From 3% sampled QA to 100% coverage across a mid-market BPO — AI quality operations

Case study · BPO · AI Ops

A mid-market BPO's QA program sampled 3% of interactions and clients still argued the scorecards. pronix.ai stood up 100% AI-scored QA with human calibration and evidence linking — 4x more coach-worthy findings and a 22% CSAT lift across the top three programs.

CLIENT
Mid-market BPO, 8,500 seats

INDUSTRY
BPO

PLATFORM
GENESYS CLOUD CX
AWS BEDROCK
SNOWFLAKE

3% → 100%
QA coverage

4x
Coach-worthy findings surfaced

+22%
CSAT on covered programs

-83%
Client scorecard disputes

01 The situation

A mid-market BPO's QA program sampled 3% of interactions and clients still argued the scorecards. pronix.ai stood up 100% AI-scored QA with human calibration and evidence linking — 4x more coach-worthy findings and a 22% CSAT lift across the top three programs.

The QA sample missed the interactions that actually mattered, coaching was reactive, and clients disputed 1 in 6 scores. Every new program required a scorecard rebuild in a spreadsheet.

3% → 100%	4x	+22%	-83%
QA coverage	Coach-worthy findings surfaced	CSAT on covered programs	Client scorecard disputes

02 How we delivered it

STEP 01

Scorecard-as-config

Modeled every client scorecard as versioned config — rubric, weights, evidence rules — so a new program went live in a day.

STEP 02

LLM scoring with citations

Every score linked to the transcript span that produced it — no black-box grades in a client review.

STEP 03

Human calibration loop

5% of interactions dual-scored by a QA analyst; drift monitored per rubric and re-trained monthly.

STEP 04

Coaching queue

Findings routed to supervisors as one-minute coaching cards tied to the exact call moment — not a weekly PDF.

STEP 05

Client scorecard portal

Read-only portal for the client with drill-down to evidence — disputes dropped by design.

03 Reference stack and assumptions

LAYER	COMPONENT	DESIGN ASSUMPTION
Contact center	Genesys Cloud CX	Voice + digital recording, transcripts and metadata streamed to Snowflake via Genesys AppFoundry connector.
LLM / scoring	AWS Bedrock (Claude 3.5 Sonnet + Haiku)	Haiku for classification / adherence; Sonnet for open-ended rubric items with citation extraction.
Scorecard config	Custom scorecard-as-code service	Rubrics stored as versioned YAML per client; new program live in one day with a scorecard PR.
Data platform	Snowflake + dbt	Every score, evidence span and dispute recorded; longitudinal drift tracked per rubric per program.
Coaching surface	Genesys native + Slack cards	One-minute coaching cards routed to supervisors with deep-links to the exact call moment.
Client portal	Retool (read-only) on Snowflake	Clients see scores, evidence and dispute status; SSO via Okta or Azure AD per client.

Written for

BPO COO · Head of Quality · Client Services Director — and the delivery leaders who have to make the same call on their own estate.

THE BOTTOM LINE

Every engagement ends in production, under one SLA. The team that builds it runs it — which is why these numbers are measured against a baseline captured before anything shipped, not estimated afterwards.

Run the same play on your estate. Book a working session with the practitioners who delivered this program, or model the business case first with our ROI calculators.

pronix.ai/book
pronix.ai/roi