SCALION

SCALION INSIGHTS • RAPPORTAGE

Van chatbot naar collega *agentic AI in de klantenservice*

Wat de aanpak van KPN ons leert over kansen, risico's en de business case van agentic AI in klantcontact: een analyse van Scalion op basis van de McKinsey & Company case study over KPN's transformatie naar AI-gedreven klantenservice.

01

Scalion Research & Advisory | Augustus 2026

Gebaseerd op: McKinsey & Company, ‘How KPN is building an agentic AI engine for customer care’ (case study, McKinsey.com).

IN HET KORT

De belangrijkste cijfers uit de case

KPN, de grootste telecomaanbieder van Nederland, bouwt samen met McKinsey en QuantumBlack een herbruikbaar agentic AI-platform voor de klantenservice. In plaats van losse chatbotfuncties zet het bedrijf in op AI-agents die via spraak hele klantvragen zelfstandig afhandelen. Onderstaande cijfers geven de omvang en de eerste resultaten van dat programma weer.

10–20%
Van alle klantinteracties die KPN tegen 2027 AI-first wil laten verlopen

83
Klanttevredenheidsscore bij AI-afhandeling, vergelijkbaar met menselijke gesprekken

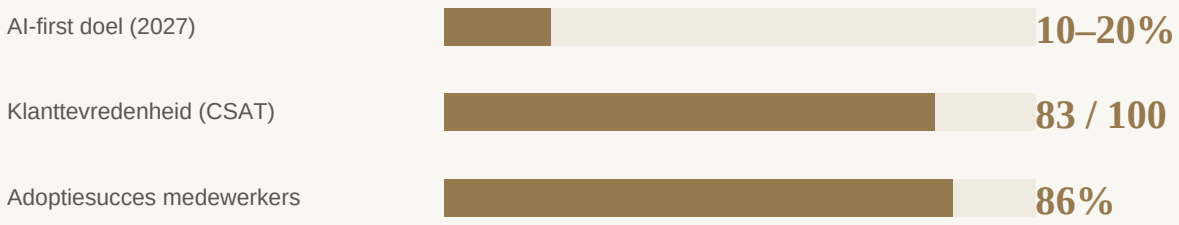
86%
Adoptiesucces onder medewerkers bij de invoering van de agents

≈ 5 mln.
Klantgesprekken per jaar waarop het programma uiteindelijk aangrijpt

100.000+
Klanten per week die de agents nu al te woord staan

< 2 sec.
Maximale reactietijd per gespreksbeurt om spraak-AI natuurlijk te laten aanvoelen

Kernresultaten in beeld



De drie meetbare resultaten van KPN's agentic AI-programma, als aandeel van het maximum. De eerste balk toont de ondergrens van de bandbreedte (10 procent).

Bron: McKinsey & Company, zie bronvermelding

CONTEXT

Van tekstuele chatbots naar spraak-agents

KPN is met bijna anderhalve eeuw geschiedenis de grootste telecomaanbieder van Nederland en profileert zich al langer als koploper bij technologische omslagen: van de uitrol van glasvezel naar twee derde van het land tot de eerste Europese aanbieder die realtime tekst tijdens mobiele gesprekken beschikbaar maakte voor alle klanten. Op het gebied van AI begon die koplopersrol met tekstuele chatbots, een agentic factuurverklaarder en een generieke vraag- en antwoordagent voor alle klanten.

Met de opkomst van steeds krachtigere spraakmodellen zag KPN een volgende stap: niet langer alleen tekst automatiseren, maar dynamische, spraak-naar-spraak interacties inzetten om wachttijden bij de

klantenservice te verkorten en de kwaliteit van dienstverlening te verbeteren. Het bedrijf past dit uiteindelijk toe op een schaal van ongeveer vijf miljoen klantgesprekken per jaar. Voor die stap sloeg KPN de handen ineen met McKinsey en zijn AI-tak QuantumBlack, met als doel niet zomaar losse AI-functies toe te voegen, maar een schaalbare, in-house agentic capability op te bouwen.

> *Wij zien AI als een manier om natuurlijkere interactie met klanten mogelijk te maken, terwijl menselijke ondersteuning beschikbaar blijft op de momenten die er echt toe doen.*
>
> **Jörg Kramer, Chief of Customer and Commercial Operations bij KPN**

Bron: McKinsey & Company, zie bronvermelding

AANPAK

Vier bouwstenen van de transformatie

De transformatie van KPN's klantcontact is opgebouwd rond vier prioriteiten: kansrijke toepassingen identificeren, een technisch fundament bouwen, oplossingen opschalen naar productie, en de organisatie voorbereiden op een nieuwe manier van werken.



De vier bouwstenen van KPN's agentic AI-transformatie, in volgorde doorlopen.

1. Kansrijke toepassingen in kaart brengen

Met een zelfontwikkeld large-language-model-raamwerk analyseerde het team grote hoeveelheden gespreks- en chattranscripten (geanonimiseerd en in lijn met Europese privacyregels) om te bepalen welke klantvragen het vaakst voorkwamen. Samen met klantervarings- en technische teams werden deze interacties vervolgens beoordeeld op impact, complexiteit en haalbaarheid, maar ook op de noodzaak om een mens in de lus te houden bij gevoelige momenten, zoals een verhuizing, waarbij het risico op klantverlies groter is. Dat leverde een scherpe prioriteitenlijst op: klantverificatie, statusvragen over bestellingen, het beheren van technicusafspraken en het oplossen van internetstoringen.

2. Een technisch fundament bouwen

Voor spraakinteracties van hoge kwaliteit moest de reactietijd per gespreksbeurt onder de twee seconden blijven. Menselijke gesprekken kennen doorgaans al reactietijden onder de seconde, waardoor elke vertraging een gesprek kunstmatig laat aanvoelen. Het team investeerde daarnaast fors in 'barge-in': klanten moeten een agent kunnen onderbreken zonder dat de context verloren gaat. Na afweging van verschillende architecturen koos KPN voor een geïntegreerd, beheerd spraak- en AI-platform,

gestructureerd in drie lagen (agent-orkestratie, het agentplatform en audio-integratie) en expliciet ontworpen voor hergebruik, zodat nieuwe toepassingen later op dezelfde basis kunnen worden gebouwd in plaats van telkens opnieuw te beginnen.



De drie samenhangende lagen van het platform, van klantgesprek tot backend-systemen.

3. Ontwerpen, opleveren en opschalen

Met korte build-test-release-cycli en vroege minimum viable products testten cross-functionele teams uitgebreid, onder meer met echte klanten in een UX-lab. Omdat AI-systemen niet altijd hetzelfde reageren op dezelfde input, kregen guardrails (de grenzen waarbinnen agents moeten blijven) en observability-tooling een centrale plek, aangevuld met een evaluatiekader dat menselijke beoordeling combineert met geautomatiseerde kwaliteitscontroles. Eenmaal live volgde een dagelijkse verbetercyclus: 's ochtends transcripten van tot honderd klantgesprekken beoordelen, 's middags prompts bijstellen op basis van wat daaruit naar voren kwam, en 's avonds de update live zetten.

4. De organisatie klaarstomen

KPN richtte een taskforce in om werkprocessen te herzien, trainingen te ontwikkelen en medewerkers te begeleiden terwijl hun rol verschoof: klantexperts richten zich sindsdien meer op complexe interacties waarin menselijk inzicht en empathie het verschil maken. Door frontliniemedewerkers actief te betrekken bij ontwikkeling en testen ontstond een sterke feedbackloop die zowel de klant- als de medewerkerservaring verbeterde, met een adoptiesucces van meer dan 86 procent tot gevolg.

Bron: McKinsey & Company, zie bronvermelding

KANSEN & RISICO'S

Wat dit voorbeeld aan mogelijkheden en aandachtspunten biedt

De KPN-case laat zien dat agentic AI in klantcontact verder reikt dan een chatbot die vragen beantwoordt, maar ook dat kansen en risico's hand in hand gaan. Onderstaande tabel geeft een snel overzicht; de toelichting volgt eronder.

KANSEN	RISICO'S
Kortere wachttijden en snellere afhandeling	Onvoorspelbaar gedrag van AI
Meer ruimte voor menselijk maatwerk	Privacy en gegevensbescherming
Schaalvoordeel door hergebruik	Afhankelijkheid van één platformleverancier
Datagedreven sturing op klantbehoeften	Gevoelige klantmomenten vragen mens in de lus
Wezenlijk andere, natuurlijke klantervaring	Verandermanagement wordt onderschat
Breed draagvlak haalbaar bij vroege betrokkenheid	Technische lat (latency, barge-in) ligt hoog

Kansen en risico's naast elkaar; zie de toelichting hieronder voor context per punt.

KANSEN NADER TOEGELICHT

Wat dit voorbeeld aan mogelijkheden biedt

Voor organisaties die eenzelfde pad overwegen, liggen hier de belangrijkste kansen:

- Kortere wachttijden en snellere afhandeling doordat AI-agents complete, veelvoorkomende interacties zelfstandig van begin tot eind afhandelen, in plaats van alleen een eerste antwoord te geven.
- Meer ruimte voor menselijk maatwerk: agents nemen routinematige taken zoals identiteitscontrole over, waardoor medewerkers zich kunnen richten op complexe of emotioneel gevoelige gesprekken.
- Schaalvoordeel door hergebruik: een gedeeld, herbruikbaar platform maakt het toevoegen van nieuwe toepassingen sneller en goedkoper dan losse puntoplossingen per use case.
- Datagedreven sturing: doorlopende analyse van gesprekstranscripten geeft continu zicht op nieuwe klantbehoeften en verbeterpunten in de dienstverlening.
- Een wezenlijk andere klantervaring: natuurlijke spraakinteractie met 'barge-in' voelt fundamenteel anders dan keuzemenu's of trage chatbots, mits de techniek de latency-eis waarmaakt.
- Breed draagvlak is haalbaar wanneer medewerkers vroeg worden betrokken bij ontwerp en test: KPN's adoptiesucces van 86 procent onderstreept dat dit geen louter technisch traject is.

RISICO'S NADER TOEGELICHT

Waar organisaties rekening mee moeten houden

Tegenover de kansen staan risico's die niet vrijblijvend zijn zodra AI-agents op grote schaal met klanten in gesprek gaan:

- Onvoorspelbaar gedrag van AI: hetzelfde type vraag kan tot verschillende antwoorden leiden, wat doorlopende guardrails, monitoring en evaluatie noodzakelijk maakt. Zonder die waarborgen zijn kwaliteit en veiligheid niet te garanderen.
- Privacy en gegevensbescherming: gespreksdata bevat persoonsgegevens; analyse en training moeten aantoonbaar voldoen aan de AVG en andere Europese privacyregels.
- Afhankelijkheid van één platformleverancier: de keuze voor een geïntegreerd, beheerd platform versnelt ontwikkeling, maar vergroot ook de afhankelijkheid van die ene leverancier.
- Gevoelige klantmomenten: bij situaties met een hoger risico op klantverlies, zoals een verhuizing, blijft een mens in de lus nodig; volledige automatisering is daar niet vanzelfsprekend wenselijk.
- Verandermanagement onderschat: zonder heldere veranderaanpak, duidelijke meetpunten en actieve betrokkenheid van medewerkers loopt adoptie het risico vast te lopen, ongeacht de kwaliteit van de technologie.
- Technische lat ligt hoog: reactietijden onder de twee seconden en natuurlijke onderbreekbaarheid ('barge-in') zijn technisch veeleisend; een zwakke uitvoering ondermijnt het vertrouwen in de oplossing direct.

Bron: McKinsey & Company, zie bronvermelding

BUSINESS CASE

Financiële mogelijkheden: wat zit erachter?

De gepubliceerde case study van McKinsey en KPN noemt geen concrete euro-bedragen voor besparingen of investeringen: de gedeelde resultaten zijn operationeel (adoptie, klanttevredenheid, streefpercentage AI-afhandeling) in plaats van financieel. Toch is op basis van de beschreven schaal en aanpak een aantal financiële mechanismen te herleiden die relevant zijn voor elke organisatie die een vergelijkbare business case wil doorrekenen:

≈ 5.000.000
Klantgesprekken per jaar (totaal)

→

10–20%
AI-first doel van KPN tegen 2027

→

500K–1M+
Gesprekken potentieel AI-afgehandeld

Indicatieve volumeberekening op basis van de door KPN gedeelde cijfers; dit is geen door McKinsey of KPN gepubliceerde prognose.

- Volume bepaalt de hefboom: bij circa vijf miljoen klantgesprekken per jaar en een streefpercentage van 10 tot 20 procent AI-first tegen 2027, gaat het uiteindelijk om potentieel honderdduizenden tot ruim een miljoen gesprekken per jaar die niet langer volledige callcentercapaciteit vragen.
- Kosten per interactie dalen wanneer afhandeltijd korter wordt en problemen, zoals internetstoringen, direct worden opgelost zonder dat daar een technicusbezoek voor nodig is.

- Marginale kosten per nieuwe toepassing nemen af: doordat het platform bewust herbruikbaar is opgezet, is de investering voor een volgende use case beduidend lager dan bij losse, eenmalige oplossingen.
- Vrijgespeelde mensuren krijgen meer waarde doordat medewerkers zich focussen op complexe of retentiegevoelige gesprekken, waar menselijk contact het grootste verschil maakt voor klantbehoud.
- Investeringen liggen vooral vooraan in het traject (in platform, guardrails, evaluatiekader en verandermanagement), terwijl de baten toenemen naarmate meer volume en meer use cases op hetzelfde fundament meeliften.

Voor organisaties die dit willen doorrekenen, geldt: bouw de business case op basis van eigen gespreksvolumes, gemiddelde kosten per interactie en het aandeel vragen dat realistisch geautomatiseerd kan worden, en behandel de bovenstaande mechanismen als richtinggevend, niet als toegezegd rendement.

MARKTONTWIKKELING

Waar staat de sector, en waar gaat dit heen?

De stap van KPN past in een bredere beweging in klantcontact en de dienstensector:

- Van tekst-chatbots naar spraak-native, agentic systemen die complete workflows afhandelen (zoals verificatie of het inplannen van een afspraak) in plaats van alleen vragen te beantwoorden.
- Telecom loopt hierin voorop omdat het gespreksvolume hoog en repetitief genoeg is om automatisering lonend te maken, terwijl de sector tegelijk gevoelig genoeg is om zorgvuldig ontwerp rond menselijke betrokkenheid af te dwingen.
- Meer grote dienstverleners (in telecom, financiële dienstverlening, verzekeringen en nutsbedrijven) zullen naar verwachting een vergelijkbaar pad volgen: een eigen, herbruikbaar platform bouwen in plaats van losse AI-functies in te kopen.
- Governance en observability worden een randvoorwaarde in plaats van een extra, naarmate agents via API's en protocollen zoals MCP steeds meer directe acties in kernsystemen uitvoeren in plaats van alleen te antwoorden.
- De combinatie van technologie én organisatieverandering bepaalt welke organisaties hier daadwerkelijk snelheid mee maken: KPN's eigen ervaring laat zien dat adoptie net zo bepalend is als de architectuurkeuzes.

Bron: McKinsey & Company, zie bronvermelding

PRAKTIJKLESSEN

Zes lessen uit de praktijk van KPN

1. Begin bij de ervaring, niet bij de technologie

Architectuurkeuzes werden gestuurd door klantervaringseisen zoals latency, natuurlijke beurtwisseling en 'barge-in', niet andersom. Het streven naar reactietijden onder de twee seconden bepaalde keuzes door de hele stack heen, van de gesprekslaag tot de orkestratie.

2. Houd het eerst simpel om snelheid te maken

Door te kiezen voor één geïntegreerd gesprek- en agentplatform hield het team de integratiecomplexiteit laag en de time-to-market kort, met als keerzijde een grotere afhankelijkheid van die ene leverancier.

3. Orkestratie is de ruggengraat van opschaling

Het coördineren van meerdere gespecialiseerde agents vraagt een robuuste orkestratielaag voor routering, overdracht en context. Dat wordt steeds belangrijker naarmate het aantal use cases groeit.

4. Bouw voor integratie, niet voor eilandjes

Agents zijn gekoppeld aan backend-systemen via gestandaardiseerde API's en het MCP-protocol, zodat ze niet alleen vragen beantwoorden maar volledige werkstromen (zoals verificatie of het inplannen van een afspraak) kunnen afronden.

5. Observability en controle zijn niet optioneel

Het platform bevat end-to-end logging, tracing en evaluatie om agentgedrag te monitoren en grenzen af te dwingen. Guardrails bepalen wat agents wel en niet mogen zeggen of doen; het evaluatiekader bewaakt kwaliteit en veiligheid op schaal.

6. Ontwerp voor hergebruik en langetermijnschaal

De architectuur is gebouwd om meerdere use cases en groeiende volumes te ondersteunen, met herbruikbare bouwstenen voor agents, data-toegang en orkestratie, zodat uitbreiding niet telkens bij nul hoeft te beginnen.

CONCLUSIE

Wat betekent dit voor jouw organisatie?

KPN laat zien dat agentic AI in klantcontact inmiddels verder gaat dan een experiment: het is een productieomgeving die wekelijks meer dan honderdduizend klanten bedient, met een klanttevredenheid die de vergelijking met menselijke afhandeling doorstaat. Tegelijk maakt de case duidelijk dat dit resultaat geen toeval is: het is de optelsom van een scherpe, datagedreven prioritering van use cases, een technisch fundament dat is gebouwd voor hergebruik in plaats van eenmalige inzet, stevige guardrails en observability, en een organisatiebrede aanpak van verandering.

Voor de meeste organisaties is de vraag niet of agentic AI in klantcontact relevant wordt, maar hoe snel en op welke manier. De KPN-case biedt daarvoor een concreet ijkpunt: begin klein en datagedreven, bouw

één herbruikbaar fundament in plaats van losse functies, en behandel medewerkersadoptie als een even groot project als de techniek zelf.

> Benieuwd wat een vergelijkbare aanpak voor jouw klantcontact kan betekenen, in kansen, risico's én rendement? Scalion denkt graag mee over de eerste stappen naar een schaalbaar, agentic klantcontactplatform. www.scalion.nl.

Bronvermelding: deze rapportage is een analyse en samenvatting door Scalion, gebaseerd op McKinsey & Company / QuantumBlack, case study ‘How KPN is building an agentic AI engine for customer care’, gepubliceerd op mckinsey.com. Cijfers en citaten in dit document zijn ontleend aan die publicatie; interpretaties, risico-inschattingen en de sectie over financiële mogelijkheden zijn duiding van Scalion en dienen niet te worden gelezen als door McKinsey of KPN bevestigde prognoses.