CROWNHILL COURT

CASE STUDY

AI-led customer care transformation for a global online learning marketplace

AI resolution 33% → 52% · CSAT held at 88% · $2m+ saved a year

September 2026

CONFIDENTIAL

Case study: AI now resolves half of all customer contacts for a global online learning marketplace saving $2m a year

Situation

- Venture-backed global marketplace connecting learners with tutors; 40,000 support conversations a month in many languages
- Legacy decision-tree workflows deflected about a third of contacts but rarely resolved them
- Refunds, cancellations and tutor changes all needed a human agent
- Level 1 support could not scale with the company's growth

What we did

- Mapped every contact reason by volume, complexity and satisfaction, then routed each to workflow, AI agent or person
- Chose the AI platform for its ecosystem and switching costs, not the demo
- Built the knowledge base, procedures and guardrails the AI needed to act, and stood up a dedicated content team
- Launched in waves, simplest intents first; redesigned the refund journey to retain before refunding
- Replaced deflection as the success measure with resolution, effort and CX scores

Measurable outcomes

- AI resolves 52% of all contacts, up from 33% with workflows alone
- ~80% satisfaction on AI-resolved conversations; 88% overall CSAT maintained
- $2m+ annual savings and rising, with Level 1 headcount held flat through growth
- Retention and gross margin up on the refund journey
- Weekly failure review and playbook now run by the client's own team

52%
of contacts resolved by AI, from 33%

88%
overall CSAT maintained

$2m+
annual savings, rising

40,000
conversations a month handled

Client unnamed at its request. Source: client management reporting, September 2026; deflection and resolution figures as reported by the client’s customer-care leadership.

The starting point – 40,000 conversations a month, a third deflected by scripted workflows and almost none of them resolved

Situation

- A venture-backed global marketplace connecting learners with tutors, growing fast and supporting customers in many languages
- Support ran on the existing helpdesk platform with decision-tree workflows: fine for narrow, linear intents, useless for ambiguous or multi-intent contacts
- Any contact with money or an account action attached (refunds, cancellations, tutor changes) went straight to a human agent
- A three-tier agent team, with Level 1 carrying the highest volume and lowest complexity, could not be hired fast enough to keep pace with growth
- CSAT was the only quality measure, so nobody could see whether a deflected contact had been resolved or simply re-contacted later

Where contacts went before the programme, share of monthly conversations



Deflected is not resolved: workflows kept a contact away from an agent but often left the customer to come back

Why it did not scale

Wrong tool for ambiguous intents
Workflows need a clear, single intent and a linear path; most real contacts have neither

No actions without a person
The platform could answer but not act, so every refund or account change queued for an agent

Blind measurement
Deflection and CSAT rewarded keeping people away from agents, not solving their problem

The brief: get half of all contacts resolved by AI without hurting satisfaction, and build something the client's own team could run after we left.

Client unnamed at its request. Source: client management reporting and contact-reason analysis, September 2026.

Our approach – four phases over 18 months, simplest and highest-volume intents first, with live traffic from month three

	Diagnose 4 weeks	Pilot 6 weeks	Wave 1 Months 3–6	Wave 2 Months 6–18
What we did	• Mapped every contact reason by volume, complexity, satisfaction and cost to serve • Inventoried systems, knowledge and data sources • Evaluated AI providers on ecosystem, integrations and switching cost	• Put simple, high-volume intents on the AI agent with live traffic and a fast fallback to agents • Defined resolution rate, user effort score and CX score alongside CSAT • Baselined every metric	• AI answers replaced scripted workflows across all contact reasons • Help-centre content rebuilt as the AI's knowledge source • Dedicated content and procedures team stood up	• AI executes refunds, cancellations and tutor changes under retention-first rules • Human approval built in for sensitive actions • Agent tiers redesigned, copilot rolled out to agents
What the client got	• Contact-reason map and routing rules • Quantified business case • Provider decision	• Working AI agent on first intents • Metric set and dashboard • Go / no-go evidence for Wave 1	• Deflection from 33% to ~45% • Content team and playbook v1 • Weekly failure review	• Deflection to 52%, CSAT held at 88% • $2m+ annual savings • Handover to the client's own team

Benefit tracking every month against the baselined business case

Weekly working session with the customer-care and IT leads; monthly steerco with the executive sponsor

Voice remains outside the programme scope; the same knowledge and procedures are designed to extend to it. Phase durations reflect the programme plan agreed with the client in the Diagnose phase.

Diagnose – every contact reason was mapped, then routed by intent clarity and complexity to a workflow, the AI agent or a person

Contact-reason map

Every reason a customer gets in touch, scored on:

- Volume
- Complexity and number of intents
- Satisfaction today
- Cost to serve
- Risk: money, policy or legal exposure

Built in cross-functional workshops with support, product and finance, then verified against transcripts and contact data

Scripted workflow

Narrow, clear intent, linear path

- One intent, one answer path, no judgement needed
- Free within the existing platform
- Examples: password reset, lesson schedule, invoice copy

AI agent

Ambiguous or multi-intent, needs a conversation

- Clarifies what the customer wants before acting
- Answers from the knowledge base; executes procedures such as refunds and tutor changes
- Priced per resolution, so only worth it where workflows fail

Human agent

Judgement, sensitivity or policy

- Legal and compliance handovers by design
- Approval step for high-value or irreversible actions
- Edge cases; 20–30% of contacts will stay here

Provider choice: we scored candidates on the ecosystem, not the demo

Claimed resolution rates
Treated as a starting point only; verified in the pilot on live traffic

Surrounding ecosystem
Knowledge, procedures, agent copilot and integrations already in the client's helpdesk

Switching cost
Content, procedures and integrations lock a client in; the decision had to be right first time

Measurement – deflection flattered the results, so we tracked resolution, customer effort and conversation quality alongside CSAT

Deflection rate counts contacts kept away from agents. It says nothing about whether the customer got an answer, how hard they had to work for it, or whether they came back a day later. Four metrics now decide whether the AI is working.

CSAT Baseline	Resolution rate Did it get solved?	User effort score How hard was it?	CX score Was it a good conversation?
What it measures	**What it measures**	**What it measures**	**What it measures**
Satisfaction survey at the end of every interaction	Share of AI conversations that needed no human agent and produced no re-contact within a set period	Composite of the number of customer inputs, total time to resolution and re-contact rate	AI-based evaluation of each transcript: customer sentiment and whether a solution was actually provided
How we used it	**How we used it**	**How we used it**	**How we used it**
Kept as the gap indicator between AI-handled and human-handled conversations, not as the target	The primary success measure; a deflected contact that re-contacts counts as a failure	Catches chatbots that resolve eventually but exhaust the customer; used to tune every journey	Reads every conversation, not the minority who answer a survey; flags generic or tone-deaf answers

Business outcomes sit above all four. The first refund automation scored well on CSAT and badly on churn, so retention and gross margin were added to the dashboard and every journey is judged against them, not satisfaction alone.

The AI is a product, not a plug-in – it needed knowledge, procedures, guardrails and a team to own them

The biggest lesson of the programme: switching the product on was the smallest part of the work. Most of the effort, and most of the results, came from the four things around it.

Knowledge
What the AI knows

- The customer help centre is the AI's primary knowledge source, so it was rebuilt article by article
- Written once in plain English; the AI translates for other languages
- Content gaps found in the weekly review go straight into the writing queue

Procedures
What the AI may do

- A step-by-step procedure per contact reason defines the actions the AI can take: issue a refund, cancel a subscription, change a tutor
- An internal flow chart per contact reason specifies the behaviour before anything is built
- Started with low-risk, high-volume intents; money-moving actions came last

Guardrails
Where a person steps in

- Human-in-the-loop approval before sensitive actions execute
- Approved response templates written by the care management team, so answers stay consistent and never improvise policy
- Legal and compliance handovers are designed in, not treated as failures

Team
Who owns it

- A dedicated content and procedures team, trained during the programme, now owns the AI day to day
- Runs the weekly unresolved-conversation review and the change queue
- Uses AI itself to draft macros and templates at scale

Together these form the AI Service Playbook: one living document of knowledge, procedures and guardrails that the client owns, so the AI keeps improving after we leave and a provider change would not start from zero.

Worked example – the refund journey was redesigned to retain first and refund last, lifting retention and gross margin

Before
First automation

Customer asks for a refund

AI issues the refund immediately

Result
High CSAT, high churn: the customer left with their money and the marketplace lost the learner and the margin

After
Retention-first procedure

1. Ask why
Understand the reason before acting: tutor fit, schedule, price, quality

2. Offer an alternative
Propose a new tutor or reschedule that solves the stated problem

3. Offer credits
If declined, offer platform credits so the value stays with the learner

4. Refund
Only then process the refund, within policy and with approval where required

What changed

Retention up
Learners who came in for a refund left with a new tutor or credits instead of leaving the platform

Gross margin up
Fewer refunds paid out; credits keep the revenue on the platform

Effort slightly up, deliberately
Two more questions than before, accepted as the price of a better business outcome; effort is tracked to make sure it stays small

Repeatable pattern
The ask-why, offer, credit, act sequence is now the template for any journey where the AI moves money

Lesson: optimise the AI for business outcomes, not for satisfaction alone. A chatbot that says yes to everything scores well and quietly costs money.

Continuous improvement – a weekly review sorts every unresolved AI conversation into four causes, each with an owner and a fix

Pull
Every conversation the AI handed to an agent or that re-contacted within the window

Classify
An LLM reads each transcript and assigns one of four causes; the team samples to check it

Route
Each cause goes to a named owner with a change request in the queue

Ship and re-measure
Fixes go live weekly; resolution rate for that contact reason is tracked the following week

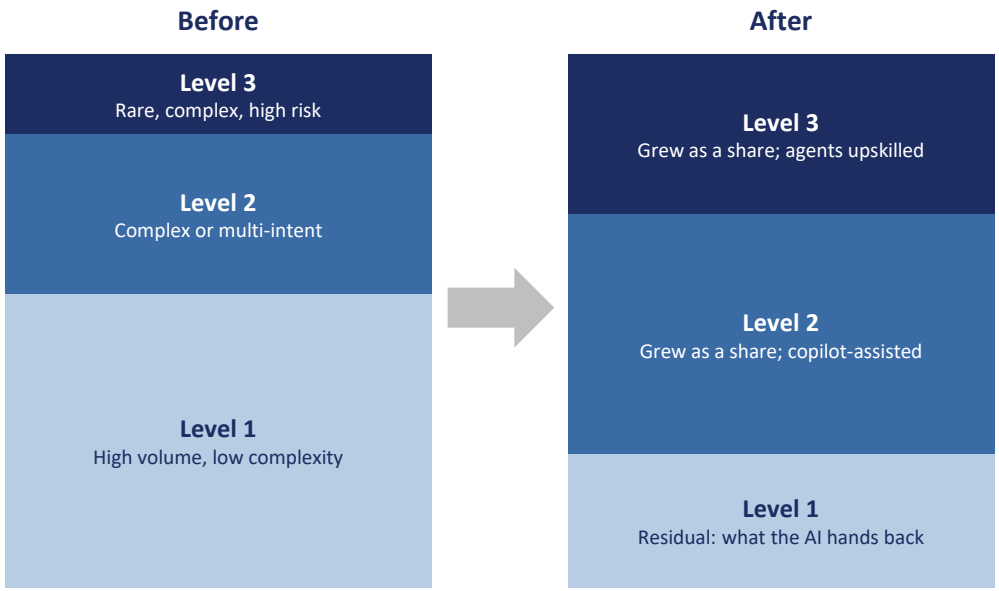
The four causes, and what each one triggers

Content gap	No action available	Generic answer	Policy handover
What happened No information existed to answer the question	**What happened** The AI understood the request but had no procedure to execute it	**What happened** The reply was correct in general but ignored this customer's credits, tutor or account	**What happened** Escalation was intentional: legal, compliance or risk appetite required a person
The fix Write or extend the help-centre article; the AI picks it up automatically	**The fix** Design the flow chart, build the procedure, decide whether it needs human approval	**The fix** Add the account context to the prompt and template so the answer is personal	**The fix** No fix. Counted separately so it never inflates the failure rate
Owner: Content team	Owner: Procedures team with product	Owner: Procedures team with engineering	Owner: By design

Why it matters: AI scales problems as fast as it scales answers. A missing step that one agent quietly covered (an apology when the customer had a point, a question the agent then asked again) shows up thousands of times a month once the AI handles it. The review is how those gaps are found before customers report them, and it feeds the same lesson back into agent handovers so a customer is never asked twice.

People – Level 1 work moved to the AI and the human team shifted up a tier, with headcount held flat through the company's growth

How the human team's work changed, illustrative share of agent-handled volume



Level 1 shrank as the AI absorbed simple contacts; Levels 2 and 3 grew as a share. Total headcount stayed roughly the same because the company kept growing; the saving is the hiring that never had to happen.

What changed for the agents

Upskilled, not replaced
Agents moved from scripted Level 1 work to the judgement cases the AI hands over; training was redesigned around them

Copilot inside the chat
An AI copilot surfaces the knowledge base and suggested answers in the agent's own screen, so Level 2 cases are handled faster

AI writes the boring parts
The training team uses an LLM to draft macros and response templates at scale, reviewed by people before use

No repeat questions
Handover rules pass what the AI already learned to the agent, so customers are not asked the same thing twice

Where automation stops. Around 20–30% of contacts will always need a person: complexity, edge cases, or policy. Automating the easier 80% delivers most of the return; pushing into Levels 2 and 3 costs more training per contact, carries more risk and yields less volume, so it is a deliberate later choice, not a target.

Languages – one English knowledge base translated by the AI, with local articles added only where resolution lagged

English help centre
Single source of truth, written and maintained once by the content team

AI translates in conversation
Customers get answers in their own language; approved templates keep tone and policy consistent

Resolution tracked per language
The same resolution, effort and CX metrics, cut by language every week

Localised articles where needed
Where a language lags, native articles are written and tested; kept only if resolution improves

What we found

- Performance is similar but not identical across languages; translation quality is not the whole story, local phrasing and expectations matter
- Japanese and French showed resolution and satisfaction gaps against English, so both are monitored closely and localised content is being tested
- Localised articles for Southeast Asian expansion markets improved resolution rates, which justified the extra writing effort there
- Consistency beats fluency: giving the AI the management team's approved templates prevents inventive answers that are fluent, confident and wrong

The operating rule

Write once in English.
Every article and procedure has one owner and one version.

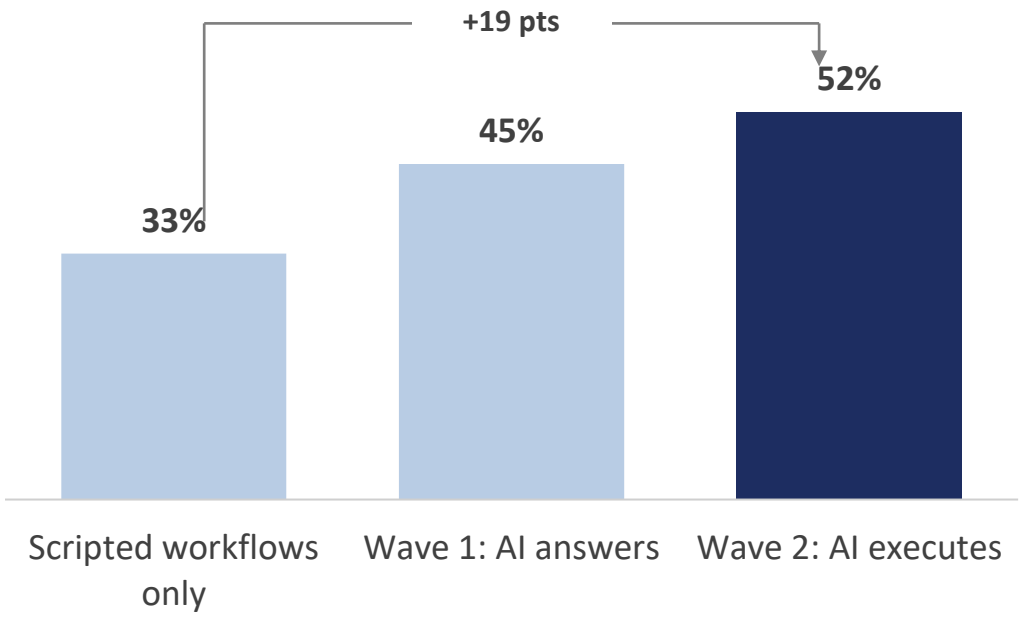
Let the AI carry the languages.
Translation is where the operating leverage comes from; a native article costs writer time in every market.

Localise on evidence, not instinct.
A language earns local content when its resolution rate falls behind, and keeps it only if the gap closes.

Why it matters for a multi-country client: the same knowledge base can serve every market from day one, and the metrics tell you which two or three languages need a local writer rather than guessing.

Results – AI resolution rose from 33% to 52% of all contacts while CSAT held at 88% and annual savings passed $2m

Share of monthly contacts resolved without a human agent, by programme stage



Source: client management reporting, September 2026. Wave 1 figure approximate.

Measurable outcomes

52% of all contacts resolved by AI
Up from 33% with scripted workflows alone; ~40,000 conversations a month

~80% satisfaction on AI-resolved conversations
Overall CSAT maintained at 88% throughout the programme

$2m+ annual savings, and rising
Level 1 headcount held flat while the company kept growing

Retention up on the refund journey
Retention-first procedure lifted retention and gross margin; effort tracked to stay low

What the client owns now: a contact-reason map, an AI Service Playbook of knowledge, procedures and guardrails, a four-metric dashboard, a weekly failure review and a trained content team. The AI keeps improving without us.

Lessons – the seven mistakes that stall AI customer-care programmes

#	The mistake	What we do instead
1	**Treating the AI as a switch to flip**	Content, procedures and integrations are most of the work; budget and staff for them from day one
2	**No dedicated team behind it**	Stand up a small content and procedures team during the pilot and hand the AI to them at the end
3	**Measuring quality with CSAT alone**	Track resolution rate, user effort and CX score alongside CSAT, and set targets on resolution
4	**Optimising satisfaction over business outcomes**	Judge every journey on retention and margin as well; a refund given instantly is a happy customer who left
5	**Launching across every contact reason at once**	Start with low-risk, high-volume, single-intent contacts; add money-moving actions only once the metrics hold
6	**Generic answers that ignore the customer's situation**	Give the AI the account context (credits, tutor, history) and approved templates, so answers are personal and consistent
7	**Asking the customer the same thing twice**	Design the handover so what the AI learned travels with the conversation to the agent

Common thread: every one of these is an operating-model mistake, not a technology one. The provider matters less than the processes, content and team around it.

What this means for your programme – the lessons and the assets transfer

What we learned

- **AI scales quality and problems alike.** Good processes improve at scale; every undocumented step becomes visible at scale too. Define the process first.
- **Deflection is not success.** Resolution rate, user effort and CX score are the measures that predict whether customers and margins are better off.
- **The AI is a product.** It needs continuous content, procedures and a team with a budget, not a one-off deployment.
- **Workflows and AI agents are complementary.** Clear, narrow intents stay on cheap workflows; ambiguous ones go to the AI; policy cases go to people.
- **Choose the provider for the ecosystem.** Switching costs are high once content, procedures and integrations exist, so the first choice has to hold.
- **Plan for a 20–30% human floor.** Full automation is not the goal; the right split between AI and people is.

What we bring to the next client

Contact-reason mapping method
Workshop format, scoring model and routing rules: workflow, AI agent or person

Four-metric measurement set
Definitions, dashboard design and targets for resolution, effort, CX and CSAT

AI Service Playbook template
Structure for knowledge, procedures, guardrails and approval rules the client will own

Failure taxonomy and weekly review
Four causes, owners and the LLM classification process, ready to run from Wave 1

Retention-first journey patterns
Ask why, offer an alternative, offer credits, then act: applied to refunds and other money-moving contacts

Provider evaluation framework
Scoring on ecosystem, integrations, pricing model and switching cost, verified in a live pilot

Typical shape of an engagement: a four-to-six week diagnostic that produces the contact-reason map, business case and provider decision, then a pilot on live traffic within the first quarter, with the client's own team trained to run the AI by the end.