Agent authorization reference architecture

runtime request · control input · evidence

CONTROL PLANE
authoritative identity, task, policy, grants and lifecycle sources

GOVERNANCE
Registry + IGA
agent owner / sponsor
tool inventory • lifecycle

IDENTITY
IdP + trust
user / workload identity
keys • status • grants

POLICY
PAP + policy store
rules • versioning
entitlements • risk inputs

TASK AUTHORITY
Trusted issuer
workflow template
signed bounded task

ELEVATION
Approval + PAM
exact-action approval
short-lived credentials

RESPONSE
Security operations
revocation signals
exceptions • reconciliation

INITIATOR + AGENT ZONE

ENTRY
Human / job
delegated user or
approved schedule

WORKFLOW
Entrypoint
authenticate + authorize
load approved template

RUNTIME
Agent A
model output untrusted
no raw target credential
proposes tool action

OPTIONAL HANDOFF
Agent B
new child task from issuer
scope / expiry cannot grow
authenticates as B

tool call

direct call

SIGNED TASK AUTHORITY
issuer • initiator / service • bound agent • allowed actions
resource selectors • expiry • approval ref • delegation depth

task ref

model → tool boundary

AUTHORIZATION + TOOL EXECUTION

TOOL ADAPTER + PEP
mandatory per-call enforcement

01 Authenticate agent; verify task + binding

02 Catalog tool; resolve action + resource + side effect

03 Canonicalize params; hash exact request

04 Ask PDP; bind approval to exact request

05 Broker grant; call target; capture outcome

DENY / ERROR ⇒ no downstream call

POLICY DECISION
PDP
user/service ∩ agent
∩ task ∩ target ∩ risk
ALLOW / DENY / STEP-UP

STEP-UP
Approval
eligible approver
payload hash + expiry

GRANT
Broker
target-scoped TTL
fallback limit recorded

PEP → target boundary

TARGET ZONE

RESOURCE SERVER
Target API
SaaS • database • queue
validates native permission
executes exact operation
returns target request ID

OBSERVATION
Target logs
authoritative state / result
correlate request ID
unknown ≠ success

EVIDENCE + CORRELATION
task chain • initiator • acting agent • canonical action / resource • policy version + decision ID
approval / grant ref • target request ID • requested / allowed / executed / confirmed / unknown

REVOCATION + RECONCILIATION
disable agent/task; revoke grants where supported
invalidate future PEP decisions; compare target logs

COVERAGE INVARIANT **No direct agent credential or alternate target route may bypass the PEP; otherwise enforcement is partial.**