SurgeX LABS

CASE STUDY · PROPTECH SAAS — LEASING OPERATIONS

Every sales tour, scored and coached.

AI coaching on real conversations — built and operated by a two-to-three-engineer AI-native pod.

CLIENT: HARBORVIEW SOFTWARE *

0	0/36	[N]
FABRICATED CITATIONS IN A 371-TOUR AUDIT	SPEAKER-ROLE SWAPS IN THE ATTRIBUTION AUDIT	TOURS ANALYZED MONTHLY ACROSS [M] COMMUNITIES [CONFIRM]

The client & the challenge

ABOUT HARBORVIEW SOFTWARE

Harborview Software records and analyzes in-person leasing tours for multifamily operators. Thousands of tours a month flow through its platform, and its customers coach, staff, and reward leasing teams based on what it reports — so a wrong claim in a coaching report is not a cosmetic bug, it is a personnel decision made on bad evidence.

THE CHALLENGE

Operators were sitting on hours of recorded tours nobody had time to hear. Managers wanted every tour scored against their own playbook; associates wanted feedback that quoted what they actually said.

Early LLM prototypes failed in the worst possible way: they scored fluently and occasionally invented quotes that were never said. On poor audio, the transcription layer itself fabricated words. And when four people speak on a walking tour, off-the-shelf diarization merged speakers — attributing one person's words to another. In a product that judges people's work, every one of those errors is disqualifying.

AT A GLANCE

CLIENT
Harborview Software *

INDUSTRY
Proptech SaaS — leasing operations

SYSTEM
AI coaching on real conversations

POD
2 engineers

STATUS
In production, multi-tenant

ENGAGEMENT
Sprint → Build → Run

SaaS
APPLIED AI

2 engineers
SURGEX LABS DELIVERY

In production, multi-tenant
SURGEX LABS DELIVERY

* Client name changed. Engagement details anonymized under NDA; detailed numbers and reference calls available on request.

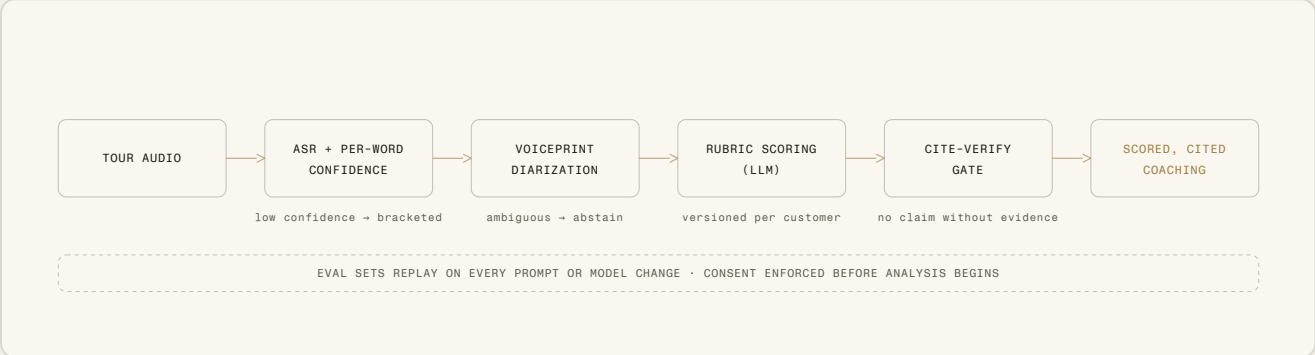
The solution

SurgeX Labs built the pipeline around one rule: no claim without evidence. Speech-to-text runs with per-word confidence — low-confidence spans are bracketed, never asserted. A voiceprint layer matches transcript words to speakers and abstains when attribution is genuinely ambiguous, because a wrong speaker label is worse than none.

Scoring runs against a rubric engine versioned per customer, and every coaching bullet must anchor to a real transcript span — a citation-verification layer rejects anything the model cannot point to. Compliance-sensitive judgments on low-confidence audio are held for human review instead of auto-failing staff. Consent is enforced before any analysis begins.

Every change to prompts or models replays against evaluation sets before it ships. The system went from prototype to multi-tenant production with a two-engineer pod, released in guarded increments behind flags.

THE EVIDENCE-GATED COACHING PIPELINE



HOW IT'S BUILT

- ASR with per-word confidence; low-confidence spans bracketed, never asserted
- Voiceprint-assisted speaker attribution for multi-speaker audio
- Cite-verify layer: coaching bullets must anchor to transcript evidence
- Rubric engine versioned per customer; eval set replayed on every change

The results

0 FABRICATED CITATIONS IN A 371-TOUR AUDIT	**0/36** SPEAKER-ROLE SWAPS IN THE ATTRIBUTION AUDIT
[N] TOURS ANALYZED MONTHLY ACROSS [M] COMMUNITIES [CONFIRM]	**2** ENGINEERS, PROTOTYPE TO MULTI-TENANT PRODUCTION

> [Pull quote pending client approval — e.g. "Coaching only works if the associate believes the quote. Zero fabricated citations is the whole product."]
>
> [NAME], [TITLE], HARBORVIEW SOFTWARE

Tell us the system you need. We'll tell you the week it ships.

hello@surgexlabs.ai
surgexlabs.ai

* Client name changed. Engagement details anonymized under NDA; detailed numbers and reference calls available on request. · This document describes a delivered engagement; bracketed figures are pending final confirmation.