

TECHNICAL WHITEPAPER · 2026

Sovereign Banking AI

BankingCore — the AI Operating System the bank owns and runs in-house

Silicon → models → core systems → agents → outcomes — one sovereign platform for banks across India & EMEA.

PUBLISHED BY

Bud BankingCore × LTM

The Banking AI Partnership · India & EMEA

Version 1.0

February 2026

CONFIDENTIAL

01 · THE IMPERATIVE

Every Bank Needs AI — But Can't Hand Staff a Public Chatbot

Every bank now needs AI — yet no bank can pour customer, payment and KYC data into a public chatbot. RBI and the DPDP Act in India, GDPR and the EU AI Act across EMEA all keep that data **inside the bank's walls**. So the data stays in and the AI stays out — while roughly **70% of staff already paste bank information into unapproved tools**. The problem is not the technology; it is the **architecture**. **BankingCore** closes the gap: one sovereign AI platform the bank **owns and runs in-house** — Bud AI OS the engine, LTM the partner that industrialises it for banking.

72 → 88%

Lower AI cost vs cloud-only

**₹1.4–2.9K
Cr**

3-yr net value (9–32x)

Weeks

To first staff assistant

100%

In-region (India / EMEA)

Examiner

-ready by design

WHY THIS IS HARD FOR A BANK — SIX REAL BARRIERS

Can't just hand staff ChatGPT

RBI/DPDP in India and GDPR across EMEA keep customer, payment and KYC data inside the bank.

~70% of staff already use unapproved AI

Knowledge trapped in silos

Staff can't quickly find the right circular, policy or customer history when they need it.

Hours lost per employee, every day

Core systems aren't AI-ready

300–500 systems with no safe, simple way for AI to read or act on them.

6–12 months to connect each, by hand

The build-vs-buy trap

Building is slow and needs scarce talent; buying means lock-in and patchy controls.

2–3 years per bank, controls rebuilt each time

AI costs spiral at scale

Sending every request to premium cloud AI is unaffordable across a whole bank.

₹19 Cr/yr for one fleet of assistants

Hard to satisfy the examiner

Scattered vendor logs can't reconstruct how an AI reached a decision, in minutes.

DPDP ₹250 Cr · GDPR 4% turnover · EU AI Act €35M

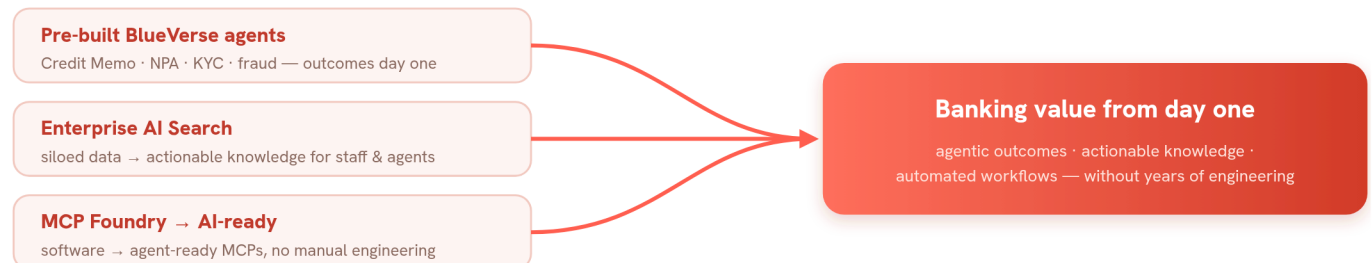
THE STRATEGIC CHOICE

The choice isn't which AI feature to buy. It's whether AI **fragments and stalls** across 40 disconnected tools — or **compounds on one platform the bank owns end to end**: silicon to use case, data never leaving the bank, every action examiner-ready. LTM × Bud BankingCore is that platform.

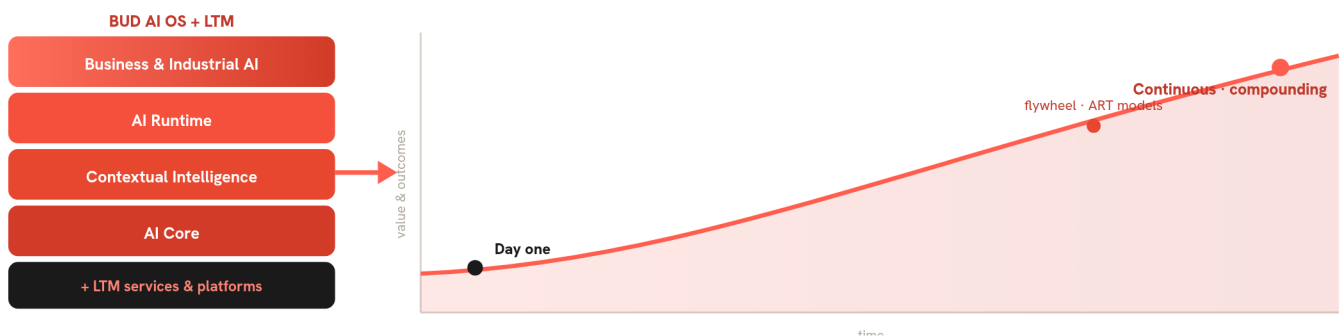
02 · WHAT THE BANK GAINS

Six Barriers, Six Outcomes

Each barrier has a direct answer — value on day one, safe automation over time, all on **one platform the bank owns**. Three pre-built capabilities deliver banking value immediately; the same owned platform then compounds that value continuously.



...AND CONTINUOUSLY — VALUE COMPOUNDS THROUGH BUD AI OS + LTM



Value keeps compounding on the same owned platform: a **self-optimising flywheel** (cheaper per call), **ART-retrained private models**, **bi-quarterly Bud × LTM upgrades**, and expanding agents & use cases — outcomes the bank never re-buys.

SIX BARRIERS, ANSWERED

1 · A private assistant, day one

Enterprise AI Search turns siloeed data into actionable knowledge for every employee and agent, in 22 languages, live in weeks. Pre-built BlueVerse BFSI agents deliver outcomes immediately.

✓ Solves: shadow AI & knowledge silos

2 · Core systems AI-ready in days

MCP Foundry converts existing banking software into agent-ready MCPs without manual engineering — agents automate across core, lending, payments & CRM in days, never touching a system of record directly.

✓ Solves: core systems aren't AI-ready

3 · One platform, not 40 tools

Every model, chip & use case on one platform the bank owns — no sprawl, no gaps the examiner will find. Build once, reuse everywhere; value compounds instead of fragmenting.

✓ Solves: the build-vs-buy trap

4 · Cheaper the more it's used

A self-optimising flywheel: every interaction lifts accuracy and cuts cost-per-call. Most work runs on the bank's own low-cost models; premium cloud only for the hardest — 72 → 88% lower cost.

✓ Solves: AI costs spiral at scale

5 · Hybrid AI, no lock-in

Bring frontier closed-source AI to work alongside the bank's private models — reducing dependence with every use. Runs on any chip from any country: no hardware, model or cloud lock-in.

✓ Solves: vendor, cloud & chip lock-in

6 · Compliant & examiner-ready

A transparent, owned AI supply chain compliant across India & EMEA — DPDP, RBI FREE-AI, GDPR, EU AI Act, DORA — with data private and every AI decision auditable in minutes.

✓ Solves: hard to satisfy the examiner

02 · THE ENGINE

Bud AI OS for Banking — The Six-Layer Engine

Bud AI OS is the software engine of BankingCore — one platform built in **six layers**, each owned and run by the bank. Together they turn raw silicon into useful banking work, safely and sovereignly. Read top-down it is how staff get value; read bottom-up it is how the bank stays sovereign and cheap.

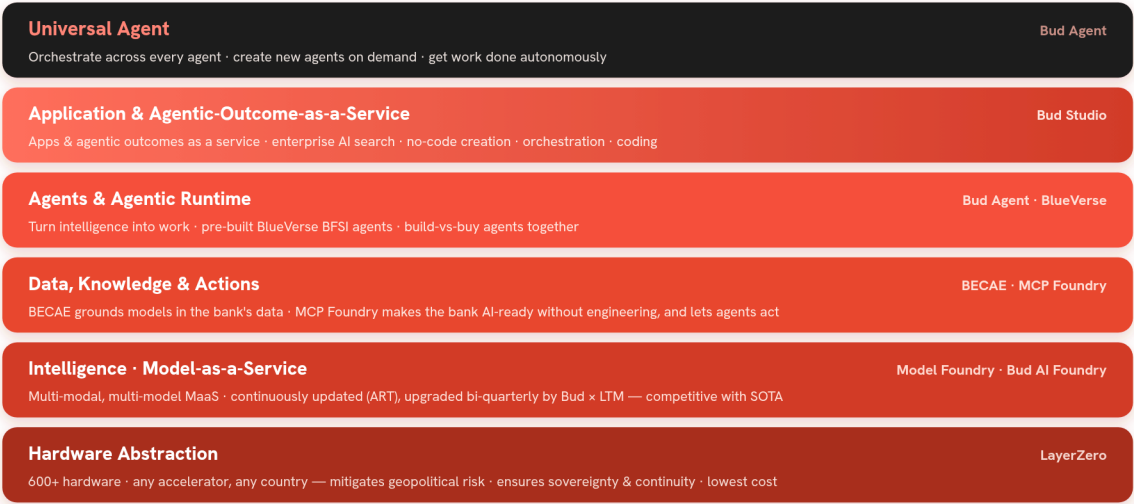
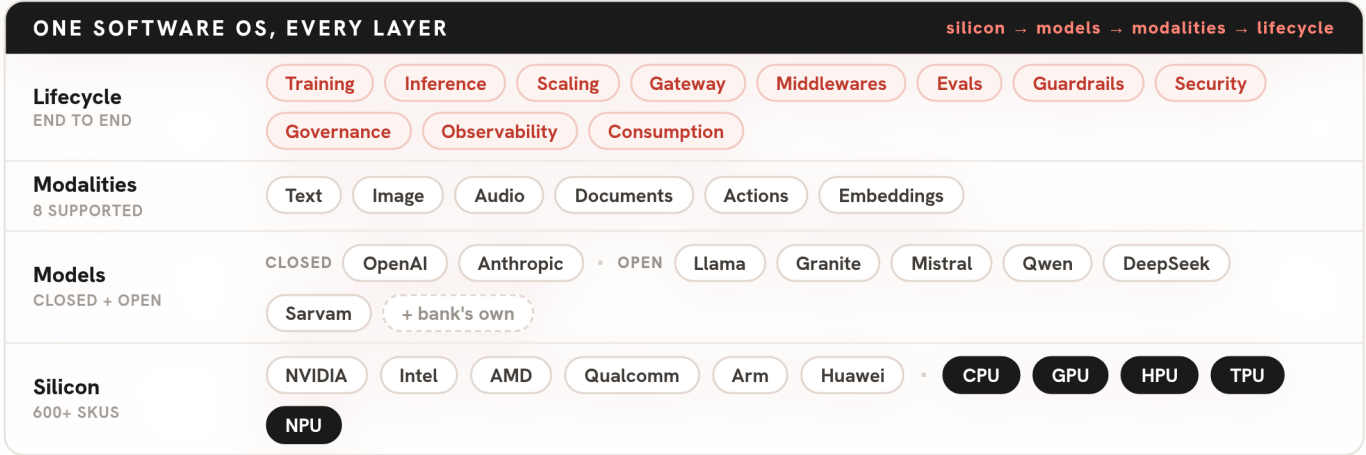


Figure 1 — Bud AI OS: six owned layers from silicon to a universal agent. Each a distinct capability; together one governed engine, owned by the bank.

Layer · component	What it delivers to the bank
Universal Agent · Bud Agent	A single agent every user can ask — it orchestrates the bank's other agents, creates new ones, and completes multi-step work autonomously within policy.
Application aaS · Bud Studio	The front door: apps & agentic outcomes as a service — enterprise AI search, no-code agent creation, orchestration and coding — one safe place for every employee.
Agents & Runtime · BlueVerse	Agents and the agentic runtime convert intelligence into useful work: pre-built BlueVerse BFSI agents (Credit Memo, NPA EWS, KYC, fraud) ready to run.
Data & Actions · BECAE + MCP Foundry	BECAE grounds models in real data so answers are accurate; MCP Foundry turns core systems into safe, typed tools — AI-readiness without engineering, and the ability to act.
Intelligence · Model Foundry	Multi-modal, multi-model Model-as-a-Service inside the bank — Llama, Granite, Mistral, Qwen, DeepSeek, Sarvam or the bank's own — kept current by ART and bi-quarterly upgrades.
Hardware · LayerZero	Runs on 600+ hardware types, any accelerator from any country — mitigating geopolitical risk, guaranteeing sovereignty and continuity, and driving the lowest TCO.

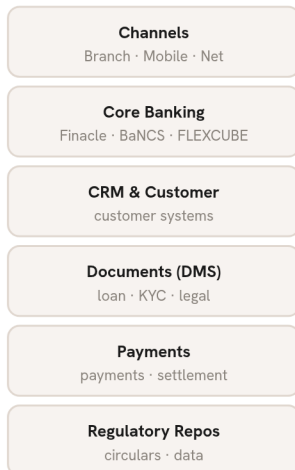
ONE PLATFORM · UNIVERSAL COVERAGE — WHAT BUD SPANS



03 · THE ARCHITECTURE

One Governed Platform Between AI and Every Bank System

Wire AI by hand and every agent touches every system: **N agents × M systems** of ungoverned, duplicated connections — no kill switch, no unified trace, no single point to enforce RBI/DPDP or GDPR/DORA. The fix is one governed platform between AI and the bank's systems. Just as the Core Banking System runs the bank's transactions, **BankingCore runs the bank's intelligence**.

BANK APPLICATION ESTATE

ONE GOVERNED CHANNEL

BANKINGCORE · making core banking intelligent**Experience & Agent Orchestration**

workspaces · bounded agents · human-in-the-loop gates · 22-language UI
Bud Studio · Bud Agent

Tool & Integration

single point of: policy · access · audit · data-class · routing
MCP Foundry

Governance & Security Control Plane

model gateway · T0-T4 routing · guardrails · zero-trust · examiner-ready audit
Bud AI Foundry · Bud Sentry

Knowledge, Models & Sovereign Compute

governed corpora · private & open models · in-region GPU · CPU-native
BECAE · Model Foundry · LayerZero

Figure 2 — Instead of $N \times M$ ungoverned connections, every agent reaches every system through one policy-enforced, audited, access-controlled layer — security, regulatory, audit & access solved once.

WHAT BANKINGCORE MUST BE**Secure**

Auditable end to end; sensitive data never leaks — zero-trust, PII-tokenized.

Sovereign

In-region and owned by the bank — across India & EMEA, never a foreign black box.

Governed

Every model, prompt & agent evaluated and controlled before and during use.

Scalable

One backend serves every use case — capacity and controls amortized, not rebuilt.

Open — no lock-in

Any hardware, any model, any agent; connect yours, a partner's, or any third party's.

Cost-effective

TCO falls as the platform learns and most work runs on the bank's own low-cost models.

04 · THE LIFECYCLE

The Agentic Development Life Cycle — Research to Scale

BankingCore is not a tool bolted on — it is the platform on which the bank's **entire agentic AI lifecycle** runs, from first research to production scale and daily consumption. Every stage lives on the same governed backend, so nothing is rebuilt, nothing escapes governance, and the bank learns once and reuses everywhere.

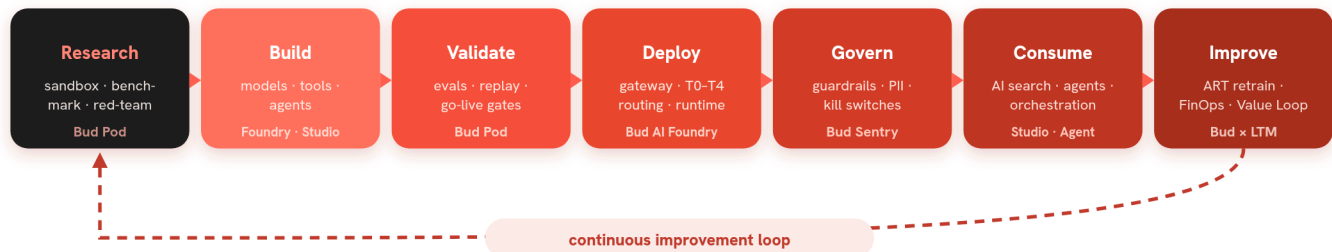


Figure 3 — The agentic development life cycle, end to end on one platform — with a continuous-improvement loop that keeps the bank's intelligence current.

Stage	On the platform	What happens — and why it is safe
Research	Bud Pod	Teams experiment, benchmark candidate models, and run red-team & replay tests in a sandbox — nothing touches production or real customer data.
Build	Model · MCP Foundry · Studio	Private domain models & adapters are trained, core systems are toolized, and agents are assembled no-code — all on shared, governed primitives.
Validate	Bud Pod	Evidence-based go-live gates: benchmark, adversarial and replay evidence required before promotion. "It worked in the demo" is not evidence.
Deploy	Bud AI Foundry	The control plane routes each request by data class (T0-T4) to the right model and infrastructure, cost-controlled and observable.
Govern	Bud Sentry	Guardrails, PII tokenization, runtime monitoring and kill switches apply to every request — deny, redact, or route to a human.
Consume & scale	Studio · Universal Agent	Staff use agents through one workspace; the universal agent orchestrates them and completes work — adoption scales without new silos.
Improve	ART · FinOps · Value Loop	Every interaction feeds the value loop; ART continuously retrains private models, and Bud x LTM upgrade the models bi-quarterly.

Intelligence that stays state-of-the-art — without the bank lifting a finger

Because the bank owns the platform, its AI keeps improving on its own: usage feeds the Value Loop, **ART continuously retrains** the bank's private models on fresh, governed data, and **Bud x LTM upgrade the underlying models bi-quarterly** — so banking intelligence stays competitive with state-of-the-art general models, while every byte stays in-region.

05 · THE PARTNERSHIP

LTM x Bud — The Banking AI Partnership

Bud BankingCore is the engine; LTM industrializes it for the bank. Bud provides the banking AI operating system in software; LTM brings advisory & readiness, in-region sovereign infrastructure, banking AI operations, and **BlueVerse BFSI Industry AI** — delivered foundation → scale → value, proven at production scale (3× faster pilot-to-enterprise · 30–40% lower infrastructure cost). **BankingCore is the engine; LTM makes it run.**

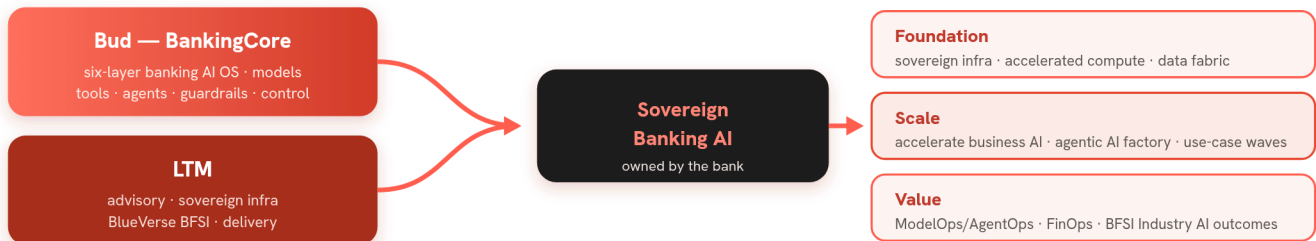


Figure 4 — BankingCore (engine) and LTM (advisory, sovereign infrastructure, BlueVerse BFSI Industry AI) combine into one sovereign banking AI capability the bank owns.

HOW THE PARTNERSHIP ALIGNS FOR BANKING

LTM pillar	What LTM brings to the bank	Pairs with Bud
Advisory — Assess & Design	Foundation strategy: AI workload profiling, readiness & strategy, architecture & roadmap, value governance & ROI	Bud Pod validation & go-live gates
CIS AI Core	In-region sovereign infrastructure (India & EMEA), accelerated compute (HPC+GPU, InfiniBand/RoCE, 30–40% cost), data-plane modernization (Unified DASE)	LayerZero · sovereign compute
Scale AI	AI-optimized infra, AI PaaS, model serving & GPU orchestration, Agentic AI Factory (FoundryAI · Agent Core · multi-agent orchestration)	Bud AI Foundry · Model Foundry
Realize AI	ModelOps/AgentOps, observability & SRE, AI FinOps, governance frameworks (Right Action™ · TRACER · ASPIRE2ADAPT)	Foundry observability & FinOps
Sovereign AI Platform	Localized & confidential pipelines, national infra & compliance, edge AI & federated learning (federated fraud)	Bud AI Foundry · Sentry · LayerZero
BlueVerse · Industry AI	BFSI industry AI — domain playbooks, multimodal apps, digital twins for banking operations & risk	Banking use-case catalogue

THE ECONOMICS — CHEAPER THE MORE THE BANK USES IT

~9.2x less control-rebuild effort across a 12-bank group — ₹321 Cr saved before any AI value	–72 → 88% lower AI cost than premium-cloud-only — and it keeps falling as the platform learns	~3x / ~6x about 3× the performance at ~6× lower cost; 82% lower hardware cost, same accuracy	₹16 Cr /yr saved on a single fleet of assistants — most work on the bank's own low-cost models
--	---	--	--

06 · THE BANKING AI FLYWHEEL

Usage Is an Asset, Not a Cost

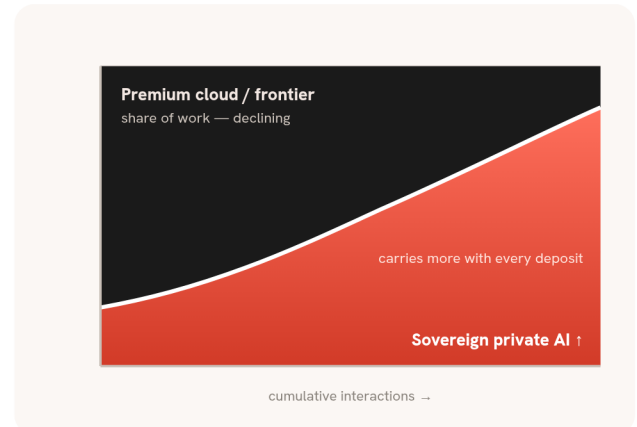
Traditional banking software **depreciates** with use. Agentic AI, architected correctly, is the first software category where **usage is an asset**. Every interaction is a **deposit** distilled into durable capital that makes the next interaction cheaper, faster and smarter. In a hybrid deployment the premium cloud / frontier model is only a **teacher** — each deposit distills its signal into the bank's private models, so its share of the work **bends down with every interaction**. Intelligence is repatriated inside the bank's boundary.

THE MECHANISM · ONE LOOP, EVERY INTERACTION



01 Serve the query. **02 Capture** the trace, label the outcome. **03 Extract** the reusable essence. **04 Deposit** into one of four capital pools — the next call draws a warmer path.

THE BANKING PAYOFF · DEPENDENCE ↓



The teacher signal compresses into a private student that serves the same banking query at a fraction of the cost — every interaction shifts load **from premium cloud to the bank's sovereign private AI**, inside the bank's own boundary.

WHAT ACCUMULATES · FOUR CAPITAL POOLS

Model ↓ \$/token

Distilled weights, preference-tuned policies, routing models — the teacher compresses into a private student serving the same query far cheaper.

Infra ↓ ms p95

KV-caches, prefix caches, disaggregated prefill/decode, locality-aware routing — the substrate learns the bank's traffic shape.

Policy ↑ success

Prompts, routes, memory, tool-use trajectories, guardrails — every resolution and incident hardens the system forward.

Primitive ↓ time-to-agent

Skills, MCP tools, A2A capabilities — every agent leaves reusable pieces; the next use case starts at 70%, not 0%.

A cascade of smaller models

Routing × distillation sends most calls to the smallest capable model; premium cloud handles only the hardest fraction.



↓ infra footprint

↑ performance

WRIGHT'S LAW, FINALLY APPLIED TO SOFTWARE

In 1936 Theodore Wright found manufacturing cost falls ~15% with every doubling of cumulative output — solar, semiconductors and batteries all followed. Software never did. Enterprise agentic AI is the **first category to deliver Wright-Law returns** — $\text{cost}(N) \approx c \cdot N^{\log_2(1-r)}$, $r \approx 15\text{--}30\%$ — and in a sovereign banking deployment, each doubling also cuts the **premium-cloud / frontier model's share of the work**.

07 · COMPLIANT BY DESIGN

Built for Banking Rules — Governed at Runtime

Compliance is a property of the architecture, not a separate workstream. Every query is classified at ingestion, and the **model gateway** — the most important control in the stack — routes it by data class: which model, which infrastructure, which controls. Sensitive data stays in-region, and every decision is examiner- and supervisor-ready.

DATA-CLASS ROUTING · the model gateway decides per request



Figure 5 — T0-T4 data-class routing: the gateway sends each request to the right model and infrastructure for its sensitivity — the most sensitive classes are architecturally barred from public AI.

BUILT FOR INDIA'S & EMEA'S RULES

- ▶ **India** — RBI data-localisation & FREE-AI (26 recommendations, 7 sutras), DPDP Act, CERT-In, MANAV
- ▶ **EMEA** — EU AI Act (risk tiers, high-risk obligations), GDPR (residency, minimisation, erasure), DORA, PSD2
- ▶ PII tokenized before any tool call; sensitive classes pinned to in-region private processing
- ▶ Data stays in-region · every AI decision auditable · examiner- & supervisor-ready

THE NON-NEGOTIABLE RED LINES

- ▶ No raw customer / KYC / payment data to public AI — enforced architecturally, T3/T4 routing
- ▶ No agent moves money or finalises credit decisions without explicit human authority
- ▶ Nothing ships on a demo — benchmark, adversarial & replay testing required (Bud Pod gates)
- ▶ Read-only tools before write tools; one control plane, no shadow point solutions

GOVERNED AT RUNTIME · MEASURED, NOT ASSERTED

Multi-tenant isolation

Multiple banks on shared infra; per-bank keys, isolated namespaces, network segmentation, quarterly red-team.

Runtime enforcement

Deny, redact, or route to a human; 8 policy packs, 15 rules across the agent lifecycle; kill switches.

Examiner-ready audit

Every decision hash-chained and mapped to RBI FREE-AI, DPDP & EU AI Act; reconstruct any decision in minutes.

Measured outcomes

Seven dimensions: 60%+ weekly use, 80%+ grounded, zero PII leaks, -50% false alerts, falling TCO, clean exams.

08 · OUTCOMES & THE PATH

Value at Every Wave — From Assistant to Autonomy

The sequencing principle is **value first, then tool depth, then autonomy**. Early wins come from high-frequency knowledge and document work — measurable gains without control over system state. Transactional agents enter only after the platform proves safe in read-only and propose-before-act modes.

Wave	Representative use cases	Benchmark signal
Wave 0 · Foundational	Knowledge Assistant · Circular Intelligence · Branch Copilot (22 languages)	~2 hrs/week saved per employee · circulars 3 days → 4 hrs · query 8 min → 90 sec
Wave 1 · Document	Loan & KYC Document Intelligence · Complaint & Grievance Triage	3 days → 2 hrs · error ~8% → <1% · KYC accuracy >98% · first response 15 → 3 days
Wave 2 · Advisory	Credit Memo Drafting · NPA Early-Warning · Fraud Triage · Collections	memo 2 days → 45 min · 6-12 mo advance notice · false positives ~50%+
Wave 3 · Native	Recovery / Auction & SCF · Federated Fraud (opt-in, anonymised)	SCF approval 3 days → 2 hrs · catch cross-bank fraud earlier · full sovereignty

ILLUSTRATIVE 3-YEAR VALUE ENVELOPE

₹180-360 Year 1 (Cr)	₹500-960 Year 2 (Cr)	₹910-1,700 Year 3 (Cr)	₹1.4-2.9K Net 3-yr (Cr)
--------------------------------	--------------------------------	----------------------------------	-----------------------------------

Representative mid-size bank (India ₹ shown); EMEA banks compute local-currency equivalents.

WHAT WE ASK THE BOARD TO APPROVE

- ▶ **Own the platform** — don't rent AI use case by use case
- ▶ **Start with a private assistant** for staff, live in weeks
- ▶ Then a **first wave** of high-value, low-risk use cases
- ▶ Keep money movement & final decisions with people
- ▶ One governance model & an examiner evidence-pack

VALUE AT EVERY PHASE — WHILE 95% OF BANK AI PILOTS NEVER REACH PRODUCTION

P0 · Days 0-30	P1 · Days 31-60	P2 · Days 61-90	P3 · Mo 4-6	P4 · Mo 7-18	P5 · Mo 19-36
Mobilise — charter, T0-T4 taxonomy, steering committee	Land platform — sovereign deploy, SSO/security, registry	First AI surface — Knowledge + Circular live, audit	Toolize — read-only Document, Complaint, Fraud	Propose-before-act — Credit Memo, NPA EWS	Autonomous & scale — bounded agents, rollout

THE BOTTOM LINE

Banks can't wait — leaders run AI at scale while most banks across India and EMEA can't even give staff ChatGPT under RBI/DPDP or GDPR. The answer is not another tool but **one sovereign platform the bank owns end to end**: a private assistant from day one, working agents over time, compliant by design, cheaper the more it's used. **Bud BankingCore is the engine; LTM makes it run.**

Bud BankingCore × LTM — Sovereign Banking AI

Owned by the bank · In-region (India & EMEA) · Examiner- & supervisor-ready · Open | Bud. Simplifying Intelligence.