AEROSENSEI — FIELD NOTES FROM THE ENTERPRISE

Token Economics

Pre-Flight Checklist

Before you hit Run on anything larger than 1,000 records.

- [ ] 01 Cost per call estimated and written down.
- [ ] 02 Total cost estimated and written down. Acceptable to your boss without a memo.
- [ ] 03 Model is the cheapest tier that does the job.
- [ ] 04 Prompt prefix is using cache_control or equivalent.
- [ ] 05 Job is using the Batch API if time allows.
- [ ] 06 Piloted on 10 records, then 100. Output quality verified by a human.
- [ ] 07 Service account in use. Personal key is not in this codebase.
- [ ] 08 Hard spend limit set on the account.
- [ ] 09 Per-request cost logged in application logs.
- [ ] 10 Daily spend alert configured.

QUICK REFERENCE — MODEL TIERS

CHEAP Haiku / GPT-4o-mini / Gemini Flash — Classification, extraction, simple summaries

MID Sonnet / GPT-4o — Reasoning, code gen, nuanced summaries (default)

PREMIUM Opus / GPT-4.1 — Multi-step reasoning, agentic flows, showcases