

CONFIDENTIAL



AI systems and cost exposure overview

Internal operating snapshot – Q1 2026

Prepared for: Board and executive committee

Prepared by: Technology

1. AI services in production

Active AI endpoints: 28

Experimental endpoints: 11

There is no central register that records for each service:

- owner
- last review date
- linked business feature
- operating cost

Tracking is distributed across teams.

2. Lifecycle and review

Deployment process:

- Team-level approval
- No standardised validation step
- No mandatory review after launch

No formal retirement process exists.

Some endpoints have not been reviewed in over 12 months.

3. Infrastructure distribution

Owned GPU capacity: 68 per cent of steady workloads

Cloud burst capacity: 32 per cent

Burst usage increases significantly during peak promotions.

No defined trigger forces review when cloud share exceeds threshold.

4. Cost profile

Annual AI-related infrastructure run rate: \$24.7m

Cloud AI spend increased 34 per cent quarter-on-quarter.

Cost is visible by environment.

It is not consistently visible by AI feature.

5. Governance controls

- Change control exists for core systems
- No dedicated AI lifecycle policy
- No centralised AI inventory
- No peak-specific AI guardrails
- No defined endpoint review cadence

Stratum Retail's AI capability is expanding.

Governance structure has not expanded at the same pace.