieuwe implementaties en verse cijfers toe. Wil je weten wat een AI-medewerker van het bord van jouw team zou halen, en ons aan dezelfde bewijsstandaard houden die je net hebt gelezen, dan is het gesprek gratis.

Praat met Triad ↗

triadagency.ai/contact

Triad

/BIJLAGE A

# Hoe we hebben geteld

Elk cijfer in dit rapport, met steekproefgrootte, meetperiode en 95%-betrouwbaarheidsinterval (Wilson-methode). De steekproefgroottes hieronder zijn de daadwerkelijk gemeten aantallen; de jaarvolumes in de tekst zijn deze gemeten tempo's geprojecteerd naar een jaar, en staan overal als projectie aangeduid.

TABEL 1 · ELK CIJFER, MET ZIJN TELLING

| CIJFER (ZOALS GEBRUIKT IN DE TEKST) | N | RESULTAAT | 95%-BI | HOE GEMETEN |
|---|---|---|---|---|
| Inboxtelling: aandeel zonder menselijk antwoord | 150 | 95,3% | 90,7 — 97,7% | elk bericht over 4 dagen, teruggelezen door iemand los van de bouw; 143 van 150 hadden geen actie nodig, 7 wel |
| Inboxtelling: juist doorgezet | 150 | 99,33% | 96,32 — 99,88% | zelfde telling; 1 echte misser, regel dezelfde week hersteld |
| Rommelfilter | 876 | ≈ 90% | 87,79 — 91,77% | apart venster; rommelberichten terecht genegeerd, 1 echte noodmelding in het venster, als urgent gemarkeerd |
| Consistentie classificatie | 444 | 100% | 99,14 — 100% | dezelfde gevallen herhaald over alle 15 categorieën, getest vóór uitrol; 0 afwijkingen |
| Herhaling na een wijziging (regressie) | 588 | 100% | 99,35 — 100% | eerdere gevallen opnieuw gedraaid na één wijziging; 0 uitkomsten veranderd die niet hadden mogen veranderen |
| Melderherkenning (intake) | 59 | 88,14% | 77,48 — 94,13% | bruto; 3 van 7 missers waren externe partijen die terecht niet gekoppeld zijn (netto 93%) |
| Intake: beslislaag agent | 320 | 99,38% | 97,75 — 99,83% | gemeten runs; beide fouten opgevangen door automatische herhaling |
| Servicelijn: runstabiliteit | ≈ 3.015 | 99,90% | 99,71 — 99,97% | opeenvolgende runs over 2 weken, 3 incidenten (runs gestopt of blijven hangen), alle zelf gedetecteerd; meet afronding, niet antwoordkwaliteit |
| Servicelijn: end-to-end proces | 697 | 99,57% | 98,74 — 99,85% | gemeten runs, venster van 2 weken, 3 fouten |
| Servicelijn: klantopzoekstap | 621 | 100,00% | 99,39 — 100% | gemeten runs, venster van 2 weken, 0 fouten |
| Servicelijn: compensatiegrenzen | 120 | 0 onterecht geweigerd | n/a | twee onafhankelijke adversariële rondes van 60 runs; 36 codes uitgegeven in 12 dagen, 0 dubbele |
| Servicelijn: kosten per gesprek | 4.939 | € 0,11 → onder € 0,01 | n/a | draaikosten vóór en na de modelwissel, kwaliteit opnieuw getest op dezelfde gevallen |
| Gepoolde controlelagen | 1.638 | 99,69% | 99,29 — 99,87% | end-to-end proces + beslislaag agent + opzoekstap; worstcasefout 0,71%, onder de 1% uit de tekst |

/BIJLAGE B

# Waar de externe cijfers vandaan komen

/01 AppFolio, *Property Manager Benchmark Survey 2026*, appfolio.com/newsroom/property-manager-benchmark-survey-2026. Cijfers gecontroleerd tegen de live pagina op 26-08-2026. De 78% betreft AI-functies in bestaande vastgoedbeheersoftware.

/02 Digital Applied, *Customer Service AI Agent Statistics 2026* (bundeling van onderzoek van Zendesk, Salesforce, Gartner, Forrester, Intercom en McKinsey), digitalapplied.com/blog/customer-service-ai-agent-statistics-2026-data. Gecontroleerd op 26-08-2026.

/03 Over de meetmethode: *Towards a Science of AI Agent Reliability*, arXiv:2602.16666, ICML 2026, het onderzoeksstandpunt dat de herhaalde-metingaanpak van dit rapport volgt. Gecontroleerd op 26-08-2026.

NOOT OVER DE METING

Steekproefgroottes en meetperiodes staan per cijfer in Bijlage A; jaarvolumes in de tekst zijn gemeten tempo's geprojecteerd naar een jaar. Alle resultaten zijn testresultaten, gemeten door Triad in eigen test- en validatieomgevingen, geanonimiseerd en samengevoegd; geen enkel cijfer in dit rapport vertegenwoordigt een specifieke klant, huurder, pand of leverancier.