



## Configuration

# Model Failover

OpenClaw handles failures in two stages:

1. **Auth profile rotation** within the current provider.
2. **Model fallback** to the next model in `agents.defaults.model.fallbacks`.

This doc explains the runtime rules and the data that backs them.

## Auth storage (keys + OAuth)

OpenClaw uses **auth profiles** for both API keys and OAuth tokens.

Secrets live in `~/.openclaw/agents/<agentId>/agent/auth-profiles.json`  
(legacy: `~/.openclaw/agent/auth-profiles.json`).

Config `auth.profiles` / `auth.order` are **metadata + routing only** (no secrets).

Legacy import-only OAuth file: `~/.openclaw/credentials/oauth.json`  
(imported into `auth-profiles.json` on first use).

More detail: [/concepts/oauth](#)

Credential types:

```
type: "api_key" → { provider, key }

type: "oauth" → { provider, access, refresh, expires, email? } (+  
projectId / enterpriseUrl for some providers)
```

## Profile IDs



OAuth logins create distinct profiles so multiple accounts can coexist.

&gt;

Default: `provider:default` when no email is available.

OAuth with email: `provider:<email>` (for example `google-antigravity:user@gmail.com` ).

Profiles live in `~/.openclaw/agents/<agentId>/agent/auth-profiles.json` under `profiles`.

## Rotation order

When a provider has multiple profiles, OpenClaw chooses an order like this:

1. **Explicit config:** `auth.order[provider]` (if set).
2. **Configured profiles:** `auth.profiles` filtered by provider.
3. **Stored profiles:** entries in `auth-profiles.json` for the provider.

If no explicit order is configured, OpenClaw uses a round-robin order:

**Primary key:** profile type (**OAuth before API keys**).

**Secondary key:** `usageStats.lastUsed` (oldest first, within each type).

**Cooldown/disabled profiles** are moved to the end, ordered by soonest expiry.

## Session stickiness (cache-friendly)

OpenClaw **pins the chosen auth profile per session** to keep provider caches warm. It does **not** rotate on every request. The pinned profile is reused until:

the session is reset ( `/new` / `/reset` )

a compaction completes (compaction count increments)

 the profile is in cooldown/disabled

---

Manual selection via `/model ...@<profileId>` sets a **user override** for that session and is not auto-rotated until a new session starts.

Auto-pinned profiles (selected by the session router) are treated as a **preference**: they are tried first, but OpenClaw may rotate to another profile on rate limits/timeouts. User-pinned profiles stay locked to that profile; if it fails and model fallbacks are configured, OpenClaw moves to the next model instead of switching profiles.

## Why OAuth can “look lost”

If you have both an OAuth profile and an API key profile for the same provider, round-robin can switch between them across messages unless pinned. To force a single profile:

Pin with `auth.order[provider] = ["provider:profileId"]` , or

Use a per-session override via `/model ...` with a profile override (when supported by your UI/chat surface).

## Cooldowns

When a profile fails due to auth/rate-limit errors (or a timeout that looks like rate limiting), OpenClaw marks it in cooldown and moves to the next profile. Format/invalid-request errors (for example Cloud Code Assist tool call ID validation failures) are treated as failover-worthy and use the same cooldowns.

Cooldowns use exponential backoff:

1 minute

5 minutes

 25 minutes  
1 hour (cap)

State is stored in auth-profiles.json under usageStats :

```
{
  "usageStats": {
    "provider:profile": {
      "lastUsed": 1736160000000,
      "cooldownUntil": 1736160600000,
      "errorCount": 2
    }
  }
}
```

## Billing disables

Billing/credit failures (for example “insufficient credits” / “credit balance too low”) are treated as failover-worthy, but they’re usually not transient. Instead of a short cooldown, OpenClaw marks the profile as **disabled** (with a longer backoff) and rotates to the next profile/provider.

State is stored in auth-profiles.json :

```
{
  "usageStats": {
    "provider:profile": {
      "disabledUntil": 1736178000000,
      "disabledReason": "billing"
    }
  }
}
```

Defaults:



Billing backoff starts at **5 hours**, doubles per billing failure, and caps at **24 hours**.

Backoff counters reset if the profile hasn't failed for **24 hours** (configurable).

## Model fallback

If all profiles for a provider fail, OpenClaw moves to the next model in `agents.defaults.model.fallbacks`. This applies to auth failures, rate limits, and timeouts that exhausted profile rotation (other errors do not advance fallback).

When a run starts with a model override (hooks or CLI), fallbacks still end at `agents.defaults.model.primary` after trying any configured fallbacks.

## Related config

See [Gateway configuration](#) for:

```
auth.profiles / auth.order  
auth.cooldowns.billingBackoffHours /  
auth.cooldowns.billingBackoffHoursByProvider  
auth.cooldowns.billingMaxHours / auth.cooldowns.failureWindowHours  
agents.defaults.model.primary / agents.defaults.model.fallbacks  
agents.defaults.imageModel routing
```

See [Models](#) for the broader model selection and fallback overview.



Powered by **mintlify**

>

---