PRINCIPALFRAME / RESEARCH 001 · VERSION 1.0

From Authorization to Alignment

Governing purpose and intent across human, service, workload and agent principals

Identity, authorization and trust for humans, machines and autonomous systems.

Author: Vaibhav Dhote · Independent research and engineering · October 2026 · principalframe.com

ABSTRACT

Identity controls have always sought to answer a small set of fundamental questions: who or what is acting, what that principal is permitted to do, and what happened as a result. These questions underpin the familiar AAA model of Authentication, Authorization and Accounting.

The controls and the principals they govern have both evolved. Authentication has moved from passwords toward MFA and phishing-resistant methods; digital principals from relatively predictable services to ephemeral workloads and now autonomous agents that can select actions in pursuit of an objective.

Foundational AAA architectures focus on establishing a principal, deciding whether access should be granted and recording activity. A separate problem has always existed: whether legitimate authority is being used for its intended purpose. For humans, organisations have largely governed that through policy, supervision and accountability; for conventional software, purpose could remain embedded in implementation.

Autonomous agents challenge that assumption. Digital principals can now exercise discretion over which technically permitted action to take.

> **The intent problem is not new. What is new is the need to govern it as a technical control for digital principals.**

This paper asks whether we can control not just what a principal may do, but whether the action it chooses is consistent with the purpose and current task objective under which its authority is exercised.

PrincipalFrame treats this as a distinct control concern: Alignment - whether a proposed Action, while exercising legitimate Authority, is consistent with approved Purpose and current Intent within applicable Constraints.

The paper introduces a 4A control model - Authentication, Authorization, Alignment and Accountability - and applies it across human, service, workload and agent principals. Using People, Process and Technology as the analytical lens, it argues that Authentication, Authorization and Accountability have established technical mechanisms, while Alignment remains distributed across governance, process and implementation.

As digital principals gain discretion, that arrangement becomes less sufficient. Purpose, task objective and constraints increasingly need machine-evaluable representations available to technical controls.

CONTENTS

Sections and principal-archetype analysis

1. The control model we already have 4
2. From human users to autonomous principals 5
3. The control gap: permission is not purpose 5
4. From Authorization to Alignment 6
5. The 4A Control Model 8
6. Applying the 4A Model Across Principal Archetypes 9
- 6.1 Human principals: where Alignment began 9
- 6.2 Service identities: Alignment embedded in implementation 9
- 6.3 Workload principals: identity becomes dynamic, purpose remains bounded 10
- 6.4 Agent principals: Alignment enters the technical control plane 10
7. The Alignment Transition 11
8. What Technical Alignment Requires 12
9. An Alignment Maturity Path 13
10. Emerging Work: The Control Plane Is Beginning to Change 14
11. Scope and Limitations 14
Conclusion 15
References and Further Reading 15

1. The control model we already have

Identity and access control has long been organised around three familiar questions:

Authentication - who or what is acting?

Authorization - what is that principal permitted to do?

Accounting - what activity took place?

Together, these form the familiar AAA model: Authentication, Authorization and Accounting. The model has endured because each layer has evolved. [3]

Authentication has moved beyond passwords to MFA, cryptographic authenticators and passkeys; Authorization from direct permissions toward role-, attribute- and policy-based decisions; Accounting from basic usage records toward richer audit, session and governance evidence. [1,2]

The established control layers

Control layer	Primary question	Common mechanisms and technologies
Authentication	**Who or what is acting?**	Passwords, MFA, passkeys/FIDO2, certificates, federation, managed identities, short-lived credentials, device or workload attestation
Authorization	**What is the principal permitted to do?**	ACLs, RBAC, ABAC, policy-based access, rules, scopes and claims, contextual access, JIT/JEA, delegated authorization
Accounting	**What happened?**	Audit logs, access logs, transaction records, session recording, SIEM correlation, identity-governance records

Modern Authorization is not simplistic: it can evaluate identity, device, resource, environmental and risk context. The question is why another control concern is needed. [2]

> **Context is not intent**
>
> Context-aware authorization can make a more informed decision about whether an action is permitted. It does not necessarily establish the objective for which the principal is performing that action.
>
> Consider an HR administrator exporting employee records. The system may confirm the user, managed device, HR role and export permission - all correctly.
>
> Those controls can still pass while the export is performed for a purpose unrelated to the authority granted. Context can tell us who, where, when and under what conditions an action occurs; it does not, by itself, establish whether exercising that permission is consistent with its intended purpose.

A sufficiently expressive authorization system can include Purpose or Intent as policy inputs. The distinction is therefore not technical capability but control concern: Authorization asks whether an action is permitted; Alignment asks whether exercising that authority is consistent with the current objective.

In mainstream enterprise identity, that second question has usually sat outside the technical control plane.

For humans, policy, supervision, approvals and retrospective accountability carried much of the burden; for conventional software, purpose was often implicit in design and workflow.

As digital principals gain discretion over which permitted action to take, that control concern increasingly needs explicit technical enforcement.

2. From human users to autonomous principals

Enterprise identity began with human users and expanded to services and cloud-native workloads with identities of their own.

The important change is not the number of identities, but the nature of the principal exercising authority.

In this paper, a principal is a human or non-human entity capable of acting within a system and exercising authority.

Agents introduce a further shift. Unlike a conventional service or workload whose behaviour is largely determined by implementation, an agent may receive an objective and select among available actions while pursuing it.

Discretion refers here to a principal's ability to select among multiple available actions or execution paths while pursuing an objective.

For the purposes of this paper, four principal archetypes provide a useful way to examine this evolution.

Principal archetype	What it represents	How purpose is typically expressed	Degree of discretion
Human	A natural person acting through a digital identity	Role, responsibility, policy, instruction and task	**High** - the individual decides how available authority is exercised
Service	A persistent application, service or automated function	Application design, configuration and defined business function	**Low** - behaviour is largely predetermined by implementation
Workload	Executing software associated with a runtime environment	Deployment, configuration and runtime context	**Low / bounded** - execution may be dynamic, while intended function remains largely predetermined
Agent	Software capable of pursuing an objective through selectable actions	Objective, instructions, delegated authority and runtime context	**Potentially high** - the agent may select how to pursue the objective

These are analytical archetypes, not rigid identity classes: a service may run as one or more workloads, and an agent may itself use services and workloads.

The key variable is discretion.

As discretion increases, a new question becomes important: how is the principal choosing to use authority it already has?

The next section tests that question using the same underlying authority in a conventional service and an autonomous agent.

3. The control gap: permission is not purpose

The limitations of the existing control model become clearer when the same authority is exercised by principals with different degrees of discretion.

Consider a customer who contacts a support service and asks:

> **Why was I charged twice?**

A conventional customer-service application may legitimately have both transaction-read and refund capability because refunds are required in some support cases. The question is not whether that capability should exist, but whether using it fits this interaction.

A predefined workflow can constrain how those capabilities are used: for this request, retrieve the transaction and explain what happened. The execution path is largely predetermined.

An autonomous agent can hold the same capabilities but choose among them.

Given the same request, it identifies the duplicate charge and decides that issuing a refund will resolve the problem.

From the perspective of the established control model, nothing necessarily went wrong.

Control	Conventional service	Autonomous agent
Authentication	The service identity is successfully established	The agent identity is successfully established
Authorization	The principal is permitted to read transactions and invoke the available functions	The principal is permitted to read transactions and invoke the available functions
Accounting	Actions are recorded and attributable	Actions are recorded and attributable
How authority is exercised	Behaviour is substantially constrained by the implemented workflow	The agent selects an available action in pursuit of the objective

AAA can all operate as designed: identity is established, the refund is authorized and the action is recorded. Those controls still do not determine whether a refund is the right response to a request for explanation.

The refund may be aligned or misaligned depending on the task policy and constraints. The point is that AAA alone does not answer whether exercising that permission is consistent with Purpose and Intent within applicable Constraints.

> **No AAA control necessarily failed. That is the gap this paper examines.**

Humans have always presented the same problem: legitimate authority can be used for the wrong purpose. Organisations have addressed that through policy, supervision, approvals and retrospective accountability.

Conventional software often constrained purpose through code and predefined workflow.

Autonomous principals weaken that assumption: greater discretion pushes the same control problem into the technical control plane.

4. From Authorization to Alignment

Authorization establishes whether a principal may perform an action. Alignment asks whether using that authority fits the objective being pursued.

Two distinctions matter before defining Alignment:

Authority - the scope of actions a principal is empowered to perform, including the source, conditions and limits of that empowerment.

Permission - a specific authorization grant or policy outcome allowing an action or access to a resource.

The core Alignment relationship is Purpose → Intent → Action. Applicable Constraints bound how that authority may be exercised:

Purpose - the approved reason for which authority exists.

Intent - the current task objective the principal is pursuing, as declared, delegated or otherwise represented to the control system.

Constraints - the applicable boundaries and conditions within which authority may be exercised.

Action - a concrete operation the principal proposes to perform or performs.

In the customer-support example:

> Purpose → Intent → Action
> within applicable Constraints
>
> Purpose: Support the customer
> Intent: Explain the duplicate charge
> Constraint: Do not make financial changes unless resolution requires it
> Proposed Action: Issue a refund

The refund may be authorized, but authorization alone does not establish that it is aligned with the customer's request.

This paper refers to that additional control concern as Alignment: the control concern that evaluates whether a proposed Action, while exercising legitimate Authority, is consistent with the approved Purpose and current Intent within applicable Constraints.

Alignment here is an identity and security control concern, not AI alignment.

Purpose-aware technical control is not new: purpose-based access control models intended use, while Usage Control (UCON) adds obligations, conditions and ongoing control. PrincipalFrame's proposition is to treat the Purpose-Intent-Action relationship, bounded by applicable Constraints, as a distinct architectural control concern that becomes more important as digital principals gain discretion. [4,5]

Healthcare provides established machine-readable purpose-of-use precedent: OASIS XSPA incorporates Purpose of Use as an access-policy attribute, supported by HL7 terminology. This captures an asserted purpose of access, not necessarily the current task objective of a discretionary agent. [15,16]

Alignment complements Authorization and does not require a separate technology.

Together with Authentication and Accountability, it yields the 4A model formalised next.

5. The 4A Control Model

PrincipalFrame extends the familiar AAA model by making Alignment an explicit control dimension and by using Accountability as the broader expression of what traditional AAA describes as Accounting.

Figure 1 shows the four complementary controls around an exercise of authority.

Figure 1

The PrincipalFrame 4A Control Model

Four complementary control concerns around an exercise of authority; they are not a fixed execution sequence.



Figure 1. Four complementary control concerns around an exercise of authority.

Alignment complements rather than replaces Authorization: an action must be legitimately permitted before its consistency with Purpose, Intent and Constraints is evaluated.

From Accounting to Accountability

Traditional AAA uses Accounting for recording activity and resource use.

PrincipalFrame broadens this to Accountability: preserving enough evidence to attribute and reconstruct an action, the authority under which it occurred and its outcome. For autonomous principals, that evidence may also need to preserve delegation and task context so the exercise of authority can be reconstructed meaningfully.

A model for all principals

The model applies to human, service, workload and agent principals alike; what changes is where each control is realised and how explicit and technically enforceable it must become.

The four controls are complementary concerns around an exercise of authority, not a fixed execution sequence.

They may be implemented by separate or shared mechanisms, and Accountability may operate continuously.

Section 6 applies the model across the four principal archetypes.

6. Applying the 4A Model Across Principal Archetypes

The following four subsections apply the same 4A and People / Process / Technology lens to the four principal archetypes. The purpose is to compare where each control is realised as the nature and discretion of the principal change, not to treat the archetypes as mutually exclusive technical identity classes.

6.1 Human principals: where Alignment began

Humans are the original discretionary principals: they interpret objectives and choose among permitted actions. The Alignment problem therefore existed long before autonomous agents, and organisations addressed it across People, Process and Technology.

4A control	People	Process	Technology
Authentication	User establishes and protects their identity and authenticators	Identity proofing, enrolment, credential recovery and assurance requirements	Passwords, MFA, passkeys/FIDO2, certificates, federation, device and risk signals
Authorization	Managers and resource owners determine legitimate need	Access request, approval, role governance, SoD, access review and privilege lifecycle	ACLs, RBAC, ABAC, policy engines, contextual access, JIT/PIM
Alignment	**Human judgement, management supervision, professional responsibility and behavioural expectations**	**Policy, acceptable-use rules, business purpose, approvals, supervision and disciplinary/legal controls**	**Primarily indirect controls such as DLP, behavioural analytics, transaction rules and monitoring; limited direct purpose/intent enforcement**
Accountability	Managers, auditors, investigators and control functions	Audit, certification, compliance review and incident investigation	Audit logs, session recording, SIEM, transaction records and identity-governance evidence

The asymmetry is clear: Authentication, Authorization and Accountability have extensive technical enforcement, while Alignment still relies heavily on human judgement, policy and oversight.

Consider an employee authorized to access payroll records. Authentication and Authorization can both be correct while the reason for viewing a particular colleague's salary remains outside the access decision.

Humans are embedded in governance: managers supervise, policies constrain behaviour, approvals gate sensitive actions and investigations can ask why an action occurred.

For humans, technology handled much of who, what and what happened; People and Process carried much of why.

6.2 Service identities: Alignment embedded in implementation

A traditional service identity represents software performing a defined function - payroll, backups, payments or integration. Its objective and behaviour are largely set by design, configuration and workflow.

4A control	People	Process	Technology
Authentication	Technical or application owner remains responsible for the identity	Credential issuance, protection, renewal, rotation, expiry and lifecycle	Passwords/secrets, certificates, service principals, managed identities and federation
Authorization	Application and resource owners determine required access	Access approval, least privilege, review and service-account lifecycle	Roles, scopes, API permissions, ACLs, policy engines and JIT access
Alignment	Business and application owners define what the service is intended to do	Requirements, architecture, design review, testing and change control	**Purpose is primarily constrained through code, configuration, workflow and the interfaces exposed to the service**
Accountability	Application, operations and security owners	Monitoring, incident response, operational review and ownership	Application logs, API logs, transaction records, service-identity logs and SIEM

Alignment is usually embedded in implementation: a payroll service does not redefine its purpose because its behaviour is constrained by what has been designed and deployed.

Business purpose becomes requirements, code and workflow; change control governs alterations.

Because behaviour is largely predetermined, the identity system rarely needs a changing task objective.

6.3 Workload principals: identity becomes dynamic, purpose remains bounded

Workloads make identity dynamic: instances can be created, scaled and destroyed while authenticating through runtime-issued, federated or short-lived credentials. [7-9]

4A control	People	Process	Technology
Authentication	Platform and application owners establish workload trust requirements	Registration, deployment, federation and workload lifecycle	Runtime-issued credentials, workload federation, managed identities, SPIFFE/SVIDs, mTLS and short-lived tokens
Authorization	Platform and resource owners determine permitted relationships	Policy definition, deployment governance, least privilege and access review	RBAC/ABAC, resource policy, scopes, workload-aware access policy and contextual controls
Alignment	Platform/application teams define the function of the workload	Architecture, CI/CD, deployment configuration, namespace boundaries and change control	**Purpose is substantially represented through code, deployment identity, configuration and runtime environment**
Accountability	Platform, application and security teams	Observability, operational review, change traceability and incident response	Workload logs, platform telemetry, distributed tracing, identity-aware logs and SIEM

A short-lived payment-reconciliation workload may be dynamic but still perform a predetermined function. Its purpose can often be inferred from code, deployment identity, configuration and allowed resources.

> **Dynamic identity does not necessarily imply discretionary behaviour.**

Services and workloads differ in how identity is established, but both can rely heavily on predetermined implementation to constrain purpose.

6.4 Agent principals: Alignment enters the technical control plane

Agents combine machine-speed execution with greater discretion: they can receive an objective, select tools and actions, and adapt the execution path.

4A control	People	Process	Technology
Authentication	Technical owners establish and maintain trusted agent identity	Agent registration, creation, credential lifecycle and trust establishment	Dedicated agent identity, workload identity, cryptographic credentials and token-based authentication
Authorization	Resource owners and sponsors determine acceptable authority	Access request, delegation, approval, least privilege and review	Roles, scopes, delegated permissions, application permissions, policy engines and task/JIT controls
Alignment	**Human sponsor defines business purpose; users or systems provide current objectives**	**Purpose definition, task approval, delegation constraints, escalation and human oversight**	**Emerging: structured task/objective representation, action gating, operation-specific authorization and runtime evaluation of proposed actions against objectives**
Accountability	Sponsors, technical owners, security and audit functions	Lifecycle governance, access review, investigation and delegation oversight	Agent-specific activity logs, tool-call records, delegation evidence, task history and identity/authorization telemetry

Authentication and Authorization are becoming explicit through agent identities, delegation and governance. Alignment is less mature: once an agent receives an objective, it may choose among several technically authorized actions. [6,10,11]

Human governance cannot inspect every machine-speed decision, while predetermined code no longer fixes the entire execution path.

Some information previously carried by governance or embedded in implementation must therefore become available to technical controls; Section 8 defines that information explicitly.

The agent does not create the Alignment problem; it changes where that problem must be solved.

7. The Alignment Transition

Across the four archetypes, Alignment follows a different path from Authentication, Authorization and Accountability.

Figure 2

Alignment Across Principal Types

Discretion falls sharply in conventional digital systems and returns with agents, changing where Alignment must be realised.

	Human Human identity	Service Service identity	Workload Workload identity	Agent Agent identity
Typical discretion	High	Low	Low / bounded	Potentially high
How Alignment is addressed	People + Process	Process + Software design	Software + Deployment/runtime context	People + Process + emerging Technology
Why the approach works — or must evolve	Human supervision, policy, judgement and organisational accountability governed Purpose.	Behaviour and Purpose were substantially constrained by implementation.	Dynamic execution still operated within a largely predetermined function.	Greater discretion makes Purpose and Intent increasingly necessary at runtime.

The change is not simply human versus machine identity. It is the return of greater discretion inside a digital principal.

Figure 2. Discretion and the dominant location of Alignment across principal archetypes.

The pattern is clear: humans relied on governance; services encoded purpose in software; workloads added dynamic identity without much new discretion; agents reintroduce discretion in a digital principal.

The gap appears because human governance sits too far from machine-speed execution while predetermined implementation no longer fully determines the path. Alignment therefore moves from governance-led, toward technology-assisted and ultimately toward runtime enforcement.

The next question is what technical Alignment must evaluate and how that capability matures.

8. What Technical Alignment Requires

Technical Alignment first requires information, not a new protocol or product.

A system needs machine-evaluable representations of the principal, its Authority, approved Purpose, current Intent, applicable Constraints and Proposed Action; Accountability also preserves the resulting Outcome.

At minimum, that information includes the relationship shown in Figure 3.

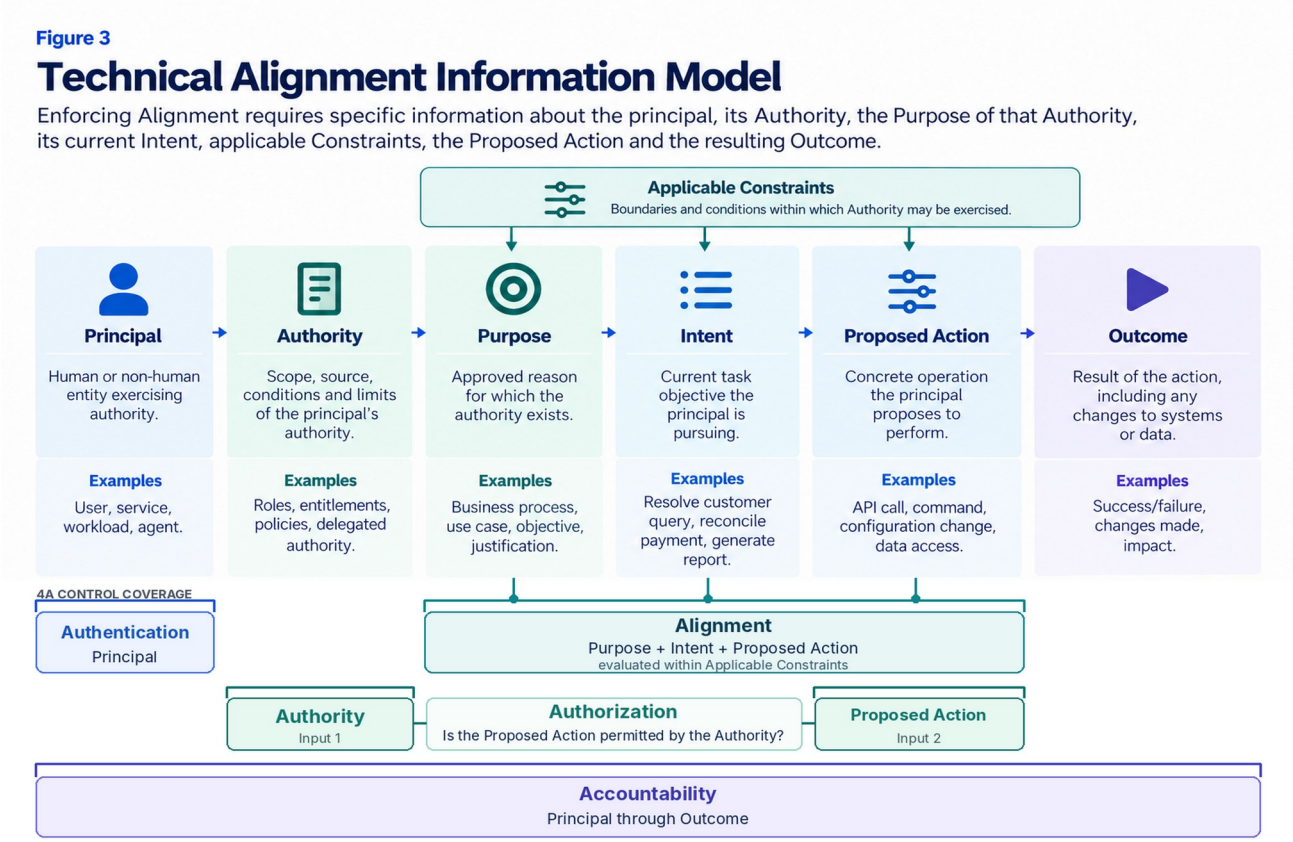


Figure 3. Purpose, Intent and Proposed Action operate within applicable Constraints; the 4A controls overlap across the information model.

Across this model, Authentication establishes the principal; Authorization establishes whether the principal has the authority required for the proposed action; Alignment evaluates that Action against Purpose and Intent within applicable Constraints; Accountability preserves the evidence needed to reconstruct the exercise of authority and its Outcome.

The running customer-service example can be expressed in the same terms:

> **Principal:**
> Customer-service agent
> **Authority:**
> May read transactions and issue refunds
> **Purpose:**
> Provide customer support
> **Intent:**
> Explain a duplicate charge
> **Constraint:**
> Do not make financial changes unless resolution requires it
> **Proposed Action:**
> Issue refund

Authorization may approve issue_refund because it falls within standing authority; Alignment separately asks whether it fits the current task objective and constraints.

Where that evaluation occurs is an architectural choice: an authorization service, policy engine, agent runtime, gateway or workflow system could enforce it using equivalent machine-evaluable representations.

Representing Intent does not make it trustworthy. Its source, integrity and authority need to be established against an approved task, delegation or policy; an agent's unverified self-assertion should not determine the objective used to assess its own actions. Otherwise, prompt injection or goal hijacking could cause a compromised principal to present a fabricated benign objective. Establishing trusted Intent is distinct from evaluating whether a Proposed Action is aligned with it.

9. An Alignment Maturity Path

Alignment is not binary. The useful question is how explicitly Purpose and Intent can participate in technical decisions; PrincipalFrame uses the following illustrative path.

Stage	Alignment state	Characteristics
1 - Governance-led	Purpose is understood primarily by people	Policies, instructions, management oversight, approvals and retrospective accountability govern intended use
2 - Technology-assisted	Technology detects or constrains indicators of misuse	DLP, behavioural analytics, transaction controls, anomaly detection and workflow restrictions provide indirect evidence or boundaries
3 - Explicit	Purpose and Intent become machine-readable	Purpose, task objective, delegation and constraints are represented as structured information available to technical controls
4 - Runtime-enforced	Alignment uses machine-evaluable Purpose, Intent and Constraints in runtime decisions	Proposed Actions are evaluated against Authority, Purpose, Intent and Constraints before or during execution

This is an illustrative path, not an industry maturity standard.

At Stage 3, machine-readable means the control information is structured. At Stage 4, machine-evaluable means it can participate directly in a decision.

Stage 1 - Governance-led

Stage 1 relies primarily on governance: the organisation may state "Use this access only for payroll administration," while the access-control system does not understand that Purpose.

Stage 2 - Technology-assisted

Stage 2 adds technical assistance: DLP, behavioural analytics and transaction rules can detect or constrain misuse without necessarily representing the current Intent.

Stage 3 - Explicit

Stage 3 makes Purpose and Intent machine-readable.

> *purpose = customer_support*
> *task = explain_duplicate_charge*
> *delegated_by = customer*
> *constraints = read_only_until_remediation_approved*

Once represented explicitly, they can become policy inputs.

Stage 4 - Runtime-enforced

Stage 4 makes Alignment part of runtime execution: before or during an action, the system can evaluate both Authorization and consistency with Purpose, Intent and Constraints.

10. Emerging Work: The Control Plane Is Beginning to Change

Alignment is not a standardised identity-control category, but current standards and product work already address several of its ingredients.

Microsoft Entra Agent ID

Microsoft Entra Agent ID introduces dedicated agent identity constructs and governance semantics; technically, an agent identity is a special service principal created through an agent identity blueprint. [10]

Its distinction between technical owners and business sponsors makes the principal, its business purpose and accountable ownership more explicit - useful foundations for Alignment. [11]

IETF WIMSE and AIMS

AIMS is now a WIMSE working-group Internet-Draft (draft-ietf-wimse-aims-00). It addresses agent authentication, authorization, delegation, human involvement and tool-to-service access, and frames the objective as ensuring the right agent has access to the right resources and tools at the right time for the right reason. It also preserves user and system context for authorization decisions and audit. [12]

AIMS Section 10.1 describes an Agent Mission received from a user, system or another agent; translating missions into authorization requirements is outside its scope. Mission closely resembles Intent but does not determine whether each selected action remains aligned with that objective. [12]

RFC 9396 specifies fine-grained authorization requests; AuthZEN standardises PDP/PEP decision exchanges; and the Transaction Tokens draft addresses authorization-context propagation across service calls. None, alone, validates Intent or defines Alignment. [17-19]

Agent Operation Authorization

Agent Operation Authorization explores a structured authorization mechanism for specific agent operations, including operation proposals, policy-enforced constraints and runtime-verifiable authorization. The latest revision, draft-liu-agent-operation-authorization-02, expired on 17 September 2026; it is therefore exploratory work rather than IETF consensus. [13]

Decoupled Authorization Model for Agent2Agent

The Decoupled Authorization Model for Agent2Agent explored dynamic, intent-based and just-in-time authorization based on an agent's specific intent and behavioural trustworthiness. The individual Internet-Draft expired on 18 August 2026 and has no formal standing in the IETF standards process, so its value here is conceptual rather than normative. [14]

These efforts address different parts of the problem:

Emerging capability	Contribution
Agent identity	Establishes the principal
Delegation	Establishes where authority came from
Sponsors and governance	Establish accountable Purpose
Operation-specific authorization	Narrows authority toward the current action
Intent-aware approaches	Begin to represent the task objective or reason associated with an action

Together, they show the technical control plane beginning to absorb concepts that previously lived mainly in governance.

11. Scope and Limitations

The 4A model is an architectural lens, not a new authentication or authorization protocol. Six limitations matter.

First, Alignment is a PrincipalFrame control abstraction used here in a narrow identity-and-security sense; it is not AI alignment. PrincipalFrame does not claim to originate purpose-aware or ongoing access control: purpose-based access control and UCON address related concerns. [4,5]

Second, Alignment and Authorization need not be implemented by separate technologies. A sufficiently expressive authorization system can evaluate Purpose, Intent or equivalent context; the distinction is between control questions, not products.

Third, Human, Service, Workload and Agent are analytical archetypes, not mutually exclusive technical identity classes.

Fourth, the Alignment maturity path is conceptual and does not yet define measurable scoring thresholds.

Fifth, agent identity and authorization are evolving rapidly; some mechanisms discussed here remain draft or exploratory.

Sixth, while trustworthy Intent provenance and integrity are necessary for effective Alignment, this paper does not prescribe how they should be established or verified. This remains an open architectural and engineering question.

Conclusion

Alignment followed a different path from the other controls.

For humans, governance carried much of the Alignment burden; services and workloads constrained Purpose through implementation. Autonomous agents bring discretion back into the digital principal.

A principal can therefore be correctly authenticated, an action legitimately authorized and the event fully auditable - and the action can still be inconsistent with the objective being pursued.

The intent problem is not new.

> **What is new is the need to govern it as a technical control for digital principals.**

The 4A model makes that concern explicit alongside Authentication, Authorization and Accountability.

The transition is about where Alignment operates, not whether it exists.

> **Alignment shifts from People and Process, through software design and implicit constraints, toward explicit and enforceable Technology controls.**

As discretion moves into digital principals, approved Purpose, current Intent and applicable Constraints increasingly need machine-evaluable representations in the technical control plane.

References and Further Reading

Validated primary and original sources supporting the technical claims in this paper. Internet-Drafts are identified as work in progress; expired individual drafts are explicitly marked as such.

[1] National Institute of Standards and Technology, "Digital Identity Guidelines: Authentication and Authenticator Management," NIST SP 800-63B-4, July 2025. Source

[2] V. C. Hu, D. Ferraiolo, R. Kuhn, A. Schnitzer, K. Sandlin, R. Miller and K. Scarfone, "Guide to Attribute Based Access Control (ABAC) Definition and Considerations," NIST SP 800-162, updated August 2019. Source

[3] G. de Laat, G. Gross, L. Gommans, J. Vollbrecht and D. Spence, "Generic AAA Architecture," RFC 2903, Experimental, August 2000. Source

[4] N. Yang, H. Barringer and N. Zhang, "A Purpose-Based Access Control Model," Third International Symposium on Information Assurance and Security, 2007, doi:10.1109/IAS.2007.29. Source

[5] J. Park and R. Sandhu, "The UCONABC Usage Control Model," ACM Transactions on Information and System Security, vol. 7, no. 1, pp. 128-174, February 2004, doi:10.1145/984334.984339. Source

[6] M. Jones, A. Nadalin, B. Campbell, J. Bradley and C. Mortimore, "OAuth 2.0 Token Exchange," RFC 8693, Proposed Standard, January 2020. Source

[7] SPIFFE, "SPIFFE Identity and Verifiable Identity Document," current specification. Source

[8] A. Schwenkschuster and Y. Rosomakho, "Workload Identity Practices," draft-ietf-wimse-workload-identity-practices-07, September 2026, work in progress; submitted to the IESG for publication. Source

[9] J. Salowey, Y. Rosomakho and H. Tschofenig, "Workload Identity in a Multi System Environment (WIMSE) Architecture," draft-ietf-wimse-arch-08, July 2026, work in progress. Source

[10] Microsoft, "Agent identities in Microsoft Entra Agent ID," Microsoft Learn. Source

[11] Microsoft, "Administrative relationships in Microsoft Entra Agent ID (Owners, sponsors, and managers)," Microsoft Learn. Source

[12] IETF WIMSE Working Group, "AI Identity Management System," draft-ietf-wimse-aims-00, September 2026, work in progress. Source

[13] D. Liu, J. Zhu and S. Krishnan, "Agent Operation Authorization," draft-liu-agent-operation-authorization-02, March 2026; expired 17 September 2026. Source

[14] M. Chen and L. Su, "A Decoupled Authorization Model for Agent2Agent," draft-chen-agent-decoupled-authorization-model-00, February 2026; expired 18 August 2026 and archived as an individual Internet-Draft. Source

[15] OASIS, "Cross-Enterprise Security and Privacy Authorization (XSPA) Profile of SAML v2.0 for Healthcare," Version 1.0, OASIS Standard, November 2009. Source

[16] HL7 International, "PurposeOfUse," HL7 Terminology, published ValueSet. Source

[17] T. Lodderstedt, J. Richer and B. Campbell, "OAuth 2.0 Rich Authorization Requests," RFC 9396, May 2023. Source

[18] OpenID Foundation, "Authorization API 1.0," OpenID Final Specification, January 2026. Source

[19] A. Tulshibagwale, G. Fletcher and P. Kasselman, "Transaction Tokens," draft-ietf-oauth-transaction-tokens-11, July 2026, work in progress. Source

PrincipalFrame is an independent, non-commercial research and engineering project. Content is provided for research and educational purposes.

PUBLICATION INFORMATION

Version 1.0 · Author: Vaibhav Dhote · Publisher: PrincipalFrame · October 2026

Cite as: Dhote, V. (2026). From Authorization to Alignment: Governing purpose and intent across human, service, workload and agent principals. PrincipalFrame Research 001, Version 1.0. PrincipalFrame.

© 2026 Vaibhav Dhote. Unless otherwise noted, original research text and diagrams are licensed under Creative Commons Attribution 4.0 International (CC BY 4.0). PrincipalFrame names and logos and third-party material are excluded.

Licence: https://creativecommons.org/licenses/by/4.0/