

Tokenmaxing Security Note

ExecuteAI ApS · CVR 45798798 · 14 September 2026

This note describes the gateway's implemented controls. It is not a certification or an availability SLA.

1. Account and API keys

New accounts require email-code verification before receiving a session or API key. API keys are emailed once and stored as SHA-256 hashes. Lost keys cannot be recovered: sign in to Account and choose Replace API key to revoke the old key and email a new one.

1a. Customer provider keys

Optional Anthropic, OpenAI and Google keys that you paste in Account are encrypted at rest with AES-256-GCM. Plaintext exists only in memory for the duration of one provider call. The portal never redisplay a stored key. Claude is served only with your Anthropic key; we do not run Claude on a shared ExecuteAI account.

2. Sign-in security

Six-digit codes expire after ten minutes and allow at most five incorrect attempts. Code requests, login and key rotation are rate limited in PostgreSQL across replicas. Portal cookies are Secure, httpOnly and SameSite=Lax; only their hashes are stored.

3. Prepaid usage

Calls require an API key and sufficient prepaid balance. A durable reservation is created before inference. Settlement, fee and usage audit are committed together. Send an Idempotency-Key for each logical call; reusing it never repeats the charge or the provider request. Duplicate calls return status 409 with the recorded state, not a cached model answer.

4. Crash recovery

Unfinished reservations expire after fifteen minutes and are reconciled automatically. Unknown provider outcomes are recorded and the customer's hold is released. These exceptional cases can incur an operator cost and are logged for investigation.

5. Pricing and records

Retail and routed-model rates use the same Artificial Analysis snapshot, stored with each call. The baseline is a comparison using the routed response's measured token counts at the originally selected model's rates. The fee is 30% of the positive difference. Net savings exclude the fee. This is not an independent measurement of the original model's output quality or token count.

6. Data and providers

PostgreSQL stores account details, payment references, token counts, rates and charges. Application tables do not store prompts or completions. Requests are forwarded to the configured model provider. OpenAI uses Azure EU deployments; other models may use global processing. See the privacy policy and model configuration before sending sensitive data.

7. Payment verification

Stripe card details do not reach the gateway. Signed webhooks verify settled USD checkout payments. Duplicate deliveries cannot credit the same session twice. Refunds reverse

8. API scope

The gateway supports text chat completions through an OpenAI-style endpoint. Responses API, embeddings, vision and tool execution are not implemented. Streaming is buffered until usage can be settled. Cache usage comes from the provider; semantic caching is not advertised.

9. Contact

ExecuteAI ApS, CVR 45798798. Contact info@tokenmaxing.online for billing reconciliation, data requests or security reports.

credited balances and disputes suspend further usage until reviewed.