

AI · DIAGNOSTIC DELIVERABLE

AI Readiness Sprint · Final Report

SAMPLE

4-week sprint · use case discovery, prioritization, build-vs-buy

Client	Empresa Anônima S.A.
Engagement ID	#ANV-AIR-2604-1115
Period	Q2 2026 · April 20 – May 18, 2026
Issued by	Anuvia · Mila Vernazza, Principal Consultant
Classification	SAMPLE · Confidential · Not for distribution

This document is a sample preview of an Anuvia diagnostic deliverable. All figures, customer names, and findings are illustrative and anonymized. No real-client data is contained within.

anuvia.com.br · contato@anuvia.com.br

EXECUTIVE SUMMARY

12 use cases inventoried • 4 prioritized • 6 deferred • 2 no-go

Across 17 stakeholder interviews and a review of 8 strategic initiatives, we inventoried **12 candidate AI/GenAI use cases**. Each was scored on five dimensions: business value, data readiness, compliance risk, build complexity, and time-to-value. Four cases were promoted to the 12-month roadmap, six were deferred (mostly pending data foundations); two were explicitly classified "no-go" due to disproportionate risk or unclear ROI.

Prioritization summary

Use case	Verdict	Value (1-5)	Data ready (1-5)	TTV (months)
Sales call summarization pipeline	Build (Q3)	4	5	2
Customer support triage with RAG	Build (Q3-Q4)	5	4	3
Inventory demand forecasting refresh	Buy + tune (Q4)	4	4	3
Document parsing for procurement contracts	Build (Q4-Q1)	4	3	4
Personalized homepage recommendations	Deferred	3	2	—
Sales rep coaching from call transcripts	Deferred	3	3	—
Automated job-description generation	Deferred	2	5	—
Lead scoring (revisit)	Deferred	3	3	—
Marketing copy generation	Deferred	2	4	—
Internal search across wikis	Deferred	3	3	—
AI-driven dynamic pricing	No-go (risk)	—	—	—
AI-driven hiring screening	No-go (compliance)	—	—	—

USE CASE · #1 OF 4

Sales call summarization pipeline

Owner sponsor	VP Sales
Business value	Each rep saves 35 min/day - est. R\$ 410k/yr productivity capture.
Data readiness	5/5 — Gong call recordings already in S3 (12-month archive).
Latency budget	Async · 1h post-call acceptable.
Compliance	Low. Calls are with B2B counterparts under existing contractual NDAs. Speaker consent banner.
Cost per inference	~R\$ 0.18 / call (10k input + 800 output tokens · Claude Sonnet).
Build vs. buy	Build. Off-the-shelf summarizers exist but the structured fields needed (next-step, blocker, deal).
12-month staging	M1 PoV (50 calls) → M2 internal pilot (1 rep team) → M3 broad rollout → M4-M12 quality feed

USE CASE · #2 OF 4

Customer support triage with RAG over knowledge base

Owner sponsor	VP Customer Success
Business value	AHT (avg handle time) reduction est. 18% on Tier-1 tickets · est. R\$ 660k/yr capacity reclaim.
Data readiness	4/5 — Zendesk KB (1,400 articles) needs deduplication + 22% of articles have stale screenshots
Latency budget	Real-time · <2s p95 (agent-assist) · <6s for in-app self-serve
Compliance	Medium. PII present in tickets (account holder names, addresses). Required: PII redaction pre-retrieval
Cost per inference	~R\$ 0.09 / generation. 8k tickets/mo → ~R\$ 8.6k/mo at full coverage.
Build vs. buy	Hybrid. Use managed Bedrock Knowledge Base for retrieval; custom orchestration for triage routing
12-month staging	M1-M2 KB cleanup + dedup → M3 PoV with 1 agent team → M4-M6 agent-assist GA → M7-M8 full coverage

USE CASE · #3 OF 4

Inventory demand forecasting refresh

Owner sponsor	VP Operations
Business value	Inventory carrying cost reduction est. R\$ 520k/yr (target: 6-week-of-supply down to 4-week).
Data readiness	4/5 — SKU-level POS data present (24 months); promo calendar partially structured; supplier I
Latency budget	Batch weekly forecasts · 4h SLA per cycle.
Compliance	Low — no PII.
Cost per inference	Negligible at batch cadence (~R\$ 12 / weekly cycle on SageMaker).
Build vs. buy	Buy + tune. Adopt managed forecasting (SageMaker Canvas or RelEx) for the base model; tun
12-month staging	M1 vendor shortlist → M2-M3 PoV on 200 SKUs → M4-M5 backtest vs. status quo → M6-M9 p

USE CASE · #4 OF 4

Document parsing for procurement contracts

Owner sponsor	Head of Procurement + General Counsel
Business value	Contract review cycle time reduction est. 50% on standard supplier contracts (avg 60 / quarter)
Data readiness	3/5 — 14k PDFs in SharePoint; OCR quality variable; no labelled training set yet (acceptable for
Latency budget	Async · 24h acceptable.
Compliance	Medium-High. Contract content is confidential. Required: in-region (sa-east-1) processing only,
Cost per inference	~R\$ 1.40 / contract (avg 30 pages, structured extraction) · 240/year → R\$ 340/yr (negligible).
Build vs. buy	Build. Extraction schemas are organization-specific (clause taxonomies). Use Claude Sonnet +
12-month staging	M1 extraction schema design w/ legal → M2-M3 PoV on 50 historical contracts (gold-set) → M

CROSS-CUTTING PRINCIPLES

Evals, guardrails, & gates

Eval-first development

- Every use case ships with a gold-set of ≥ 50 examples scored by domain owner before any LLM call.
- Eval suite runs in CI; regression on the gold-set blocks deploy.
- Quarterly eval refresh — gold-set drifts as the world drifts.

Guardrail layers (defense in depth)

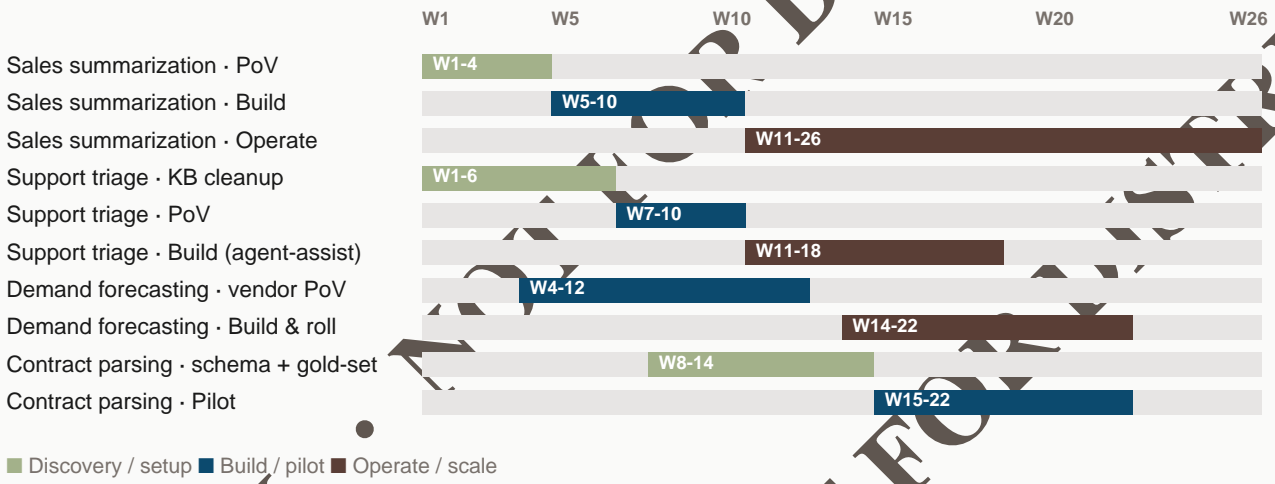
- Input layer: PII detection + redaction · prompt-injection classifier.
- Model layer: system prompt + tool allow-list + structured output enforcement.
- Output layer: schema validation · toxicity / brand-safety check · citation requirement.
- Observability: every generation logged with redacted I/O · sample rate 100% for first 30 days.

Gate criteria · **Discovery** → **PoV** → **Build** → **Operate**

- Discovery → PoV: hypothesis written; gold-set exists; owner identified.
- PoV → Build: eval \geq baseline AND latency budget met AND cost-per-inference within model.
- Build → Operate: shadow run ≥ 2 weeks · zero P0 incidents · runbook complete.
- Operate → Sunset: usage below threshold or eval regression sustained ≥ 4 weeks.

12-MONTH ROADMAP

Discovery → PoV → Build → Operate gates



BUILD vs. BUY SUMMARY

Decision matrix · 4 priority use cases

Use case	Verdict	Rationale	Vendor short-list
Sales summarization	Build	Structured field extraction is org-specific; eval suite is the moat.	
Support triage (RAG)	Hybrid	Buy retrieval (Bedrock KB) + build orchestration (LangChain)	Bedrock, KB, Pinecone (alt)
Demand forecasting	Buy + tune	Well-trodden problem; mature managed ops	Tableau, SageMaker Canvas, RelEx, ToolsGroup
Contract parsing	Build	Clause taxonomy is org-specific; reviewer-in-loop is mandatory.	

Cross-cutting platform decisions

- Model provider: Anthropic Claude (Sonnet/Haiku tiers) — best fit for structured-output + safety profile.
- Hosting: AWS Bedrock for in-region (sa-east-1) compliance with LGPD.
- Vector store: Bedrock Knowledge Base for managed; OpenSearch (managed) as escape hatch.
- Eval tooling: open-source *promptfoo* + custom gold-set runner integrated in CI.
- Observability: Langfuse self-hosted for trace inspection + cost attribution.

APPENDIX · METHODOLOGY

Methodology & references

Use cases were sourced from (1) 17 stakeholder interviews across 6 functions, (2) review of the 2026 strategic plan, (3) inventory of current vendor evaluations, (4) review of in-flight engineering experiments. Each use case was scored on a 1-5 scale across 5 dimensions; scoring rubrics are documented in the engagement workspace.

Scoring rubric (abridged)

Dimension	1 (low)	5 (high)
Business value	Vague qualitative benefit	Quantified P&L impact, owner-validated
Data readiness	Data does not exist	Clean, labelled, accessible, governed
Compliance risk	Severe (PII / regulated)	No regulated data touched
Build complexity	Novel research problem	Off-the-shelf with prompt tuning
Time-to-value	>12 months	<3 months

— End of report — Anuvia · contato@anuvia.com.br · anuvia.com.br