European Business Wallets as the Legal Control Plane for Zero Trust AI Agents

An action and policy whitepaper on trusted AI and the role of business wallets for AI adoption in regulated industries

Dr. Carsten Stöcker
CEO, Spherity GmbH
carsten.stoecker@spherity.com

1 June 2026

Prepared for EU policy makers, the European Business Wallet workstream, the EU AI Office, cyber agencies, national digitalisation authorities, frontier AI platforms, and regulated industrial buyers

Document status: infrastructure-neutral academic and policy draft

Scope note. This paper is an architecture and policy analysis. It is not legal advice, certification advice, or assurance for a specific deployment. It uses current public drafts of the European Business Wallet framework and recent eIDAS 2.0 implementing acts as of June 2026. Where a legal text is still under legislative negotiation, the paper treats it as a draft policy signal, not as adopted law.

Abstract

Anthropic's recent paper, Zero Trust for AI Agents, raises the security baseline for enterprise agent deployments. Its maturity model moves from cryptographic agent identifiers and short-lived authentication artifacts to hardware-backed identity, attestation, continuous authorization, provenance, drift detection, and automated response. This is a significant contribution. It also exposes a structural gap. Technical Zero Trust controls can identify a workload, constrain a session, and record a tool call. They do not, by themselves, prove which legal person is accountable for an autonomous action, whether that legal person granted a valid mandate, whether the mandate covers the exact action, or whether the proof remains valid across borders and corporate domains.

This paper argues that the European Business Wallet (EBW), built on the eIDAS 2.0 trust framework and extended through the current European Commission EBW proposal, is the missing legal control plane for trusted industrial AI. The EBW provides the primitives required for enterprise-grade agentic systems: legal-person identity through European Business Wallet owner identification data (EBWOID), unique legal identifiers such as the European Unique Identifier (EUID), electronic attestations of attributes, wallet secure cryptographic devices, wallet unit attestations, digital mandates, automated machine-to-machine interaction, verifiable trust lists, revocation, and signed transaction evidence. These primitives map directly to the Advanced tier of agentic Zero Trust Architecture (ZTA).

The paper develops an infrastructure-neutral reference architecture for EBW-backed Zero Trust AI agents. It shows how EBW electronic attestations, mandates, status proofs, and cryptographic authorization artifacts can operate as a decentralized legal permission layer for Model Context Protocol (MCP) tool calls, Agent2Agent (A2A) interactions, Dataspace Protocol transactions, Tractus-X-style AI Service KITs, and conventional API gateways. The architecture is intentionally protocol-neutral and carrier-neutral: MCP, A2A, dataspace connectors, and API gateways provide interaction surfaces, while EBW supplies the legal-person identity, mandate, revocation, and non-repudiation plane required for regulated B2B and B2G automation.

The central claim is not that Europe will win the global AI race through larger models. The claim is narrower and more defensible. Europe has a structural opportunity to lead trusted industrial AI by binding autonomous agents to legal-person identity, verifiable authority, and regulator-grade evidence. Without that identity plane, B2B and B2G deployment of autonomous agents remains blocked by liability, compliance, audit, procurement, and corporate accountability risks that cannot be solved by model performance or runtime sandboxing alone.

Keywords: Zero Trust Architecture; AI agents; European Business Wallet; eIDAS 2.0; legal-person identity; EBWOID; EUID; electronic attestations of attributes; digital Power of Attorney; Agent2Agent Protocol; Model Context Protocol; Dataspace Protocol; Tractus-X AI Service KIT; Eclipse Data Space Connector; Verifiable Data Registry; AI Service Passport; format-neutral per-action authorization; post-quantum cryptography; secure multi-party computation; AI Act; industrial data spaces.

Contents

1 Introduction: the identity gap in autonomous enterprise AI

Autonomous AI agents change the unit of cybersecurity control. Traditional software executes pre-defined logic. Agentic systems interpret objectives, select tools, persist context, call APIs, compose workflows, and coordinate with other agents. This changes the security question from whether a human user may access a resource to whether a non-human actor may perform a business action under a particular legal and organisational authority.

Anthropic's Zero Trust paper correctly identifies the technical consequences. Static API keys, shared service accounts, network-location trust, session-level authorization, and friction-only controls fail when attackers and defenders operate at AI speed. The paper therefore pushes the maturity baseline toward cryptographic agent identity, short-lived authentication artifacts, hardware-bound authentication material, identity-based isolation, JIT/JEA privilege, comprehensive traceability, behavioural baselines, drift detection, and automated defensive operations [1].

The same shift creates a legal and institutional problem. A high-risk enterprise agent does not merely call a database. It may submit regulatory data, update a product passport, trigger a payment, order components, negotiate terms, access a regulated data space, run a pharmacovigilance workflow, instruct a connected machine, or exchange data with a public authority. In these contexts, the decisive question is not only "which workload called this API?" It is "which legal person is accountable, which mandate authorized the agent, which role or licence applies, which policy was active, and what verifiable evidence proves this after the fact?"

This paper treats that question as the core adoption barrier for B2B and B2G agentic AI. Frontier AI platforms can continue to improve model capability, tool orchestration, and runtime isolation. These advances are necessary. They do not resolve cross-border corporate attribution, agency, non-repudiation, revocation, or auditability. Without a verifiable legal identity plane, autonomous agents remain trapped in pilot settings or narrowly internal workflows because risk owners cannot assign liability with enough certainty.

The European Business Wallet addresses precisely this missing plane. It is not an AI product. It is a legal and cryptographic infrastructure for organisations. In the current Commission proposal, the EBW is a digital solution for economic operators and public sector bodies to receive, store, manage, combine, and present owner identification data and electronic attestations of attributes. The Council compromise draft further preserves the core ideas of EUID-based owner identification, issuance from authentic sources, common protocols, automatic interaction without manual intervention, mandate management, machine-readable directories, and verifiable presentation to relying parties [5], [6].

The architectural claim of this paper is direct: Anthropic's Advanced ZTA tier naturally requires an EBW-type trust fabric when actions cross corporate or public-sector boundaries. Agent identity must be bound to legal-person identity. Authorization must be evaluated at the action level, not only at session start. Delegation must be explicit, cryptographically attenuated, revocable, and auditable. Trust lists and revocation registries must be programmatic. Evidence chains must preserve provenance across multi-agent workflows. Runtime security must be linked to legal authority and AI governance evidence.

2 Analytical frame and source basis

The analysis combines five source families. First, it uses Anthropic's Zero Trust for AI Agents as the agentic ZTA baseline [1]. Second, it uses NIST ZTA concepts, including the movement from perimeter trust to resource-centric, identity-based, dynamic authorization [2], [3]. Third, it uses eIDAS 2.0 and implementing acts on wallet integrity, protocols, person identification data, electronic attestations of attributes, trust lists, and wallet certification [4], [7–10]. Fourth, it uses the current EBW proposal and Council working texts as the emerging legal-person wallet framework [5], [6].

Fifth, it uses the EU AI Act, NIST AI RMF, ISO/IEC 42001, W3C Verifiable Credentials, and NIST post-quantum cryptography standards as adjacent governance and cryptographic references [11–18].

The paper uses the term EBW as an infrastructure category. It does not depend on any vendor implementation. It also does not assume that the final EBW regulation will preserve every wording from current drafts. The argument is structural. If autonomous agents need cross-border legal identity, digital mandates, real-time attribute verification, revocation, and automated M2M proof exchange, then the resulting architecture is an EBW-type architecture even when implemented through different national providers, qualified trust service providers, or sectoral infrastructures.

2.1 Anthropic's tier model as the current agentic ZTA baseline

Anthropic's model is best read as a maturity model for agent security. Foundation is the minimum viable level. Enterprise is the practical target for large organisations. Advanced is the baseline for regulated, national-security, and high-impact deployments. The Advanced tier contains the controls that matter most for autonomous B2B action: hardware-backed identity, attestation, continuous authorization at each action, JIT/JEA privilege, full provenance chains, drift detection, real-time policy evaluation, and orchestrated response [1].

Table 1: Anthropic's tier model interpreted as an agentic Zero Trust maturity model and mapped to the legal trust gap.

ZTA tier	**Agent security intent**	**Relevant control features**	**Residual gap at corporate boundary**
Foundation	Remove trivial identity and portable-authority failure modes.	Cryptographic identifiers per agent instance; short-lived authentication artifacts; deny-by-default RBAC; identity-based isolation; basic action logging.	Technical identity exists, but legal-person attribution and corporate mandate are not yet proven.
Enterprise	Scale controls across real enterprise deployments.	X.509 or equivalent certificate lifecycle; mTLS; ABAC; sandboxing; immutable audit trails; distributed tracing; statistical anomaly detection.	Service identity and runtime evidence improve, but cross-company authority remains local and non-standard.
Advanced	Protect regulated and high-impact agentic workflows.	Hardware-backed identity; attested issuance; continuous authorization; JIT/JEA; full provenance; drift detection; SOAR integration; automatic revocation.	The controls require a legal and decentralised trust fabric when the agent acts for a company outside its own enterprise boundary.

A narrow reading of ZTA would treat Advanced controls as a stronger version of enterprise IAM. That reading is insufficient for agentic AI. The Advanced tier does not merely need more certificates. It needs an attribution chain from the legal person to the acting agent, and from the agent to the exact business action. Otherwise, the system can prove that a workload acted, but not that the workload was legally authorised to bind or represent a company.

2.2 Trusted AI as a deterministic governance architecture

Trusted AI, as used in this paper, is not an abstract ethics label. It is a deterministic governance architecture. Its purpose is to make AI-mediated actions verifiable, policy-bound, auditable, and attributable to accountable legal persons. The concept has five pillars.

This framing requires an important qualification. Large language models and many generative AI systems are probabilistic technologies; their outputs may vary across runs, contexts, prompts, model versions, sampling parameters, and retrieval states. Deterministic governance therefore cannot mean that the model layer itself always produces identical or fully predictable outputs. In regulated environments, the attainable assurance target is more precise: statistically meaningful and verifiable performance, safety, robustness, and drift thresholds, combined with deterministic external controls that decide whether an action may proceed. Predictable, reproducible, and explainable governance outcomes must be engineered around the probabilistic model through fixed policy predicates, cryptographic identity and mandate checks, versioned configurations, reproducible TEVV procedures, runtime monitoring, risk scoring, audit logging, and mandatory escalation when compliance or security thresholds are not met. This is a material technical challenge in regulated industries, safety-critical systems, and critical infrastructure, where unbounded agent discretion, weak explainability, or non-reproducible high-impact actions are incompatible with legal accountability and operational safety [1], [11–13], [20].

Table 2: Trusted AI pillars expressed as technical, policy, and audit controls.

Trusted AI pillar	**Control objective**	**ZTA implementation meaning**	**Policy and audit meaning**
Legal anchoring	Every regulated AI action is linked to a legal person, not only to a cloud account or model endpoint.	EBWOID and EUID identify the corporate principal; wallet unit attestation binds keys to a wallet unit.	Regulators, courts, customers, insurers, and auditors can identify the accountable organisation.
Delegated autonomy	An agent acts only within a narrow, revocable authority.	Digital mandate or Power of Attorney attestation defines actions, resources, purpose, limits, time, delegation depth, and escalation rules.	Corporate agency is machine-verifiable. Over-delegation and expired authority can be denied before action.
Lifecycle provenance	The AI service has a verifiable operating history.	AI Service Passport records model version, tool set, TEVV assertions, policy version, runtime controls, drift status, incidents, and changes.	Supports technical documentation, record keeping, due care, product-service accountability, and post-market monitoring.
Deterministic policy enforcement	Policy decisions are reproducible and explainable.	Policy engine evaluates identity, mandate, status proofs, context, risk, data class, and tool action before each invocation.	Governance moves from manual review to versioned, testable, auditable rules.
Audit-ready accountability	Every consequential action produces evidence that can be verified after the fact.	Signed authorization artifacts and action receipts include hashes, status checks, evidence chain, and policy decision.	Provides transparency, non-repudiation evidence, incident reconstruction, and regulator access without relying on informal logs.

The link between these pillars and ZTA is systematic. Legal anchoring strengthens identity. Delegated autonomy operationalises least agency. Lifecycle provenance extends traceability beyond single tool calls. Deterministic policy enforcement implements continuous authorization. Audit-ready accountability transforms logs into evidentiary records. The EBW is relevant because it supplies these controls in a form that can operate across organisational and Member State boundaries.

3 The European Business Wallet as a legal control plane

The EBW should be understood as a legal control plane for organisational identity and authority. It is not a replacement for IAM, PKI, access-management protocols, workload identity, confidential computing, or security monitoring. It supplies the cross-domain legal evidence that those systems lack. The EBW can say which organisation exists, which identifier binds to it, which attributes are verified from authentic sources, which representative or system has a mandate, which attestation or proof is current, and whether a relying party should accept a proof.

3.1 Legal-person identity: EBWOID, EUID, and authentic sources

European Business Wallet owner identification data (EBWOID) is the identity root for the wallet owner. In the Commission proposal and Council compromise, EBWOID contains at least the official name of the economic operator or public sector body as recorded in a relevant register or official record, and a unique identifier. Where an economic operator has an EUID, the EUID is to be used as that unique identifier. For entities without an EUID, the draft provides for a created unique identifier under implementing specifications [5], [6].

This design matters for AI agents because it separates legal identity from technical service identity. An X.509 certificate may authenticate a workload endpoint. It does not normally prove the current legal existence, register state, representative structure, sector role, or mandate of the corporate principal. EBWOID closes that gap by grounding the organisational identity in authentic sources and issuing it in a machine-verifiable form. EUID provides a cross-border company-law identifier where available. For public-sector bodies and economic operators outside normal company registers, equivalent unique identifiers are required to avoid identity fragmentation.

EBWOID is also a better anchor for corporate accountability than proprietary tenant identifiers. Tenant IDs are useful inside one platform. They are not authoritative in a multi-platform, multi-state, multi-sector economy. EBWOID is designed as portable legal-person identification data bound to an EBW unit. It can therefore support regulated data spaces, product passports, public procurement, KYC/KYB, corporate banking, tax interactions, and agent-to-agent commerce without forcing every relying party to re-build the same identity proofing process.

3.2 Automated M2M interaction and digital mandates

The EBW drafts are especially relevant for AI agents because they include automated interaction. The Commission proposal requires common protocols and interfaces that allow interaction with European Business Wallets automatically without manual intervention or through direct user action [5]. The Council compromise retains the same control objective and adds a European Digital Directory with a machine-readable API for automated system-to-system communication [6]. This is the normative bridge from human wallet use to machine-to-machine trust exchange.

The same drafts introduce mandate logic. An authorised representative is defined as a person or legal person acting on behalf of the EBW owner on the basis of an authorisation. A mandate is the authorisation granted by the EBW owner that enables that representative to act on behalf of the

owner in operating EBW functions [5], [6]. The Council text also preserves functions for authorising multiple users and managing or revoking authorisations. In policy terms, this is the digital Power of Attorney layer for organisations. In agentic AI terms, it is the legal form of least agency.

The policy importance is clear. An autonomous agent should never inherit full corporate authority because it can access a platform or because it was called by a high-privilege process. It should receive a specific mandate or digital Power of Attorney attestation. That mandate should state the legal principal, actor, permitted action, prohibited action, resource class, purpose, time window, delegation depth, geographic or legal domain, risk threshold, and revocation mechanism. This makes corporate agency verifiable at runtime.

3.3 Wallet secure cryptographic devices and wallet unit attestations

eIDAS 2.0 implementing acts and the EBW annex use wallet-specific security concepts that map directly to Advanced ZTA. Wallet solutions include wallet instances, wallet secure cryptographic applications, and wallet secure cryptographic devices. The secure cryptographic device is a tamper-resistant environment used to protect critical assets and execute cryptographic functions. Wallet unit attestations describe the wallet unit or allow authentication and validation of its components [7], [8].

The EBW annex requires wallet back-ends to use at least one wallet secure cryptographic application and wallet secure cryptographic device to manage critical assets. It also requires wallet unit attestations to contain public keys, with corresponding private keys protected by a wallet secure cryptographic device. It further requires revocation policies and publicly available validity status when wallet unit attestations are revoked [5].

This is not a narrow wallet design detail. It is the security root for enterprise agent authority. If an agent mandate is issued from or bound to an EBW, the relying party must know that the wallet unit is authentic, that the relevant keys are protected, that signatures are bound to valid wallet material, and that revoked wallet units are detected. Otherwise, agent authorization becomes another copyable bearer-authority system.

3.4 Verifiable Data Registries, trust lists, and revocation

A scalable EBW-backed agent architecture requires a trust registry layer. The term Verifiable Data Registry (VDR) is used here in a technology-neutral sense. A VDR may be a distributed ledger, a DID registry, a signed web registry, a national register interface, an eIDAS trust list, a status list, or a sectoral registry. Its function is to provide verifiable information about issuers, schemas, attestation and proof status, wallet providers, relying parties, public keys, service endpoints, and revocation. It need not be blockchain-based.

The eIDAS implementing act landscape already contains key registry functions. It includes wallet relying party registration, trusted lists, lists of certified wallets, verification of electronic attestations of attributes, qualified certificates, electronic signatures and seals, qualified time stamps, preservation services, and qualified electronic ledgers [7–10]. For agents, these must become programmatic. A policy engine must be able to determine, at action time, whether an issuer is trusted, whether a wallet unit is valid, whether a mandate was revoked, whether an attribute is current, and whether the relying party is entitled to request the data.

Table 3: EBW components as security properties for agentic ZTA.

EBW trust component	Security property	Agentic ZTA function	Failure if absent
EBWOID + unique identifier	Legal-person binding and cross-border uniqueness.	Identifies the corporate or public-sector principal behind the agent.	Actions are tied only to accounts, tenants, or certificates with weak legal meaning.
Electronic attestations of attributes	Machine-verifiable claims from trusted issuers or authentic sources.	Proves licences, roles, permits, market roles, compliance states, or representative authority.	Policy engines cannot distinguish authorised economic actors from self-declared actors.
Digital mandate / PoA attestation	Attenuated authority from legal principal to agent or representative.	Implements least agency and action-level delegation.	Agents inherit broad authority or rely on informal internal configuration.
Wallet secure cryptographic device	Protection of private keys and critical assets.	Prevents simple export of the legal-authority signing key.	Compromised hosts can exfiltrate authority keys.
Wallet unit attestation	Authenticity and validity of wallet unit components.	Allows relying parties to validate the wallet and bound keys.	Verifiers cannot distinguish genuine wallet proofs from cloned or invalid environments.
Trust list / VDR	Issuer, schema, relying party, status, and revocation discovery.	Automates cross-domain proof verification.	Every integration becomes bilateral and non-scalable.
Action receipt and audit log	Integrity, chronological evidence, non-repudiation support.	Records who authorised what action, under which evidence and policy.	Incident reconstruction relies on local logs without legal authority proof.

4 EBW-backed Zero Trust architecture for AI agents

The architecture is based on a separation of concerns. The model chooses or recommends actions. The runtime executes controlled workflows. The tool gateway enforces policy. The EBW trust plane provides legal identity and authority proofs. The AI governance plane supplies lifecycle, TEVV, risk, and drift evidence. The audit plane records signed action evidence. These planes must be integrated, but not collapsed into one opaque product stack.

4.1 Control planes and separation of concerns

Table 4: Reference control-plane architecture for EBW-backed Zero Trust AI agents.

Plane	Core function	Primary artifacts	Security boundary
Model reasoning plane	Plans steps, interprets user intent, selects candidate tools.	System prompts, tool schemas, retrieved context, model outputs.	Untrusted for authority. Reasoning may propose actions but cannot grant permission.
Runtime and workload plane	Executes agent code in containers, sandboxes, microVMs, or confidential environments.	Workload identity, mTLS certificate, attestation evidence, logs.	Protects execution and limits blast radius.

Table 4: Reference control-plane architecture for EBW-backed Zero Trust AI agents.

Plane	Core function	Primary artifacts	Security boundary
Tool / MCP gateway plane	Mediates each API or tool call.	Tool manifest, policy decision, parameter validation, authorization artifact, response.	Policy enforcement point. No tool call executes without permit.
EBW legal trust plane	Provides legal-person identity, mandates, attributes, trust lists, revocation.	EBWOID, mandate attestation, role attestation, wallet unit attestation, status proof.	Determines legal authority and cross-domain trust.
AI governance plane	Provides lifecycle and risk evidence.	AI Service Passport, TEVV assertions, drift attestation, incident status, model version.	Determines whether the AI service is within approved operating state.
Audit and evidence plane	Creates verifiable records after action.	Signed action receipts, timestamps, trace IDs, input/output hashes, status snapshots.	Supports audit, incident response, AI Act evidence, and legal non-repudiation.

The model reasoning plane is deliberately not the authority plane. Prompt instructions, hidden model reasoning, or internal model confidence cannot create legal authority. Authority is externalised into electronic attestations, mandates, status proofs, and deterministic policy decisions. This is the main architectural move. It converts the agent from an autonomous account with broad permission into a policy-bound delegate with proof-carrying actions.

4.2 MCP and API tool-level enforcement

The Model Context Protocol and comparable tool invocation frameworks are natural enforcement points. Tool descriptors, schemas, and server metadata expose capabilities to agents. That same interface can be abused through tool poisoning, rug-pull attacks, malicious descriptors, tool chaining, and confused deputy patterns. The correct response is not to treat MCP as a trust layer. MCP is a tool transport and interface layer. EBW-based attestations and proof objects should supply the legal and authorization layer around MCP.

For frontier AI platforms and enterprise AI stacks, the implementation pattern is straightforward. Every tool server or API gateway becomes a wallet-relying party or integrates a wallet verifier. The gateway validates EBWOID, mandate attestations, role or licence attestations, wallet unit attestations, AI Service Passport status, risk and drift attestations, revocation status, and proof-of-possession before executing the tool call. The policy decision is evaluated for the exact action, resource, parameters, purpose, and current risk state. The gateway then returns a signed action receipt.

Table 5: EBW-backed enforcement flow for MCP and comparable API/tool gateways.

Step	Operation	Required verification	Output
1	Agent selects a tool and proposes an action.	No authority is inferred from model output.	Candidate action vector.
2	Tool gateway builds an authorization challenge.	Challenge includes nonce, audience, resource, action, parameter digest, policy version, and requested evidence.	Proof request.

Table 5: EBW-backed enforcement flow for MCP and comparable API/tool gateways.

Step	Operation	Required verification	Output
3	Agent or agent-wallet presents the required evidence.	EBWOID, agent identity, mandate, role/licence, AISP, drift/risk, wallet unit attestation.	Verifiable presentation.
4	Verifier resolves trust and status.	Trust lists, VDRs, issuer status, revocation registries, wallet validity, relying-party registration.	Status proof bundle.
5	Policy engine evaluates context.	Legal principal, action, resource, purpose, data class, risk score, drift state, time, location, delegation depth.	Permit, deny, or escalate.
6	Gateway enforces decision.	Tool call executes only if permit is current and bound to the exact action.	Tool result or denial.
7	System records evidence.	Action receipt includes hashes, evidence references, policy result, timestamp, and trace identifiers.	Audit-ready record.

This pattern is platform-neutral. OpenAI, Anthropic, Google, Microsoft, Schwarz Digits, Mistral, and enterprise agent platforms can implement it as an SDK, gateway, sidecar, managed tool registry, or policy plug-in. The important point is not the packaging. The important point is that a tool call should be executed only when the caller presents legal identity, delegated authority, current risk state, and proof-of-possession for the exact action.

4.3 Protocol convergence: MCP, A2A, and dataspace connectors

The prior section treats MCP and API gateways as immediate enforcement points. That is necessary but incomplete. Industrial agentic systems will not use a single protocol. They will combine agent-to-tool protocols, agent-to-agent protocols, and dataspace protocols. The relevant architectural question is therefore not which protocol wins. The relevant question is where legal identity, delegated authority, revocation, and action evidence attach across heterogeneous protocol boundaries.

MCP standardizes how an AI application connects to tools, resources, prompts, and external systems. It is best read as an agent-to-tool and agent-to-context interface [23]. A2A standardizes peer interaction among independent or opaque agents. Its specification defines Agent Cards, tasks, messages, artifacts, security schemes, protocol bindings for JSON-RPC, gRPC, and HTTP+JSON, authorization-required task states, and authorization scoping requirements [21], [22]. The Dataspace Protocol standardizes cross-organisational data sharing through catalog publication, contract negotiation, agreement verification, and transfer-process governance [24], [25]. Tractus-X AI Service KITs then apply this dataspace logic to AI systems by describing AI services as assets connected to EDC connectors and by defining AI-specific asset descriptions and usage policies [26], [27]. This is material for procurement and assurance planning because production assurance must check both the protocol layer and the legal authority layer [28].

These layers are complementary. MCP governs structured tool use. A2A governs peer-agent collaboration and long-running task state. Dataspace Protocol governs cross-company discovery, contract negotiation, and usage control. Tractus-X-style AI Service KITs turn AI services into discoverable, policy-bearing dataspace assets. None of these layers alone proves that an autonomous action is legally attributable to a specific company, that a digital mandate is current, that the agent may act for a regulated role, or that the resulting action record has legal non-repudiation properties. EBW is therefore not a substitute protocol. It is the legal and cryptographic control plane that each protocol should call when a material action is proposed.

Table 6: Protocol convergence matrix for EBW-backed Trusted AI.

Interaction layer	**Representative protocol or component**	**What the protocol does**	**What it does not prove by itself**	**EBW attachment point**
Agent-to-tool / agent-to-context	MCP; API gateway; tool server	Exposes tools, resources, prompts, and structured interfaces to AI systems.	Does not establish legal-person attribution, corporate mandate, sector role, or revocation status.	Verifier checks EBWOID, mandate, role, risk/drift evidence, status, and proof-of-possession before tool execution.
Agent-to-agent	A2A	Publishes Agent Cards, manages tasks, messages, artifacts, security schemes, task state, and in-task authorization.	Does not by itself prove that the peer agent acts for a legal person or has authority for a regulated action.	Agent Card attestation links provider EBWOID, service endpoint, declared skills, AI Service Passport, status location, and policy obligations. In-task authorization is satisfied by a mandate or approval artifact.
Dataspace negotiation and transfer	Dataspace Protocol; EDC connector	Publishes catalog entries, negotiates contracts, verifies agreements, and governs transfer processes.	Does not by itself prove agent-level delegation or AI-system lifecycle assurance.	Connector policy engine consumes EBW evidence, mandate scope, market-role attributes, and AI Service Passport references.
Industrial AI service publication	Tractus-X-style AI Service KIT	Describes AI systems as dataspace assets and expresses AI-specific usage policy, including AI agents and RAG services.	Does not replace legal identity, revocation, or non-repudiation evidence.	AI service asset description references EBWOID, AISP, TEVV evidence, policy conditions, assurance level, and status endpoints.

Legal control plane	EBW; VDR; trust lists; revocation registry	Provides legal-person identity, mandates, attestations, status, and proof verification.	Does not execute tools, transport messages, or perform model inference.	Protocol-neutral verifier and policy decision point for MCP, A2A, dataspace, EDC, API, and future A2A protocols.

A2A is especially important because it moves the trust boundary from a tool invocation to a peer-agent relationship. The A2A Agent Card is a useful capability manifest, but a self-description is not yet legal authority. In regulated environments the Agent Card should be signed or otherwise integrity-protected and accompanied by an EBW-backed Agent Card attestation. The attestation should bind the service provider, legal-person identity, endpoint, supported interfaces, declared skills, AI Service Passport reference, security scheme, policy obligations, revocation location, and assurance level. A relying agent should not treat an unsigned or legally unanchored Agent Card as a sufficient basis for high-impact collaboration.

A2A in-task authorization also maps directly to EBW, but it should be described in carrier-neutral terms. The protocol can move a task into an authorization-required state and lets the client obtain authorization out of band or through negotiated extensions [21]. For low-risk cases this may be a protocol-native approval or ordinary enterprise access mechanism. For regulated B2B/B2G cases the authorization artifact should carry, or reference, EBW legal-person identity, a mandate or digital Power of Attorney, role or licence attributes, action-specific scope, proof-of-possession, and current status evidence. The evidence may be carried by any approved EBW-compatible presentation, certificate assertion, signed envelope, compact proof object, signed protocol metadata, or out-of-band verifier response. The security requirement is independent of the carrier: it must be audience-bound to the originating agent or relying gateway, encrypted where sensitive, linked to the task identifier, action vector, parameter digest, nonce, and freshness window, and verifiable against trust lists and revocation registries. This prevents authority forwarding through an A2A chain and reduces the confused-deputy risk described by frontier ZTA frameworks.

Dataspace Protocol and Tractus-X-style AI Service KITs add the industrial deployment plane. A company can publish an AI service as a dataspace asset, describe its expected inputs, outputs, model/service characteristics, legal constraints, and usage policy, and expose it through an EDC connector. EBW adds the missing legal authority plane: the service provider is identified through EBWOID, the consuming company proves its role and mandate, the AI Service Passport supplies lifecycle and TEVV evidence, and the connector policy engine evaluates both the dataspace usage policy and the EBW mandate before contract negotiation or transfer. This is the correct integration point for automotive, Manufacturing-X, energy, logistics, product-passport, pharma, and public-sector data spaces. The design remains neutral to whether the EBW evidence is transported in the A2A message, an MCP metadata envelope, an EDC control-plane message, an HTTP header, a signed body, or an out-of-band verifier call.

The practical design rule is therefore: transport and interaction protocols should remain lightweight and interoperable; legal authority and non-repudiation should remain external, verifiable, and revocable. EBW should not be embedded as a proprietary agent runtime. It should be exposed as verifier, issuer, wallet, trust-list, VDR, status, and policy services that A2A servers, MCP servers, EDC connectors, API gateways, and SOAR agents can invoke through standardized interfaces. This preserves protocol competition while making the legal control plane testable and portable.

4.4 Per-action authorization artifact

Session-level authorization is too coarse for autonomous agents. A session can contain many actions with different legal effects, data classifications, tool risks, and compliance obligations. Therefore, a

regulated agent action should carry, or be accompanied by, a per-action authorization artifact. The term artifact is used deliberately. It is neutral with respect to carrier type: credential, attestation, certificate, presentation, signed object, proof envelope, or future format. The following model is a semantic control checklist for standardization work, not a proposal for a new carrier format.

Table 7: Format-neutral semantic fields for per-action authorization of regulated AI-agent actions.

Authorization data group	**Required semantic content**	**Security purpose**	**Policy decision use**
Legal principal	EBWOID reference; EUID where available; issuer; trust-list reference; identity status.	Binds the action to an accountable legal person.	Reject if the legal person, issuer, or owner identification status is invalid.
Agent actor	Stable non-human identity; workload identity; wallet unit attestation reference; proof-key binding; runtime attestation.	Prevents shared identities and copyable authority.	Reject if the acting agent cannot prove possession of the bound key or runtime state.
Mandate and role	Mandate or digital Power of Attorney reference; parent mandate digest; delegating subject; delegation depth; role or licence attestation references.	Proves that the legal person delegated a bounded authority.	Reject if action exceeds mandate scope or role semantics.
Action vector	Operation; tool or service; method or interface; resource; parameter digest; data class; purpose; jurisdiction.	Binds authority to one concrete action.	Evaluate exact action, not a broad session category.
Temporal and anti-replay controls	Issuance time; not-before time; expiry or freshness rule; nonce; request and trace identifiers; maximum status age.	Limits replay, stale status use, and long-lived authority.	Reject expired, reused, stale, or incorrectly scoped artifacts.
AI governance state	AI Service Passport reference; model/service version digest; tool manifest digest; policy version; TEVV reference; risk and drift attestations.	Links authorization to AI governance and runtime compliance.	Escalate or deny when AI service state exceeds thresholds.
Evidence binding	Input digest; retrieved-context digest; output digest; decision digest; status snapshot digest.	Creates reconstructable audit evidence.	Support incident review, AI Act logging, and dispute analysis.

Proof and status	Signature, MAC, attestation proof, zero-knowledge proof, or equivalent proof; algorithm identifiers; trust-list and status references; verifier identity; decision proof.	Ensures integrity, authenticity, crypto agility, and non-repudiation evidence.	Accept only if proof, issuer trust, and status are valid at decision time.

Element	**Symbol**	**Required semantics**	**Binding rule**
Legal principal	L	Legal person identified through EBWOID, EUID where available, issuer trust, and current status.	L is bound to the mandate chain and to the relying-party decision.
Agent actor	a	Stable agent identity, workload or wallet binding, runtime attestation, and proof key.	a must prove possession of the bound key K_a.
Protocol event	P	MCP tool call, A2A task or message, dataspace negotiation or transfer, API call, or future protocol event.	The artifact is audience-bound to the verifier and event-bound to P.
Action vector	V	Operation, resource, purpose, parameter digest, data class, jurisdiction, and impact class.	Policy evaluates V exactly; no implicit session inheritance.
Mandate chain	$M = (m_0, \ldots, m_n)$	Representative authority, digital Power of Attorney, role or licence, delegation depth, and expiry/freshness rule.	$\forall i \in [0, n-1]$: $\mathrm{Scope}(m_{i+1}) \subseteq \mathrm{Scope}(m_i)$.
Context and status	C, S_r	Time, policy version, risk state, drift state, trust-list status, revocation status, and assurance level.	$\mathrm{Fresh}(S_r, t) \Leftrightarrow t - \mathrm{issuedAt}(S_r) \leq \Delta_{\max}$.
Cryptographic proof	π	Signature, MAC, attestation envelope, zero-knowledge proof, or equivalent proof-of-possession object.	π binds H(P $\|$ V $\|$ C $\|$ S_r $\|$ nonce) to K_a.

The core property is action binding. The authorization is not valid for a session, another tool, another resource, another purpose, or another parameter set. A proof object that cannot be replayed outside its named audience, freshness window, and parameter digest removes an entire class of privilege-forwarding attacks.

4.5 Evidence chaining and confused deputy mitigation

The confused deputy problem occurs when a higher-privilege actor is induced to use its authority for a lower-privilege actor or attacker. In multi-agent systems this problem is amplified. Agents routinely request help from other agents. A manager agent may ask a worker agent to act. A compromised worker may ask a manager to execute a harmful action. A tool may accept a request because it sees a high-privilege caller, without checking the original requester or the delegated purpose.

Copyable bearer artifacts and broad session grants do not solve this problem. They often make it worse because authority can be forwarded or reused outside the original context. The cryptographic mitigation requires proof-carrying delegation. Each delegated action must carry a signed or otherwise cryptographically verifiable chain from the legal principal to the acting agent. Each hop must attenuate authority. The effective permission set of a derived authorization artifact must be the intersection of the parent mandate, the child mandate, policy constraints, current risk state, and resource-specific rules. Delegation must never expand authority.

A secure chain therefore has the following form: EBWOID → representative authority → mandate to agent operator → mandate to agent instance → per-action authorization artifact → action receipt. Each link contains, or cryptographically references, the hash of the previous link, the permitted scope, expiry or freshness rule, audience, revocation endpoint, and proof-of-possession binding. A relying tool accepts the action only if the full chain validates, all status checks are current, and the final action vector is within the intersection of all upstream scopes.

Table 8: Cryptographic mitigation of confused-deputy and privilege-inheritance failures.

Attack path	**Weak design**	**EBW-backed cryptographic control**	**Residual risk**
Manager-to-worker delegation	High-privilege session grant or local trust between agents.	High-privilege agent must verify original requester, mandate chain, purpose, and requested action before deriving a new action artifact.	Compromised policy engine or issuer.
Worker-to-manager confused deputy	Manager trusts worker instruction without original-request evidence.	Manager verifies original legal principal, mandate chain, action vector, and proof-of-possession before acting.	Social or operational pressure to override controls.
Tool call forwarding	Copyable bearer artifact without audience, nonce, or parameter binding.	Authorization artifact includes audience, nonce, freshness, parameter digest, and proof-of-possession binding.	Implementation error in gateway validation.
Cross-agent authority relay	Authority artifacts readable and reusable by intermediate agents.	Sensitive artifacts are encrypted or selectively disclosed so only the intended verifier can use them.	Metadata leakage through protocol logs.
Revoked or stale mandate	Long-lived authorization artifact, stale local ACL, or cached trust decision.	Tool server validates status and revocation within the required freshness window before execution.	Registry availability or denial-of-service.

4.6 Evidence chaining for multi-agent workflows

Evidence chaining is the mechanism that allows multi-agent workflows without implicit trust. In a regulated workflow, one agent may gather evidence, another may evaluate it, another may draft a filing, and another may submit it. Each step should have a task-specific proof chain and a signed action receipt. The chain preserves provenance and authority without giving every sub-agent the full authority of the orchestrating process.

Table 9: Evidence chain for cross-company multi-agent workflows.

Evidence object or attestation	Issuer or controller	Subject	Function in the chain
Legal-person identity attestation	Competent authority or authorised source	Company or public body	Establishes legal principal.
Representative authority attestation	Company register, qualified issuer, or EBW authority process	Natural person or organisational role	Shows who may delegate authority.
Agent mandate / digital Power of Attorney	Legal person or authorised representative	Agent operator or agent instance	Delegates bounded authority.
Market role or licence attestation	Regulator, sector body, or authorised issuer	Legal person or service	Proves regulated role or licence.
AI Service Passport	AI provider, operator, auditor, or governance function	AI service or agent system	Records provenance, TEVV, version, controls, and residual risk.
Runtime or drift attestation	Monitoring service or security control	Agent runtime or AI service	Proves current operating state.
Per-action authorization artifact	Policy decision point or wallet-controlled actor	Specific action	Binds exact authority to exact action.
Action receipt	Policy enforcement point or relying service	Executed action	Creates non-repudiation and audit evidence.

5 AI Act alignment and the AI Service Passport pattern

The EU AI Act regulates AI systems through risk-based obligations. For high-risk AI systems, the key operational requirements include risk management, data governance, technical documentation, record keeping, transparency to deployers, human oversight, and accuracy, robustness, and cybersecurity [11]. These requirements are not identical to ZTA. They are complementary. ZTA controls whether an agent may act. AI governance controls whether the AI service is designed, documented, monitored, and operated within approved risk bounds.

5.1 Mapping AI Act obligations to EBW-backed controls

Table 10: Alignment between AI governance obligations and EBW-backed action evidence.

AI governance requirement	Agentic risk	EBW-backed Trusted AI control	Evidence object
Risk management	Agent actions may create health, safety, fundamental-rights, financial, or operational harm.	Risk attestations and policy thresholds are evaluated before each high-impact action.	Risk score attestation; policy decision record.

Table 10: Alignment between AI governance obligations and EBW-backed action evidence.

AI governance requirement	Agentic risk	EBW-backed Trusted AI control	Evidence object
Technical documentation	Models, tools, data flows, and limitations may be opaque.	AI Service Passport links service identity, intended purpose, model version, tool manifest, TEVV, and operating limits.	AISP attestation and versioned documentation hash.
Record keeping / logging	Logs may show technical events but not legal authority.	Action receipts bind legal identity, mandate, policy decision, hashes, and timestamps.	Signed action receipt and trace bundle.
Transparency to deployers	Deployers may not understand capabilities, constraints, or required oversight.	AISP and mandate disclose approved use, prohibited actions, escalation triggers, and residual risks.	Instructions-for-use attestation; deployer notice hash.
Human oversight	Fully autonomous action may exceed the intended mandate.	Mandates encode escalation thresholds and human approval attestations for high-risk actions.	Approval attestation and action receipt.
Accuracy, robustness, cybersecurity	Prompt injection, tool poisoning, drift, runtime compromise, and supply-chain attacks may affect decisions.	Runtime attestation, tool manifest signing, drift attestations, status checks, and revocation block unsafe actions.	Runtime health attestation; tool integrity attestation; drift attestation.

5.2 AI Service Passport as lifecycle evidence

An AI Service Passport is a proposed control artifact for trusted industrial AI. It is not a legal term in the AI Act. It is a practical pattern for lifecycle provenance. It records the identity, configuration, risk state, validation, monitoring, and operational history of an AI service in a verifiable format. It is analogous in function to a digital product passport, but focused on an AI service, its model supply chain, runtime configuration, TEVV evidence, and operational controls.

The AISP should include at least: provider and deployer legal identity; service identifier; intended purpose; high-risk classification state; model family and version or hash; tool and connector manifest; data-source classes; guardrails and limitations; TEVV methods and results; approval records; runtime controls; security posture; incident history; change log; drift baseline; current risk score; revocation or suspension status; and retention policy. Each assertion should be signed by an accountable issuer. The passport should be versioned, status-checkable, and linked to the authorization artifacts that rely on it.

The AISP is the missing bridge between enterprise AI governance and per-action ZTA. A policy engine cannot evaluate whether an agent should perform a regulated action unless it can determine whether the AI service is approved for that purpose, whether the relevant TEVV evidence exists, whether the current runtime matches the approved configuration, and whether a recent drift or incident has changed the risk state. The AISP converts those questions into machine-verifiable attributes.

5.3 Drift attestations and deterministic governance

Drift monitoring should not remain a passive dashboard. For agentic ZTA, behavioural and model drift must become authorization inputs. A short-lived drift attestation can state that an agent or AI service remains within an approved baseline for tool usage, output characteristics, error rates, data-access volume, and anomaly indicators. If drift exceeds policy thresholds, new authorization artifacts are not issued or are revoked. Tool gateways deny or escalate actions that require those artifacts.

This design enforces deterministic governance. The agent does not decide that its drift is acceptable. The monitoring system attests to the current state. The policy engine evaluates that attestation against a versioned policy. The tool gateway enforces the decision. The audit plane records the evidence. This is the same architectural pattern as financial transaction screening or AML risk scoring, adapted to AI-agent operations.

6 Post-quantum cryptography, sMPC, and cryptographic agility

EBW-backed agentic ZTA will create long-lived public-key evidence: legal identity attestations, mandates, role attestations, trust lists, wallet unit attestations, AI Service Passports, action receipts, timestamps, and audit records. Some evidence must remain verifiable for years. A conventional classical-only signature stack is therefore a strategic risk. Even if large-scale quantum computers are not available today, long-lived authenticity and confidentiality requirements require migration planning now.

NIST finalized its first post-quantum cryptography standards in 2024. FIPS 204 specifies ML-DSA, a lattice-based digital signature standard. FIPS 205 specifies SLH-DSA, a stateless hash-based digital signature standard that provides algorithmic diversity and a conservative backup to lattice signatures. FIPS 203 specifies ML-KEM for key establishment and encryption use cases [14–16].

6.1 PQC profile for EBW-backed agent systems

Table 11: Post-quantum cryptographic profile for EBW-backed AI-agent evidence.

Cryptographic object	Recommended transition profile	Reasoning	Operational constraint
High-volume action receipts	Hybrid classical + ML-DSA signatures during transition.	Action receipts are numerous and need efficient verification with quantum-resistant assurance.	Verifier libraries, signature sizes, and storage costs must be measured.
Long-lived trust lists and root attestations	Hybrid or SLH-DSA-backed signatures where size and performance permit.	Long-lived trust anchors need algorithm diversity and conservative security assumptions.	SLH-DSA signatures are larger and may not suit all high-volume paths.
Mandate and role attestations	Hybrid signatures and clear algorithm identifiers.	Mandates may remain legally relevant after expiry in audit or litigation.	Relying parties must support dual validation.
Transport security	Hybrid key establishment using ML-KEM where supported.	Protects against harvest-now-decrypt-later risks for confidential exchanges.	Requires TLS/library and compliance profile updates.

Table 11: Post-quantum cryptographic profile for EBW-backed AI-agent evidence.

Cryptographic object	Recommended transition profile	Reasoning	Operational constraint
Firmware, tool manifests, and runtime images	ML-DSA for frequent signing; SLH-DSA for highly conservative long-term artifacts.	Tool integrity underpins agent safety and confused deputy mitigation.	Supply-chain tooling must verify PQC signatures in CI/CD and runtime.
Archived audit bundles	Preservation profile with timestamp renewal, algorithm migration, and evidence re-sealing.	Audit evidence must survive algorithm deprecation.	Requires long-term validation policy and archive governance.

6.2 Secure multi-party computation and threshold key management

Hardware-backed keys remain central to Advanced ZTA. HSMs, TPMs, secure elements, confidential computing, and wallet secure cryptographic devices all reduce key-exfiltration risk. They also create operational constraints: hardware availability, certification cycles, cloud dependency, vendor lock-in, remote operation, disaster recovery, and migration latency. Secure multi-party computation (sMPC) and threshold cryptography provide a complementary pattern.

In threshold cryptography, a private key is split across multiple parties using secret sharing. The cryptographic operation, such as signing, is performed through a threshold protocol so that the private key does not need to be reconstructed. NIST describes this as a way to preserve key secrecy even if some parties are compromised, and to distribute trust across operators rather than depending on one critical key location [19].

For EBW-backed agents, threshold key management can support high-assurance signing by wallet back-ends, policy-gated signing, multi-operator approval, geographic distribution of shares, cloud independence, and rapid key rotation. It can also support cryptographic agility because key shares and signing policies can be migrated without treating one hardware key container as the only root of trust. However, this is not a claim that every sMPC design automatically satisfies eIDAS certification or qualified-signature requirements. The legal assurance depends on the concrete implementation, certification scheme, threat model, and regulatory acceptance.

6.3 Key rotation, revocation, and long-term validation

Agentic ZTA requires aggressive freshness windows for operational authorization, but legal evidence requires durable verifiability. These two requirements appear to conflict. The solution is to separate operational authority artifacts from evidence records. Operational artifacts should be short-lived and revocable. Evidence records should be immutable, timestamped, status-snapshotted, and preserved with long-term validation metadata. Key rotation should therefore preserve the chain of accountability while reducing future exposure.

A robust EBW agent profile should require key identifiers, algorithm identifiers, validity periods, revocation locations, trust-list references, status-list snapshots, timestamp renewal, and migration procedures. For high-risk sectors, policy should mandate dual signatures during PQC transition and require proof that archived evidence can be re-validated after algorithm deprecation.

7 The global race for trusted industrial AI and Europe's USP

The global AI race is usually framed around model capability, compute scale, data access, inference cost, agentic tool use, and runtime safety. These variables matter. Yet they do not solve the adoption problem in regulated industry. B2B and B2G organisations do not deploy autonomous agents into core processes only because a model is powerful. They deploy when liability, audit, access control, data protection, cyber resilience, procurement, insurance, and compliance can be controlled.

This is where Europe has a specific institutional advantage. eIDAS 2.0, trust services, European Digital Identity Wallets, and the proposed European Business Wallet create a legal identity and trust-service infrastructure that goes beyond Web PKI and proprietary platform identity. Web PKI authenticates domains and endpoints. Enterprise IAM authenticates users, workloads, and tenants. EBW can authenticate legal persons, attributes, mandates, and corporate authority across borders. That is the trust layer that autonomous B2B agents require.

The economic thesis is therefore not that Europe must out-train every global foundation model. The thesis is that Europe can define the trust fabric through which industrial AI is allowed to act. In pharma, energy, automotive, manufacturing, financial services, public administration, mobility, customs, and product-passport ecosystems, agents will need to prove identity, mandate, role, licence, purpose, counterparty trust, and current risk state. These requirements are not optional. They derive from regulation, contract law, sector oversight, safety obligations, and board accountability.

Without this verifiable identity plane, autonomous enterprise AI remains blocked at the boundary between experimentation and legally relevant action. A procurement agent cannot bind a company if its authority is unclear. A data-space agent cannot access restricted datasets if its market role is self-declared. A regulatory submission agent cannot sign or seal evidence without legal non-repudiation. An AI service consumer cannot trust a supplier agent without proof that the supplier, agent, mandate, and AI service state are genuine. Runtime sandboxing can reduce technical harm, but it cannot create corporate authority.

Table 12: Europe's USP in trusted industrial AI: legally binding trust infrastructure rather than model capability alone.

Global AI competition dimension	**Dominant current emphasis**	**Adoption blocker in regulated industry**	**European EBW-based differentiator**
Model capability	Reasoning, multimodality, tool use, code generation.	A capable model can still lack authority, provenance, and liability allocation.	Legal-person identity and mandate proofs govern when the model may act.
Runtime safety	Sandboxing, guardrails, prompt-injection defences, tool allow-lists.	Technical containment does not prove the corporate principal or legal authority.	EBW adds legal identity, verifiable delegation, and action evidence.
Platform identity	Tenant IDs, API keys, workload certificates, and platform access scopes.	Identity is often platform-local and hard to verify across companies or Member States.	EBWOID and EUID provide cross-domain legal-person identity.
Compliance operations	Manual review, checklists, internal approvals.	Manual controls do not scale to machine-speed agent-to-agent workflows.	M2M attestation exchange, trust lists, and revocation enable automated compliance gates.

Table 12: Europe's USP in trusted industrial AI: legally binding trust infrastructure rather than model capability alone.

Global AI competition dimension	Dominant current emphasis	Adoption blocker in regulated industry	European EBW-based differentiator
Cyber resilience	Detection, response, isolation, key management.	Security logs may lack legal authority and audit sufficiency.	Action receipts combine technical traceability with legal mandate proof.

8 Stakeholder-specific control requirements

The same architecture implies different responsibilities for different stakeholder groups. The following requirements are phrased as control objectives rather than product requirements. They are intended for public-sector drafting, cybersecurity certification, procurement, and platform engineering.

8.1 Frontier AI platforms, A2A providers, and dataspace tool providers

Frontier AI platforms should treat EBW verification as a protocol-neutral permission framework. A platform does not need to become a business wallet provider. It needs to integrate wallet verification, status checking, and policy enforcement at the tool, agent, and dataspace boundary. Every MCP tool invocation, A2A task transition, EDC connector access, or API call that can affect external data, money, regulated records, safety, production systems, or third-party rights should require proof of legal identity, mandate, role, risk state, and proof-of-possession.

A practical platform control profile includes signed tool descriptors, signed A2A Agent Cards, immutable AI service asset descriptions, tool-server identity, relying-party registration, EBW verifier APIs, policy decision points, policy enforcement points, action receipts, and exporter functions for audit evidence. Tool and A2A logs should carry request IDs, task IDs, trace IDs, evidence references, policy decision hashes, status snapshots, and revocation freshness. The model should not be able to bypass the tool gateway or A2A gateway by constructing direct network calls.

8.2 European Commission, EBW workstream, and EU AI Office

The EBW workstream should define an AI-agent profile. This profile should specify semantic evidence objects for agent identity, digital mandate, role and licence attributes, AI Service Passport, runtime attestation, drift attestation, per-action authorization, and action receipts. It should also specify M2M presentation flows, relying-party authentication, revocation freshness, privacy-preserving selective disclosure, and audit export without prescribing a single carrier format.

The EU AI Office should treat legal-person identity, delegated authority, and action-level logs as enabling controls for high-risk and regulated agentic AI. The AI Act requires governance evidence. EBW-backed action records can make that evidence verifiable. The AI Office should not need to operate the EBW. It should define how EBW evidence can support AI Act documentation, logging, post-market monitoring, incident reporting, and human oversight.

8.3 BSI, ENISA, and cyber authorities

Cyber authorities should treat EBW-backed agentic AI as a critical trust infrastructure topic. Certification profiles must cover wallet secure cryptographic devices, wallet secure cryptographic applications, wallet unit attestations, back-end key management, relying-party verification, trust-list integrity, revocation availability, PQC migration, secure MCP gateways, tool signing, runtime attestation, and evidence preservation. Testing must include confused deputy attacks, mandate revocation races, stale status proofs, tool manifest substitution, replay, cross-tenant context poisoning, and key compromise.

8.4 Industrial B2B and B2G buyers

Industrial buyers should make per-action authority proof a procurement baseline for agentic systems that leave internal experimentation. Minimum requirements should include unique agent identity, no shared service accounts, no static API keys, EBW-compatible legal-person identity, explicit agent mandates, action-level policy checks, signed tool manifests, revocation tests, audit export, risk and drift attestations, incident playbooks, and PQC transition plans. The procurement question should not be "does the model perform?" alone. It should be "can every material action be attributed, authorised, constrained, revoked, reconstructed, and defended?"

Table 13: Stakeholder-specific control requirements for EBW-backed agentic ZTA.

Stakeholder	Immediate control objective	Non-negotiable technical constraints	Deliverable or policy artifact
Frontier AI platforms, A2A providers, and MCP/tool providers	Turn tool, agent, and dataspace access into proof-carrying, action-level authorization.	No direct execution without gateway; no long-lived copyable bearer authority; signed Agent Cards and tool manifests; audience-bound attestations; revocation checking; task-level logs.	MCP/A2A/Dataspace EBW verifier profile and action receipt API.
EC EBW workstream	Define agent-compatible EBW profiles.	M2M presentation; mandate schemas; status and revocation semantics; policy-readable attributes.	EBW AI-agent profile and conformance tests.
EU AI Office	Map AI Act evidence to verifiable action records.	AISP profile; TEVV assertions; risk/drift attestations; record retention and incident evidence.	Guidance for high-risk agentic AI evidence.
BSI / ENISA	Certify and test the cyber trust fabric.	WSCD/WSCA assurance; trust-list availability; PQC roadmap; confused deputy and revocation race testing.	Security profile and evaluation methodology.
Industrial buyers	Procure accountable, auditable agentic systems.	No shared accounts; explicit mandates; per-action logs; revocation tests; incident export; least agency.	Procurement control baseline and acceptance tests.

9 Standardization agenda for EBW-backed agentic ZTA

The EBW proposal and eIDAS implementing acts provide the trust infrastructure. Agentic AI requires a dedicated profile on top. The profile should be narrow enough to implement, but strong enough to prevent proprietary fragmentation. The following agenda identifies the minimum standardization objects.

9.1 Evidence object and attestation schemas

The first task is semantic schema definition. At least ten evidence object or attestation types are needed: Agent Identity Attestation, Agent Mandate or digital Power of Attorney, Market Role or Licence Attestation, AI Service Passport, Runtime or Drift Attestation, MCP Tool Manifest, A2A Agent Card Attestation, Dataspace AI Service Asset Description, Per-Action Authorization Artifact, and Action Receipt. These objects should be composable and carrier-neutral. They should include issuer identity, subject identity, validity period or freshness rule, revocation location, proof method, algorithm identifier, schema version, privacy controls, audience binding, and policy semantics. They should support selective disclosure where the relying party does not need full organisational or model details.

9.2 Policy semantics and deterministic evaluation

Policy languages can vary, but policy semantics must be deterministic. A verifier must be able to decide whether an evidence chain permits an action without relying on model interpretation. The policy engine should evaluate identity, authority, action, resource, purpose, time, risk, data class, location, and revocation. Policies should be versioned, signed, tested, and linked to action receipts. They should support deny, permit, escalate, step-up, and quarantine decisions.

9.3 Revocation and status freshness

Revocation is a core security property, not an administrative afterthought. The profile should define freshness windows by risk class. Low-risk read-only actions may tolerate cached status for a short interval. High-risk write, payment, submission, or safety-related actions should require live or near-live status. A revoked mandate, suspended wallet unit, invalid role attestation, or failed drift attestation must immediately remove the ability to act within the relevant scope.

9.4 Privacy and data minimisation

EBW-backed agent authorization must not become a surveillance layer. Selective disclosure and unlinkability remain important. A relying tool usually needs to know that a company has a valid role, that an agent has a mandate, and that the action is within scope. It does not always need the full corporate register record, full ownership structure, or full AI Service Passport. The profile should separate verification of eligibility from unnecessary disclosure.

9.5 Minimum viable implementation path

A minimum viable EBW-backed agent implementation can start with internal B2B workflows and a limited evidence set: legal-person identity, agent identity, mandate, tool manifest, and action receipt. The next maturity step adds role attestations, VDR-based trust discovery, revocation, risk/drift attestations, AISP linkage, and signed A2A Agent Cards. The advanced step adds cross-company

A2A workflows, dataspace AI service publication through EDC connectors, PQC hybrid signatures, threshold key management, live policy federation, and sector-specific regulatory reporting.

Table 14: Maturity path for EBW-backed Zero Trust AI agents.

Maturity step	Evidence set	**Enforcement point**	**Evidence output**	**Use cases**
Step 1: Controlled pilot	Legal-person identity, agent identity, narrow mandate.	Single enterprise tool gateway.	Signed action receipt.	Internal process automation; read-only data access.
Step 2: Regulated enterprise	Add role/licence attestation, tool manifest, revocation, risk score.	MCP/API gateway with PDP/PEP.	Evidence chain and status snapshot.	Regulated data space, DPP updates, KYB, procurement.
Step 3: Cross-company A2A	Add counterparty wallet verification, AISP, drift attestation.	Mutual tool gateways and VDR trust discovery.	Bilateral action receipts.	Supplier agent interaction, compliance exchange, agentic commerce.
Step 4: High-assurance infrastructure	Add PQC hybrid signatures, threshold keys, long-term validation.	Certified wallet and security infrastructure.	Archive-grade evidence bundle.	Critical infrastructure, public-sector automation, safety-relevant workflows.

A Unresolved Structural Vulnerabilities in Frontier Agentic ZTA Frameworks

This appendix formalises the open loops left by frontier agentic ZTA frameworks when agents act across corporate and public-sector boundaries. The analysis is not a criticism of runtime Zero Trust. It shows where runtime Zero Trust ends and legal trust infrastructure begins.

A.1 Formal model

Let L denote a legal person. Let W denote an EBW unit controlled by, or validly representing, L. Let K_W denote a wallet-bound signing key protected by a wallet secure cryptographic device or an equivalent certified key-management environment. Let a denote an AI agent with key K_a. Let T denote a tool, API endpoint, A2A peer, or dataspace connector. Let V denote the action vector. Let $M = (m_0, \ldots, m_n)$ denote the mandate and delegation chain. Let R denote the relevant status and revocation registries. Let P denote a deterministic policy engine. Let G denote the AI-governance evidence set, such as an AI Service Passport, TEVV reference, risk score, and drift attestation. In the acceptance predicate, θ_r denotes the configured risk threshold and θ_d denotes the configured drift threshold.

Define the per-action authorization artifact α as a format-neutral proof object with the following abstract shape:

$$\alpha = \langle L, W, a, K_a, T, V, M, G, S_r, \pi_a, \pi_e \rangle.$$

The verifier accepts α only if the following predicate evaluates to true:

$$\begin{aligned}\mathrm{Accept}(\alpha) \Leftrightarrow\ & \mathrm{ValidLPID}(L) \wedge \mathrm{ValidWallet}(W) \wedge \mathrm{PoP}(a, K_a, \pi_a) \\ & \wedge \mathrm{ValidChain}(M) \wedge \mathrm{Scope}(V) \subseteq \bigcap\{\mathrm{Scope}(m) \mid m \in M\} \\ & \wedge \mathrm{Fresh}(S_r) \wedge \mathrm{Within}(G, \theta_r, \theta_d) \\ & \wedge \mathrm{Policy}(L, a, V, M, G, S_r) = \mathrm{permit}.\end{aligned}$$

An action receipt ρ is valid evidence only if it binds the accepted authorization artifact, the final action vector, the policy decision, the status snapshot, the timestamp, the verifier identity, and the trace identifier:

$$\rho = \mathrm{Sign}_e(H(\alpha) \parallel H(V) \parallel \mathrm{decision} \parallel H(S_r) \parallel t \parallel \mathrm{verifierID} \parallel \mathrm{traceID})\,.$$

The receipt therefore preserves a cryptographic link from legal person to wallet unit, mandate chain, agent, action, policy decision, and execution evidence. The representation can be serialized in different standards; the acceptance predicate is the invariant.

A.2 Vulnerability matrix and EBW closure

Table A1: Structural vulnerabilities in frontier agentic ZTA and EBW closure mechanisms.

Structural vulnerability	**Why frontier ZTA alone remains open**	**EBW closure mechanism**	**Resulting security property**
Cross-border legal attribution gap	Workload identity proves a technical subject, not the legal person accountable for the action.	EBWOID and EUID bind the action to a legal person using authentic-source identity data.	Attribution from action to legal principal.
Corporate authority ambiguity	Platform access scopes, local IAM roles, and service accounts do not prove board, representative, or organisational authority.	Digital mandate / PoA attestation proves the delegated authority and scope.	Machine-verifiable agency.
Session-to-action mismatch	A session artifact may authorize broad or changing actions after initial login.	Per-action authorization artifacts bind authority to action, resource, purpose, parameters, and expiry.	No trusted session; only authorised actions.
Confused deputy escalation	High-privilege agents may use their authority for low-privilege or malicious requesters.	Evidence chains preserve original requester, legal principal, delegated scope, and proof-of-possession.	Authority cannot be forwarded without attenuation and verification.
Over-delegation in multi-agent workflows	Manager agents can pass full context to sub-agents unless the framework enforces scope reduction.	Mandate derivation uses scope intersection and explicit delegation depth.	Least agency by construction.
Stale authority and revocation latency	Local caches and long-lived authorization artifacts continue after role or mandate revocation.	Trust lists, VDRs, and revocation registries are checked at action time with risk-based freshness.	Revoked authority loses effect before action.

Table A1: Structural vulnerabilities in frontier agentic ZTA and EBW closure mechanisms.

Structural vulnerability	**Why frontier ZTA alone remains open**	**EBW closure mechanism**	**Resulting security property**
Unverifiable counterparty	An agent can interact with a spoofed supplier, rogue tool server, or unregistered relying party.	Relying-party certificates, wallet unit attestations, and digital directory APIs authenticate counterparties.	Mutual A2A trust.
Runtime state disconnected from authority	An agent may keep valid attestations after drift, compromise, or failed attestation.	Risk, drift, runtime, and AISP status attestations become authorization preconditions.	Operational state controls legal authority.
Weak audit and non-repudiation	Logs may be mutable, incomplete, or missing legal authority evidence.	Action receipts bind evidence chain, status snapshot, policy decision, timestamp, and hashes.	Regulator-grade evidence and dispute handling.
Post-quantum fragility	Long-lived attestations and receipts may rely on classical signatures only.	Hybrid PQC signatures, ML-DSA, SLH-DSA for selected roots, and algorithm migration policy.	Long-term authenticity resilience.

A.3 Proof sketches

Attribution lemma. If legal-person identity, wallet-unit validity, mandate-chain validity, and agent proof-of-possession all hold, then the request is attributable to the legal principal for the bounded action, subject to current status and policy checks:

$$\begin{aligned}&\mathrm{ValidLPID}(L) \wedge \mathrm{ValidWallet}(W) \wedge \mathrm{ValidChain}(M : L \to a) \wedge \mathrm{PoP}(a, K_a, \pi_a)\\ &\quad \wedge \mathrm{Accept}(\alpha) \Rightarrow \mathrm{Attributable}(\mathrm{action}, L).\end{aligned}$$

Least-agency lemma. A derived mandate is valid only if it attenuates authority. The effective authority is the intersection of upstream authority, derived authority, policy context, and current status:

$$\forall i \in [0, n-1] : \mathrm{Scope}(m_{i+1}) \subseteq \mathrm{Scope}(m_i).$$

$$\mathrm{EffectiveScope} = \mathrm{Scope}(M_1) \cap \mathrm{Scope}(M_2) \cap \mathrm{PolicyContext} \cap \mathrm{CurrentStatus}\,.$$

Anti-confused-deputy lemma. A confused-deputy attack requires a privileged actor to perform an action without binding original requester, purpose, audience, and parameters. In the EBW profile, acceptance requires all of those bindings:

$$\begin{aligned}\mathrm{Accept}(\alpha) \Rightarrow\ &\mathrm{requester}(\alpha) \in M \wedge \mathrm{audience}(\alpha) = T \wedge \mathrm{purpose}(\alpha) = \mathrm{purpose}(V)\\ &\wedge H(\mathrm{params}(\alpha)) = H(\mathrm{paramsSubmitted}) \wedge \mathrm{nonce}(\alpha) \notin \mathrm{Used}(T).\end{aligned}$$

A malicious lower-privilege agent must therefore forge a mandate, compromise a proof key, or corrupt the policy engine; prompt manipulation alone is not sufficient.

Revocation lemma. If every high-risk action requires a fresh status proof for wallet unit, mandate, role attestation, and AI governance state, then the maximum silent-use interval for revoked authority is bounded by the status freshness window:

$$\mathrm{Fresh}(S_r, t) \Leftrightarrow t - \mathrm{issuedAt}(S_r) \leq \Delta_{\max}.$$

$$\text{Revoked}(x, t_0) \wedge t \geq t_0 + \Delta_{\max} \Rightarrow \neg\, \text{Accept}(\alpha).$$

Non-repudiation evidence lemma. A signed action receipt that binds the evidence chain, action vector, parameter digest, policy decision, status snapshot, timestamp, verifier identity, and trace identifiers gives materially stronger evidence than a local log:

$$\rho = \text{Sign}_e(H(\alpha) \parallel H(V) \parallel H(\text{params}) \parallel \text{decision} \parallel H(S_r) \parallel t \parallel \text{verifierID} \parallel \text{traceID})\,.$$

Legal non-repudiation remains a matter of the applicable legal regime, but the cryptographic record gives the technical evidence needed to test whether the action was authorized, current, and executed under a valid mandate.

Post-quantum continuity lemma. For long-lived identity evidence, attestations, and receipts, a hybrid classical/PQC proof profile improves migration continuity when it records algorithm identifiers, timestamps, status snapshots, and re-sealing procedures:

$$\pi = \pi_1 \parallel \pi_2, \quad \pi_2 \in \{\text{ML-DSA}, \text{SLH-DSA}\}.$$

Here π_1 is the classical proof. The system remains dependent on correct implementation, preservation services, and trust-list migration. The formal benefit is not absolute future-proofing; it is reduced dependence on a single classical signature assumption.

A.4 Protocol composition corollary: MCP, A2A, and Dataspace Protocol

Let $\mathcal{P}$ denote the set of interaction protocol event classes used by an agentic system, including MCP, A2A, Dataspace Protocol, and API events. Each protocol can carry requests, responses, task state, tool calls, or contract events. None of these transport or coordination properties is equivalent to legal authority.

$$\forall p \in \mathcal{P} : \text{SecureInteraction}(p) \neq \text{LegalAuthority}(p).$$

$$\text{MaterialAction}(p, V) \Rightarrow \text{VerifyEBW}(L, a, M, G, S_r, V) = \text{permit}.$$

The protocol composition corollary is therefore simple: MCP, A2A, Dataspace Protocol, and future A2A protocols should remain interaction layers. EBW supplies the external legal and cryptographic authority predicate that must evaluate before material action.

The closure property is compositional. If EBW evidence is bound to an MCP event, then a tool call is constrained by legal mandate and action policy. If EBW evidence is bound to an A2A event, then a peer-agent task cannot silently inherit or forward authority outside the mandate chain. If EBW evidence is bound to a dataspace event, then a connector can negotiate and enforce usage policies only for legally identified participants and authorised AI services. If the same action receipt records all three bindings, cross-protocol workflows remain auditable even when the runtime stack changes from MCP to A2A to EDC and back.

A.5 Boundary conditions

An EBW architecture does not solve every AI safety problem. It does not prove that a model output is true. It does not prevent all prompt injection. It does not replace sandboxing, data-loss prevention, secure software supply-chain controls, incident response, or human oversight. It closes a different class of gaps: legal identity, authority, delegation, revocation, cross-domain attribution, and evidentiary auditability. These are precisely the gaps that block autonomous agents from moving from pilots into accountable B2B and B2G operations.

References

[1] Anthropic. Zero Trust for AI Agents: A security framework for deploying autonomous AI agents in the enterprise. 2026. User-provided source document.

[2] NIST. Special Publication 800-207: Zero Trust Architecture. National Institute of Standards and Technology, 2020. https://csrc.nist.gov/pubs/sp/800/207/final

[3] NIST. Special Publication 800-207A: A Zero Trust Architecture Model for Access Control in Cloud-Native Applications in Multi-Location Environments. National Institute of Standards and Technology, 2023. https://csrc.nist.gov/pubs/sp/800/207/a/final

[4] Regulation (EU) 2024/1183 of the European Parliament and of the Council amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework.

[5] European Commission. COM(2025) 838 final. Proposal for a Regulation of the European Parliament and of the Council on the establishment of European Business Wallets, 19 November 2025.

[6] Council of the European Union. ST 7659/26 INIT. Proposal for a Regulation on the establishment of European Business Wallets. Presidency compromise text, 2026.

[7] Commission Implementing Regulation (EU) 2024/2979 of 28 November 2024 on the integrity and core functionalities of European Digital Identity Wallets.

[8] Commission Implementing Regulation (EU) 2024/2982 of 28 November 2024 on protocols and interfaces to be supported by the European Digital Identity Framework.

[9] Commission Implementing Regulation (EU) 2024/2977 of 28 November 2024 on person identification data and electronic attestations of attributes issued to European Digital Identity Wallets.

[10] Commission Implementing Regulation (EU) 2025/1569 of 29 July 2025 on qualified electronic attestations of attributes and electronic attestations of attributes provided by or on behalf of a public sector body responsible for an authentic source.

[11] Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence, 13 June 2024.

[12] NIST. Artificial Intelligence Risk Management Framework (AI RMF 1.0). NIST AI 100-1, 2023. https://www.nist.gov/itl/ai-risk-management-framework

[13] ISO/IEC 42001:2023. Information technology - Artificial intelligence - Management system. International Organization for Standardization, 2023.

[14] NIST. FIPS 204: Module-Lattice-Based Digital Signature Standard. National Institute of Standards and Technology, 2024. https://csrc.nist.gov/pubs/fips/204/final

[15] NIST. FIPS 205: Stateless Hash-Based Digital Signature Standard. National Institute of Standards and Technology, 2024. https://csrc.nist.gov/pubs/fips/205/final

[16] NIST. NIST Releases First 3 Finalized Post-Quantum Encryption Standards. 13 August 2024; updated 29 August 2025. https://www.nist.gov/news-events/news/2024/08/nist-releases-first-3-finalized-post-quantum-encryption-standards

[17] W3C. Verifiable Credentials Data Model v2.0. W3C Recommendation, 15 May 2025. https://www.w3.org/TR/vc-data-model-2.0/

[18] W3C. Decentralized Identifiers (DIDs) v1.0. W3C Recommendation, 19 July 2022. https://www.w3.org/TR/did-core/

[19] NIST. Multi-Party Threshold Cryptography Project. National Institute of Standards and Technology. https://csrc.nist.gov/projects/threshold-cryptography

[20] Author-supplied Trusted AI concept materials. From pilots to AI-industry trust infrastructure. 2025. User-provided source deck.

[21] Agent2Agent Protocol. Agent2Agent (A2A) Protocol Specification, Version 1.0. Linux Foundation project contributed by Google. https://a2a-protocol.org/latest/specification/

[22] Agent2Agent Protocol. A2A and MCP: Detailed Comparison. https://a2a-protocol.org/latest/topics/a2a-and-mcp/

[23] Model Context Protocol. Specification and introduction to tools, resources, and prompts. https://modelcontextprotocol.io/specification/2025-06-18

[24] Eclipse Dataspace Protocol. Dataspace Protocol 2025-1. https://eclipse-dataspace-protocol-base.github.io/DataspaceProtocol/2025-1/

[25] International Data Spaces Association. Dataspace Protocol overview and standardisation path. https://internationaldataspaces.org/offers/dataspace-protocol/

[26] Eclipse Tractus-X. AI Service KIT Adoption View, Release 26.03. https://eclipse-tractusx.github.io/docs-kits/kits/ai-service-kit/adoption-view/

[27] Eclipse Tractus-X. Release 26.03 changelog and AI Service KIT 0.0.1 listing, 18 March 2026. https://eclipse-tractusx.github.io/blog-changelog/release-26-03/

[28] Eclipse Tractus-X SIG-Release. AI Service KIT reference implementation issue #1597, 2026. https://github.com/eclipse-tractusx/sig-release/issues/1597