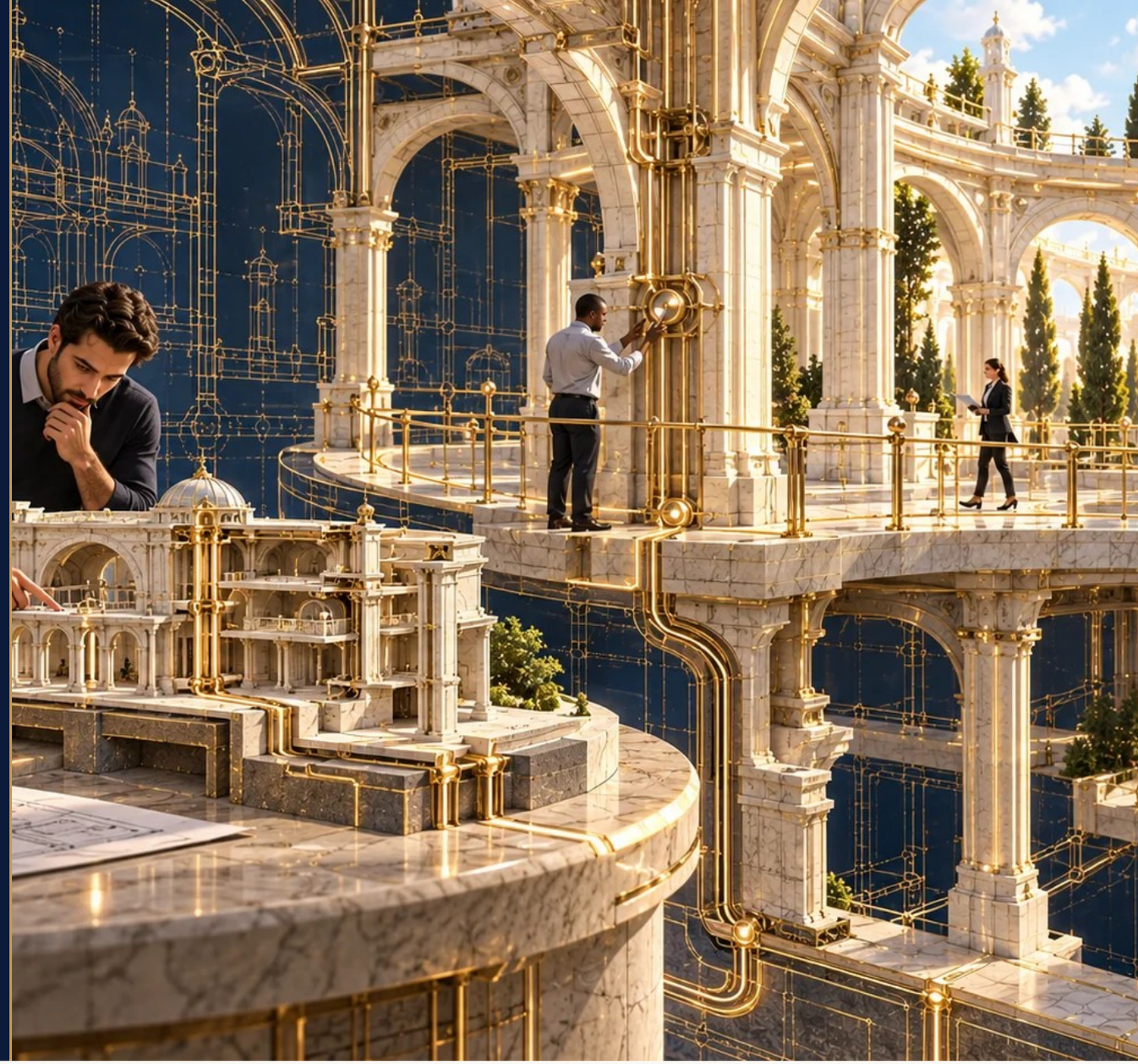
OPENAI ARCHITECTURE PRIMER

Private Safety Processing

Customer-controlled retention, protected review, and bounded signals

A visual guide to how eligible Zero Data Retention projects support offline automated safety review without OpenAI retaining customer prompts or responses.

FIRST-PARTY DOCUMENTATION / INDEPENDENT LUMINITY SYNTHESIS



Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

OpenAI Private Safety Processing

Source verified Nov 5, 2026

OpenAI Select Partner

A Protected Path for Offline Safety Review

Eligible Zero Data Retention projects can retain referred interactions in customer storage and expose only bounded safety signals.

AT A GLANCE

1
required retention posture: Zero Data Retention

30
days: minimum customer storage lifecycle

2
linked flows: retention and protected review

4
supported storage clouds: AWS, Azure, GCP, OCI

Request and retention
01

A selected interaction is protected and written asynchronously to customer-controlled regional storage.

Automated safety review
02

An attested runtime decrypts the record, returns bounded signals, and keeps detailed output protected.

IN THIS SET

Three chapters

01 **Request and retention**
How referral, encryption, routing metadata, and customer storage form the retention boundary.

02 **Protected safety review**
How attestation-gated decryption produces bounded signals without human access to content.

03 **Activation decision**
The eligibility, storage, lifecycle, and operating checks required before a project is ready.

Eligibility, storage support, permissions, and lifecycle requirements can change. Recheck the live OpenAI documentation before deployment.

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

Source: https://developers.openai.com/api/docs/guides/private-safety-processing · verified Nov 5, 2026

OpenAI Select Partner

Request and Retention

A referred interaction is protected before it crosses into customer-controlled regional storage.

01

Can the project use PSP?

✓ Zero Data Retention approved	REQUIRED
✓ Dedicated customer storage	REQUIRED
✓ Regional storage connection	VALIDATED
! Human review of retained content	NOT PERMITTED
! Inference waits for storage write	NO

Project-level policy → **Applies to all project API traffic**

02

How a referred interaction is stored

Input · prompt + response | **Selection · classifier or sampling** | **Destination · customer cloud**

1 **Infer** — The API returns the model response

2 **Refer** — Policy selects the interaction

3 **Protect** — HPKE and customer key controls apply

4 **Store** — The gateway writes to customer storage

SELECTION

A referral does not establish a policy violation

INDEX

OpenAI keeps metadata and a storage reference

FAILURE

A failed protected write has no OpenAI content fallback

03

Four architectural boundaries

API Interaction — INFERENCE

The prompt and response complete on the inference path before the storage task finishes.

Referral — SELECTION

A classifier signal or approved sampling policy selects the record for protected retention.

Protected Record — ENCRYPT

OpenAI HPKE protects the content and customer EKM authorization applies when enabled.

Customer Storage — CUSTOMER

AWS, Azure, Google Cloud, or OCI holds the referred content in regional customer storage.

BOUNDARY

- Referral does not prove a violation
- No OpenAI content copy
- Storage write is asynchronous
- Separate project for ZDR without PSP

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

Source: https://developers.openai.com/api/docs/guides/private-safety-processing · verified Nov 5, 2026

OpenAI Select Partner

Asynchronous Safety Review

Protected records return only to an approved automated review path with bounded plaintext output.

01

What may leave the runtime?

✓ Fixed-shape safety decision	ALLOWED
✓ Approved operational metadata	ALLOWED
✓ Detailed reviewer output	ENCRYPTED RETURN
! Prompt or response plaintext	PROHIBITED
! Human access to retained content	DISABLED

Attested identity → **Hardware-backed key release**

02

How protected review operates

Record · customer storage | **Trust · identity + measurements** | **Output · bounded + encrypted**

1 **Retrieve** — The gateway fetches the encrypted record

2 **Attest** — Identity and measurements authorize the key

3 **Review** — Automation decrypts and evaluates the record

4 **Release** — Only approved signals leave in plaintext

KEY
The scheduler and gateway never receive the decryption key

RUNTIME
Hardware attestation gates automated access

OUTPUT
Free-form text and transcript excerpts remain outside the plaintext boundary

03

Four control surfaces

Customer Storage — RECORD

The encrypted safety record remains under customer permissions and regional storage controls.

Key Boundary — TRUST

Attestation gates review access. PSP decryption is separate from EKM and uses a customer-controlled asymmetric key.

Review Runtime — REVIEW

The approved hardware-attested environment performs automated safety review without human access.

Two Output Paths — OUTPUT

Bounded signals support action. Detailed results return encrypted with the original expiration.

CONTROL RULE

- No plaintext at the gateway
- PSP decryption is separate from EKM
- Bounded signals only
- Detailed output returns encrypted

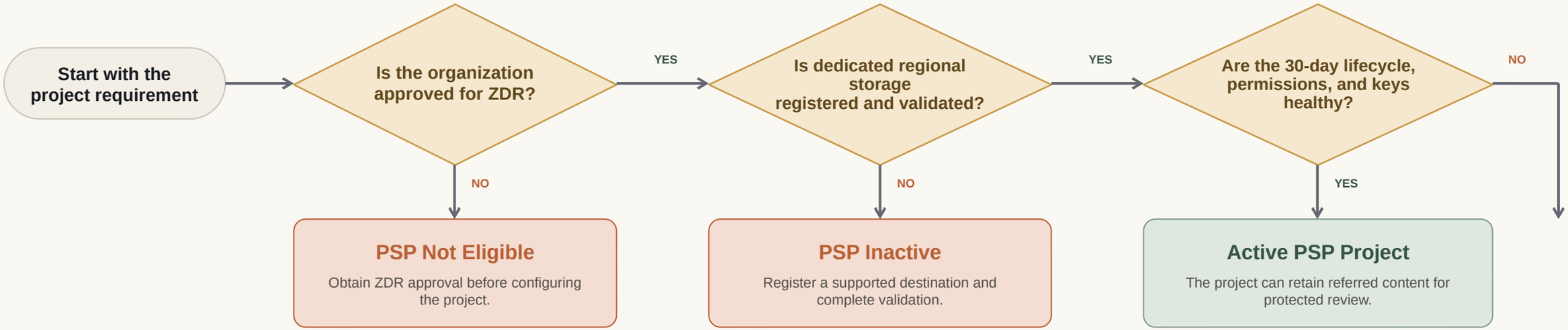
© 2026 Luminity Digital, Inc. | Independent synthesis of official OpenAI documentation.

Source: https://developers.openai.com/api/docs/guides/private-safety-processing · verified Nov 5, 2026

OpenAI Select Partner

PSP Activation Decision

A project is ready only when eligibility, storage, lifecycle, and operating controls all hold.



Repair Before Use

Restore lifecycle, permissions, or key authorization.

OPERATING CONTROLS AFTER ACTIVATION

30-Day Retention — RETENTION

Keep referred records for at least 30 days.

Access and Keys — ACCESS

Maintain storage access and customer key authorization.

Evidence — MONITOR

Treat registration status, cloud logs, and Compliance API events separately.

Response — OPERATE

Repair unhealthy configuration and investigate safety concerns.

ARCHITECTURE RULE

- Validation is a point-in-time check
- Disconnect resets project retention policy
- Existing cloud objects remain

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

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

Source: https://developers.openai.com/api/docs/guides/private-safety-processing · verified Nov 5, 2026

OpenAI Select Partner