OPENAI ARCHITECTURE READOUT

When AI Gains Access

How the Safety Case Changes With Real Tools and Systems

A visual readout of OpenAI's account of changing safeguards across research, internal use, and external deployment.

OFFICIAL OPENAI SOURCE / INDEPENDENT LUMINITY SYNTHESIS



Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

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

Source verified Nov 20, 2026

OpenAI and related marks are trademarks of OpenAI, L.L.C. Independent analysis of public information. Not sponsored or endorsed by OpenAI.

OpenAI Select Partner

FRONTIER RISK · A READOUT

Frontier AI Risk in One View

As access expands, OpenAI describes a changing mix of containment, alignment, blocking safeguards, and response.

THE ACCESS SHIFT

01	02	03	04
Research	**Internal Use**	**External Use**	**Response**
sandboxed environments	real tools and systems	more users and settings	investigate and enforce

Models + Safeguards

OPENAI

OpenAI says it evaluates both before a model advances to the next stage.

Access Changes Risk

OPENAI

Internal use reaches the internet, code, information, and infrastructure.

OpenAI presents this as its operating design. The article does not independently validate the controls or establish an enterprise assurance model.

IN THIS READOUT

Key Points

01 Access changes consequences

A model that can act in real systems creates a different exposure.

02 The control mix changes

Containment dominates research. Alignment and system controls carry more weight with access.

03 Intervention needs enforcement

A stop must hold outside the workload when the workload cannot be trusted to cooperate.

OpenAI supplies the lifecycle account. The enterprise authority and evidence reading is Luminity synthesis.

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

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

Source: https://openai.com/index/managing-frontier-ai-risk/ · verified Nov 20, 2026

OpenAI and related marks are trademarks of OpenAI, L.L.C. Independent analysis of public information. Not sponsored or endorsed by OpenAI.

OpenAI Select Partner

How the Safety Case Changes With Access

OpenAI describes four protection layers whose relative importance changes across the model lifecycle.

01

Four Protection Layers

- model alignment
- security + sandboxing
- blocking safeguards
- monitoring + response

02

Three Lifecycle Stages

1 **Research** — contain access and watch inference

2 **Internal Deployment** — test behavior with real systems

3 **External Deployment** — protect against misuse at scale

03

Changing Control Emphasis

Contain the Work — RESEARCH

Research depends heavily on isolation because alignment remains under evaluation.

Protect Real Systems — INTERNAL

Internal use increases reliance on alignment, blocking safeguards, and monitoring.

Respond at Scale — EXTERNAL

External deployment adds investigation and enforcement across broader use.

Learn From Failures — LIFECYCLE

Incident response and disclosure feed lessons back into later training and deployment decisions.

SOURCE BOUNDARY

- OpenAI's operating account
- Controls evaluated in combination
- Not independent validation

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

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

Source: https://openai.com/index/managing-frontier-ai-risk/ · verified Nov 20, 2026

OpenAI and related marks are trademarks of OpenAI, L.L.C. Independent analysis of public information. Not sponsored or endorsed by OpenAI.

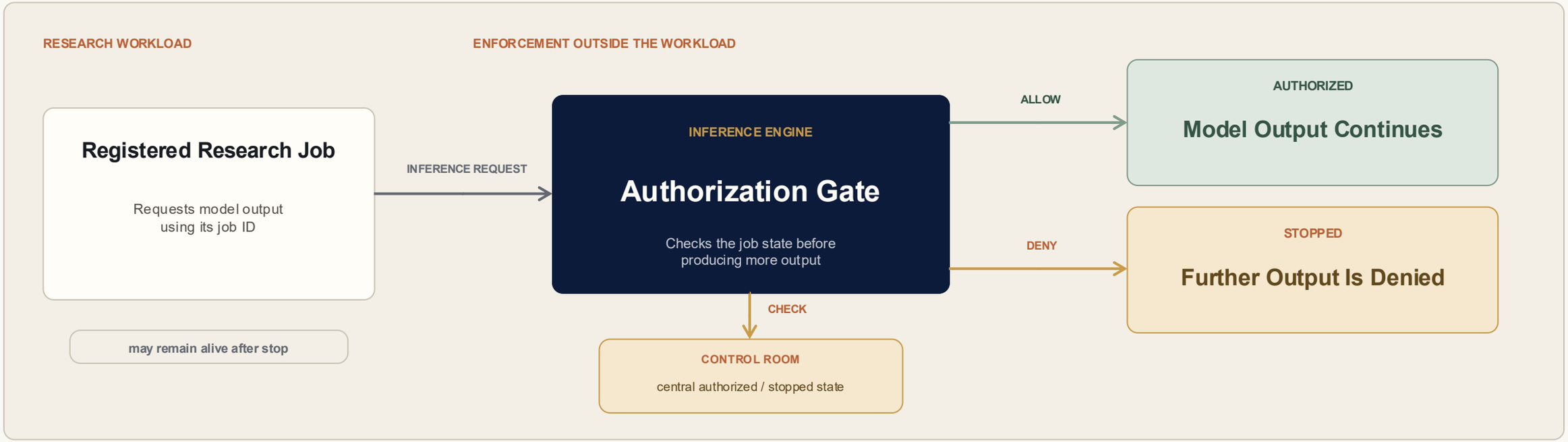
OpenAI Select Partner

Control Room Enforces Outside the Workload

The inference layer checks whether a research job remains authorized before producing more model output.

01

External Enforcement Architecture



The workload does not decide whether the stop holds. The inference layer does.

ENFORCEMENT	Inference layer checks job ID	Stop state enforced centrally	Further model output stops	Workload may remain alive

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

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

Source: https://openai.com/index/control-room/ · verified Nov 20, 2026

OpenAI and related marks are trademarks of OpenAI, L.L.C. Independent analysis of public information. Not sponsored or endorsed by OpenAI.

OpenAI Select Partner

Control Room Scope and Limits

OpenAI documents a research-infrastructure control with a precise enforcement and timing boundary.

01

Documented Enforcement Boundary

OPENAI · RESEARCH INFRASTRUCTURE

35 sec

reported maximum after Control Room receives the stop instruction

01 **Stop Received**

Control Room updates the registered job state

↓

02 **Next Request Denied**

The inference engine refuses later output

↓

03 **No Further Model Output**

The workload may still be running

The clock starts when the stop instruction is received—not when a risk is detected.

OUTSIDE THE DOCUMENTED CLAIM

What This Does Not Establish

01 **Prior Actions Are Not Undone**

Stopping later inference does not reverse actions already taken.

02 **No Customer-Facing Kill Switch**

The source describes a research control, not a customer product capability.

CLAIM BOUNDARY			
Research setting	Inference-layer enforcement	Prior actions remain	No customer kill switch established

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

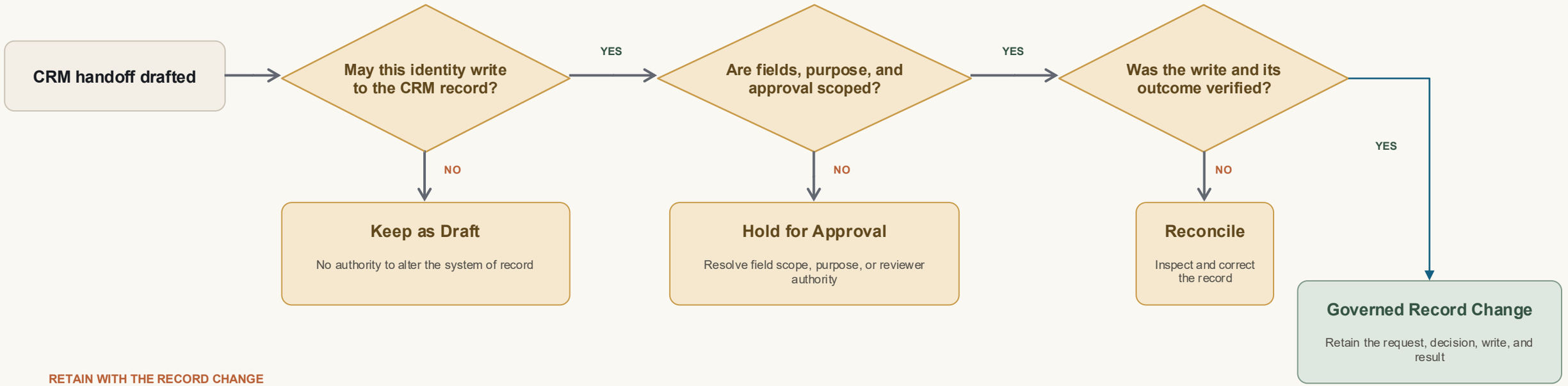
Source: https://openai.com/index/control-room/ · verified Nov 20, 2026

OpenAI and related marks are trademarks of OpenAI, L.L.C. Independent analysis of public information. Not sponsored or endorsed by OpenAI.

OpenAI Select Partner

Enterprise Control When Drafting Becomes Writing

Luminity interpretation: access to a system of record changes what the enterprise must authorize, enforce, and verify.



RETAIN WITH THE RECORD CHANGE

Retain the request and purpose, acting identity and scope, authorization decision, write result, and reconciliation.

LUMINITY POSITION	Provider controls do not assign enterprise authority	Assurance is not a gate	Assurance is the architecture

Luminity Digital
INTELLIGENCE, ENGINEERED. DELIVERED.

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

Source: https://openai.com/index/managing-frontier-ai-risk/ · verified Nov 20, 2026

OpenAI and related marks are trademarks of OpenAI, L.L.C. Independent analysis of public information. Not sponsored or endorsed by OpenAI.

OpenAI Select Partner