OPENAI ARCHITECTURE PRIMER

Where Agent Permission Is Enforced

Reasoning, policy, approval, credentials, and execution boundaries

A visual guide to the controls that deny, escalate, or equip a proposed agent action before execution.

NIST TENETS / OPENAI BEHAVIOR / LUMINITY SYNTHESIS



Permission is enforced across several control boundaries

A model can propose a tool call. Separate controls decide whether the action is denied, escalated, or equipped for bounded execution.

AT A GLANCE

2
architectural planes:
reasoning and authority

3
gate outcomes:
deny, escalate, or equip

5
NIST tenets owned
by this edition

1
downstream authorization
recheck before side effects

TWO DISTINCTIONS PREVENT FALSE PERMISSION

Proposal versus authorization

Instructions and model reasoning can select an action. Current identity, policy, credentials, environment, and resource checks decide whether it may execute.

Approval versus permission

Human or automatic review can approve a request. The decision does not expand roles, override network policy, or replace downstream authorization.

DOCUMENTED BOUNDARY

OpenAI documents distinct controls for roles, review, sandboxes, networks, vaults, and MCP calls. The reviewed sources do not define one universal cross-system policy engine.

THREE CHAPTERS

01 **Dual-plane architecture**

Keep action selection inside the reasoning plane and permission decisions inside an external authority plane.

02 **Controlled action path**

Validate the request, evaluate policy and review, release bounded credentials, and preserve downstream checks.

03 **Enforcement decision**

Return an explicit disposition: deny, hold for the right authority, or equip one controlled execution path.

NIST defines the target separation. OpenAI documents current primitives. Luminity composes the authority plane.

© 2026 Luminity Digital, Inc. | Local review draft

NIST tenets 03, 04, 05, 07 and 08 + OpenAI RBAC, WIF, automatic review, sandbox, vault, and MCP docs · verified Nov 5, 2026

OpenAI Select Partner

Reasoning plane and authority plane

The model proposes an action. External controls determine whether permission and execution capability reach the target.

WHAT EACH LAYER CONTRIBUTES

Model and instructions	PROPOSE ACTION
Identity and role scope	BOUND PLATFORM ACCESS
Automatic review	MODEL-MEDIATED GATE
Sandbox and network	RESTRICT EXECUTION
Resource server	AUTHORIZE SIDE EFFECT

Authority-plane policy

Enterprise-owned decision logic combines current identity, mandate, resource, action, risk, environment, and separation-of-duties requirements.

CONTROLLED CROSSING



INDEPENDENT ENFORCEMENT POINTS

RBAC and mapping

Roles, key permissions, and federation mappings bound OpenAI Platform access.

Vault or broker

Credentials stay outside instructions and can reach only approved paths or application tools.

Sandbox and network

Isolation, files, permissions, and outbound endpoints limit available execution capability.

Resource server

The MCP server or connected application applies its own authorization before side effects.

KEEP CONTROLS DISTINCT Instruction ≠ policy decision ≠ approval ≠ credential release ≠ sandbox permission ≠ downstream authorization

© 2026 Luminity Digital, Inc. | Local review draft

NIST policy-separation and deterministic-decision tenets + OpenAI RBAC, WIF, sandbox, vault, and MCP docs · verified Nov 5, 2026

OpenAI Select Partner

Controlled action path

Every checkpoint answers a different question before bounded authority reaches an external resource.

WHAT EACH CHECKPOINT ESTABLISHES

Tool and arguments valid	REQUEST IS WELL FORMED
Identity and scope current	PLATFORM ACCESS BOUNDED
Risk threshold evaluated	REVIEW PATH SELECTED
Credential path constrained	SECRET EXPOSURE LIMITED
Resource authorizes call	SIDE EFFECT PERMITTED

Review routing

Enterprise risk policy decides which requests require deterministic denial, automatic review, or the correct human authority.

ACTION PATH



CONTROL SURFACES AROUND THE CALL

Identity and mandate

Bind the accountable principal, active workload, approved task, and current delegated scope.

Action and audience

Match tool, operation, resource, destination, data class, and permitted audience.

Environment boundary

Keep network restrictions, sandbox permissions, and credential-host rules independently active.

Downstream recheck

The MCP server or connected application validates identity, scope, and resource access again.

APPROVAL LIMIT Approval does not expand roles, override network policy, disclose a secret, configure sandboxing, or replace downstream authorization

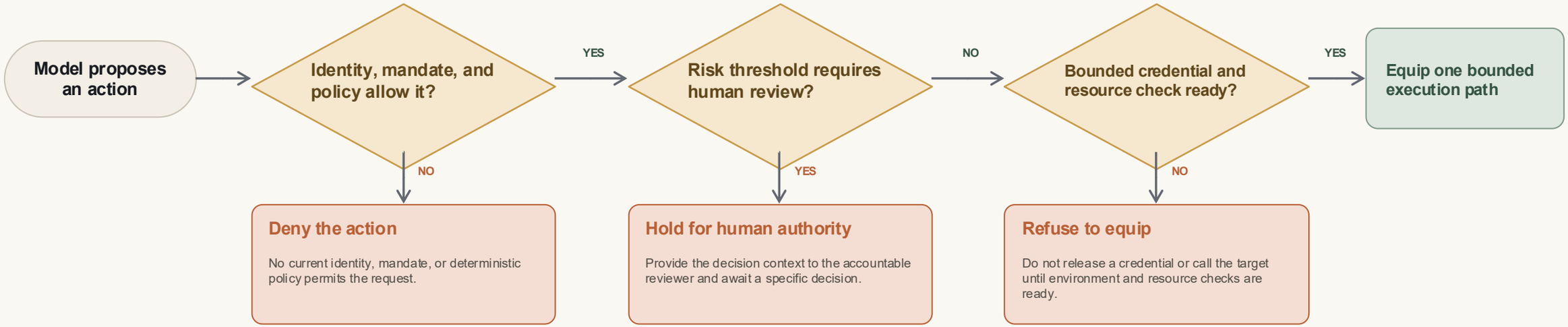
© 2026 Luminity Digital, Inc. | Local review draft

OpenAI automatic review, sandbox security, vault, MCP approval, RBAC, and WIF documentation · verified Nov 5, 2026

OpenAI Select Partner

Governed permission decision

A proposed action crosses the execution boundary only when current policy, review requirements, credentials, environment, and downstream checks agree.



REQUIRED OVERLAYS

Identity and scope

Principal, workload, mapped account, role union, mandate, and delegated limits.

Policy and duties

Action, resource, context, risk threshold, conflict checks, and fail-closed rule.

Credential and environment

Audience, expiry, approved host, network path, sandbox permissions, and revocation.

Downstream decision

Resource-server authorization, execution result, denial, approval, and policy-decision record.

ARCHITECTURE RULE — **No model or reviewer decision expands standing permission.**

OPENAI BOUNDARY — Automatic review can decide supported requests; it does not override network restrictions or downstream authorization.

© 2026 Luminity Digital, Inc. | Local review draft

NIST human-review and deterministic-decision tenets + OpenAI automatic review, sandbox, vault, and MCP docs · verified Nov 5,2026

OpenAI Select Partner