OPENAI ARCHITECTURE PRIMER

Codex Model Catalogs

A versioned capability contract

A visual guide to source precedence, authenticated discovery, client-version matching, provider overrides, cache identity, and inference authority.

FIRST-PARTY DOCUMENTATION / INDEPENDENT LUMINITY SYNTHESIS



The Catalog Is a Versioned Capability Contract

One response controls model discovery and client behavior for one authenticated caller and one client version.

SIX PROVIDER DECISIONS

01
Source
local or remote

02
Authenticate
scope credentials

03
Version
match the client

04
Assemble
complete response

Cache 05
Partition freshness by provider, authenticated identity, client version, and validation state.

Authorize 06
Keep catalog visibility separate from route truth and inference access.

The six-decision model is Luminity synthesis. OpenAI documents the source, response, cache, and inference-access mechanics.

IN THIS PRIMER

Key Points

01 Resolve the source

Local configuration takes precedence. Remote discovery follows the selected authentication and provider path.

02 Assemble compatible metadata

Match the client version, constrain the route inventory, and return one complete ModelsResponse.

03 Preserve the inference boundary

Catalog visibility informs the client. The deployed provider still authorizes every inference request.

© 2026 Luminity Digital, Inc. | Independent synthesis of official OpenAI documentation.

Source: https://learn.chatgpt.com/docs/enterprise/roll-out-a-gateway#model-catalog-reference · verified Nov 5, 2026

OpenAI Select Partner

Resolve Source and Response Contracts

Local precedence and authenticated discovery feed one client-versioned ModelsResponse.

01

Source — Local or Remote

local file · precedence | provider path | explicit URL

1 **Config** — catalog source

2 **Auth** — credential path

3 **Fetch** — remote contract

4 **Response** — complete payload

Local Precedence — SOURCE

A configured local catalog becomes the startup snapshot and skips remote discovery.

Credential Scope — AUTH

Remote discovery sends the selected provider credentials and configured headers.

02

Response — Compatibility Contract

authenticated caller | client version | provider identity

1 **Match** — client version

2 **Defaults** — upstream metadata

3 **Constrain** — deployed routes

4 **Return** — ModelsResponse

Unique Slugs — ROUTE

Each advertised slug remains a stable route key in the complete response.

Client Behavior — BEHAVIOR

Metadata sets compatible defaults for reasoning, context, tools, and service tiers.

03

The Compatibility Record

Source Identity — SOURCE

Record local or remote origin, upstream baseline, and validation time.

Client Version — VERSION

Retain the client build used to select the compatible response.

Provider Identity — CALLER

Retain the provider and authenticated identity that scoped discovery.

Validation State — EVIDENCE

Retain route and capability tests before the catalog reaches clients.

© 2026 Luminity Digital, Inc. | Independent synthesis of official OpenAI documentation.

Source: https://learn.chatgpt.com/docs/enterprise/roll-out-a-gateway#model-catalog-reference · verified Nov 5, 2026

OpenAI Select Partner

Own Cache, Route, and Authority State

Freshness, route truth, capability evidence, and inference access remain separate provider records.

01

Cache Identity

- ✓ provider identity — SOURCE
- ✓ authenticated identity — CALLER
- ✓ client version — COMPATIBILITY
- ✓ upstream ETag — FRESHNESS
- ✓ validation timestamp — EVIDENCE

Provider | Caller | Client Build

02

Route and Capability State

picker inventory | deployed route | inference authority



ROUTE TRUTH

Every advertised slug needs an owned and deployed provider route

INFERENCE

Catalog visibility does not grant account, region, or model access

03

Release Record

Source + Version — IDENTITY

Retain catalog origin, provider identity, client version, upstream ETag, and validation time.

Route + Region — ROUTING

Retain the unique slug-to-route map, deployed region, and provider ownership.

Behavior + Authority — CONTROL

Retain capability tests and the separate inference authorization decision.

BOUNDARY

- Cache hit ≠ current truth
- Visibility ≠ deployed route
- Metadata ≠ capability evidence
- Catalog ≠ inference authority

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

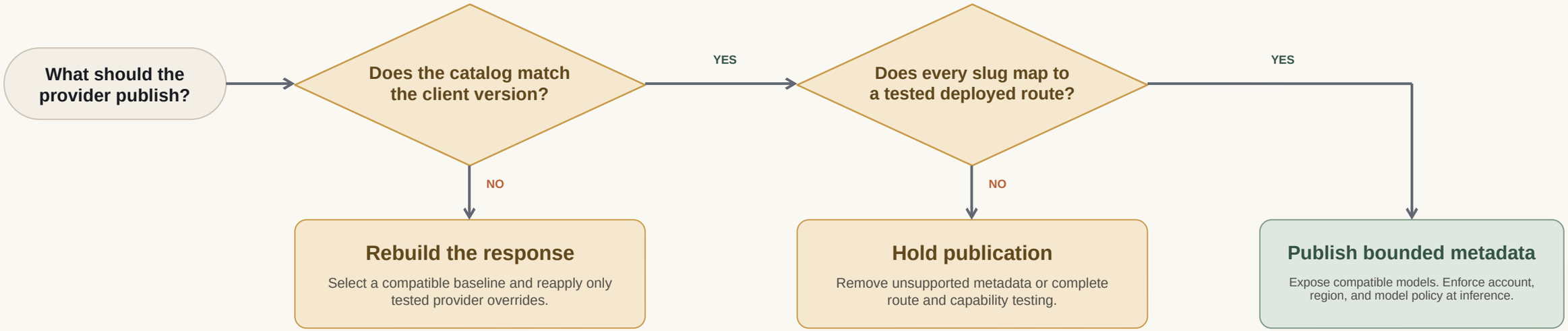
© 2026 Luminity Digital, Inc. | Independent synthesis of official OpenAI documentation.

Source: https://learn.chatgpt.com/docs/enterprise/roll-out-a-gateway#model-catalog-reference · verified Nov 5, 2026

OpenAI Select Partner

Governed Catalog Rollout

Verify compatibility and route truth before metadata reaches clients. Authorize inference separately.



RETAIN THIS RECORD WITH EVERY CATALOG RELEASE

Source + Version — IDENTITY

Retain catalog source, client version, provider identity, ETag, and validation time.

Route Truth — ROUTING

Retain the unique slug-to-route map, deployed region, and provider owner.

Capability Evidence — BEHAVIOR

Retain tests for reasoning, context, modalities, tools, and service tiers.

Inference Decision — AUTHORITY

Retain the separate account, region, and model-policy authorization.

BOUNDARY

- Source determines discovery
- Version determines compatibility
- Route truth needs evidence
- Inference uses separate authority

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

© 2026 Luminity Digital, Inc. | Independent synthesis of official OpenAI documentation.

Source: https://learn.chatgpt.com/docs/enterprise/roll-out-a-gateway#model-catalog-reference · verified Nov 5, 2026

OpenAI Select Partner