Yepai E-commerce Digital Employee Standard White Paper

Definitions, capability levels, evaluation methods, governance controls, and implementation guidance for e-commerce digital employees

Standard ID: YEP-DE-EC-2026-001
Version: v1.1
Publication date: 2026-06
Issued by: Yepai Team
Document type: Industry standard white paper
Intended audience: e-commerce executives, digital transformation leaders, customer service, operations, marketing and after-sales leaders, system architects, risk and compliance teams, AI product and delivery teams, research organizations, and ecosystem partners.

Release Statement

Copyright Notice

Use Statement

This standard defines the terminology, role model, reference architecture, capability requirements, autonomy levels, evaluation methods, governance controls, conformance classes, and implementation requirements for e-commerce digital employees. It applies to enterprise practice, procurement evaluation, partner review, project acceptance, supplier alignment, delivery acceptance, and risk governance.

Disclaimer

This document defines an industry standard framework and implementation baseline. It does not replace applicable laws, regulatory obligations, mandatory audit procedures, or independently issued third-party

certification. Metric thresholds in this document are recommended baselines or illustrative examples. Enterprises shall calibrate them according to business scale, marketplace rules, customer segments, regulatory obligations, risk appetite, and system maturity.

Document Attributes

Attribute	Description
Document name	Yepai E-commerce Digital Employee Standard White Paper
Standard ID	YEP-DE-EC-2026-001
Document type	Industry standard white paper
Scope	Highly digitized e-commerce workflows, information interactions, decision support, cross-system data execution, and human-AI operations
Out of scope	Heavy physical-world control, highly subjective judgment workflows without SOPs, and high-risk scenarios without auditability or accountable fallback owners
Recommended uses	Procurement evaluation, supplier review, system acceptance, governance baseline, role modeling, capability assessment, and conformance statement
Standard object	Role-level agentic systems for e-commerce, not a single model, interface, virtual avatar, or point automation tool

Normative Language

This document uses the following normative language:

Term	Meaning
Shall	Mandatory baseline requirement for the relevant standard level or production scenario
Shall not	Explicitly prohibited behavior or configuration
Should	Recommended requirement; if not met, an equivalent compensating control and risk rationale should be documented
Recommended	A practice that may be calibrated according to business scale, risk appetite, and maturity
May	Permitted practice, not a mandatory requirement

In this document, "digital employee" and "e-commerce digital human" refer to the same standard object: a role-based agentic software labor system that may be authorized to perform e-commerce business responsibilities. A virtual avatar, voice interface, video presenter, or human-like front-end may serve as an interaction layer, but it is not sufficient to qualify a system as a digital employee.

Table of Contents

List of Figures and Tables

Executive Summary

E-commerce is moving from a model in which humans operate systems and AI provides assistance to a model in which AI systems are authorized to perform portions of role-based work. This shift is driven by two simultaneous forces. First, AI adoption has become mainstream: the Stanford AI Index 2026 reports that organizational AI adoption reached 88% in 2025, while generative AI was used in at least one business function by 70% of organizations [R15]. Second, delegation remains early: the same source reports that AI agent deployment was still in the single digits across nearly all business functions [R15]. The gap between adoption and delegation is the problem this standard addresses.

Agentic commerce is increasingly described as a shopping model in which AI agents act on behalf of consumers to discover, compare, negotiate, and execute transactions. McKinsey estimates that agentic commerce could represent a global opportunity of USD 3 trillion to USD 5 trillion by 2030, with nearly USD 1 trillion of US B2C retail revenue potentially orchestrated through agentic commerce [W1]. Bain also projects that AI agents may influence or conduct 15% to 25% of US e-commerce sales by 2030 [W2]. These shifts mean that merchants need more than better chatbots. They need e-commerce digital employees that can be defined, authorized, evaluated, audited, governed, and continuously improved.

This standard defines an e-commerce digital employee as follows:

> **An e-commerce digital employee is a software labor system with explicit e-commerce role responsibilities, KPI/SLA obligations, authorization boundaries, and governance constraints. It can perceive business events, interpret rules, make decisions, invoke tools, and complete cross-system execution within an e-commerce environment, while remaining continuously measurable, auditable, interruptible, and improvable.**

This definition has five necessary conditions:

1. **Role-based**: The system shall correspond to a clearly defined e-commerce role or responsibility domain, such as customer service, after-sales and refunds, product operations, marketing operations, content production, or fulfillment coordination.
2. **Authorizable**: The system shall have explicit read/write permissions, monetary thresholds, marketplace action boundaries, approval conditions, and revocation mechanisms.
3. **Closed-loop capable**: The system shall be able to invoke APIs, RPA, GUI automation, or protocol interfaces within its authorization boundary. Recommendation generation alone is not sufficient.
4. **Evaluable**: The system shall prove its capability through offline tests, sandbox tests, shadow runs, gray releases, and production evaluations. A demo is not sufficient evidence.
5. **Governable**: The system shall include data minimization, permission isolation, audit trails, human oversight, emergency shutdown, version rollback, and consumer protection controls.

The following external industry figures establish the adoption, capability, risk, and economic background for this standard. These figures shall not be used as conformance thresholds.

Signal	Verified figure or source position	Implication for this standard
Enterprise AI adoption	Organizational AI adoption reached 88% in 2025; generative AI was used in at least one business function by 70% of organizations [R15]	Procurement language must distinguish general AI adoption from authorized digital labor
Agent deployment maturity	AI agent deployment remained in the single digits across nearly all business functions [R15]	The standard should define controlled progression from assistance to delegated execution
Capability trajectory	OSWorld task success rose from roughly 12% to 66.3%, still leaving about one in three structured attempts unresolved [R15]	Authorization levels must be tied to evidence, not benchmark enthusiasm
Productivity pattern	AI productivity gains are largest in structured, measurable work, with reported gains in customer support, software development, and marketing output [R15]	E-commerce pilots should start with high-frequency, rule-governed, monitorable workflows
Risk signal	Documented AI incidents rose from 233 in 2024 to 362 in 2025 [R15]	Audit, shutdown, rollback, and accountable ownership are release prerequisites
Economic potential	McKinsey estimates generative AI could add USD 2.6 trillion to USD 4.4 trillion annually across analyzed use cases, with about 75% of value concentrated in customer operations, marketing and sales, software engineering, and R&D [R16]	E-commerce standards should focus on operating roles, not only front-end interaction

This standard establishes an eight-layer framework:

Layer	Problem addressed	Standard output
Definition	Aligns terminology across digital employees, AI agents, RPA, copilots, and digital humans	Terms, conformance criteria, scope
Role	Links capability to e-commerce responsibility and accountability	Role description, SOP, KPI/SLA, authorization boundary
Architecture	Defines the system as a multi-component engineering system, not a single model	Intake, perception, knowledge, policy, decisioning, tools, audit, oversight
Capability	Determines whether the system is fit for the role	Six capability standards and metrics
Level	Determines the degree of autonomy that may be authorized	L1-L5 levels and human roles

Layer	Problem addressed	Standard output
Evaluation	Proves capability, stability, and reproducibility	Five-stage evaluation and runtime metrics
Governance	Controls financial loss, compliance, experience, and brand risk	Data, permission, audit, shutdown, oversight, rollback
Implementation and evolution	Guides adoption from pilot to scaled operations	90-day path, maturity model, continuous improvement

Core principle: stronger capability requires stronger governance; broader authorization requires stronger evidence; digital employees shall convert high-frequency, rule-governed, verifiable e-commerce work into governable digital labor.

---

1. Industry Background and Problem Definition

1.1 E-commerce Is Entering Agent-Mediated Commerce

Traditional e-commerce is designed around human browsing, comparison, and purchasing. Search, recommendation, advertising, product detail pages, carts, payment, and after-sales processes are all built around human clicks and human decisions. Agentic commerce changes this assumption: more shopping tasks will be delegated to AI agents acting on behalf of consumers. McKinsey defines agentic commerce as commerce in which AI agents shop on behalf of people and can anticipate needs, navigate options, negotiate, and execute transactions [W1]. Visa describes AI agents as autonomous systems that process information, reason, take action, and use tools or APIs in an observe-decide-act loop, and reports that 47% of consumers are interested in using agents for commerce [R8].

This shift affects both the buyer side and the seller side:

- Buyer agents will reshape product discovery, comparison, conversion, and repeat purchase.
- Merchants will need structured product data, real-time inventory, pricing, fulfillment policies, returns policies, and trusted payment capabilities.
- Platforms will need to manage agent access, identity, authorization, preferences, ad fairness, and anti-abuse controls.
- Payment and risk systems will need to recognize authorized agent behavior rather than only human behavior.

An e-commerce digital employee is therefore not a simple transfer of general office agents into commerce. It is part of the merchant-side operating infrastructure for the agentic commerce era.

1.2 The Adoption-to-Delegation Gap

The near-term challenge is not whether enterprises will use AI. The challenge is whether AI systems can be delegated role-level work with explicit boundaries, measurable performance, and accountable supervision. The AI Index 2026 reports 88% organizational AI adoption in 2025 and 70% use of generative AI in at least one business function, while AI agent deployment remains in the single digits across nearly all business functions [R15]. This means most organizations have adopted AI capabilities without yet converting those capabilities into governed digital labor.

In e-commerce operations, this gap appears in a familiar pattern:

1. A customer-service bot answers questions, but a human still checks orders, logistics, refunds, and exceptions.
2. A content model drafts titles or scripts, but a human still maps marketplace rules, uploads assets, and handles rejection.
3. A marketing assistant generates copy or keyword ideas, but a human still interprets spend, risk, promotion rules, and budget authority.
4. An RPA workflow updates fields, but a human still diagnoses ambiguous business context and decides when to stop.

These tools improve local productivity, but they do not by themselves create a role-level operating unit. A digital employee standard is needed because delegation is not a feature. Delegation is an operating contract: role, authority, metrics, evidence, oversight, and rollback.

1.3 Why E-commerce Needs Role-Based Digital Employees

E-commerce is a natural environment for digital employees, but it also demands standardization.

E-commerce is suitable because:

1. **High frequency and repetition**: Customer inquiries, order lookups, logistics tracking, returns, product listing, price monitoring, and content rewriting are frequent and rule-intensive.
2. **Digitized data**: Orders, inventory, products, users, reviews, ads, logistics, and after-sales records are already digitized.
3. **Cross-system workflows**: A single task may span IM, OMS, ERP, WMS, PIM, CRM, ad platforms, payment systems, and logistics systems.
4. **Fast feedback loops**: Response time, conversion, refund turnaround, complaints, and financial loss can be measured within short cycles.
5. **Global operating pressure**: Cross-border commerce must handle multiple languages, time zones, currencies, marketplace rules, and long-tail after-sales issues.

Standardization is required because:

1. **Concept inflation**: Chatbots, digital human videos, RPA scripts, copilots, and AI plug-ins are often all marketed as "digital employees," making procurement evaluation unreliable.

2. **Evaluation distortion**: A demo that can answer questions does not prove stable tool use, exception handling, marketplace compliance, or KPI accountability in production.
3. **Unclear accountability**: When AI issues an incorrect refund, changes a price incorrectly, misleads a customer, or leaks personal information, enterprises need an auditable chain of responsibility.
4. **Governance lag**: Many organizations deploy capability first and add permissions, audit, shutdown, and human oversight later. The sequence is backwards.
5. **Organizational misfit**: Without role responsibilities, supervision roles, and performance adjustments, a digital employee cannot become an operational asset.

1.4 Gaps in Current Approaches

Approach	Strength	Gap	Standard position
RPA script	Stable and low-cost for fixed processes	Lacks semantic understanding and dynamic decision-making; brittle when interfaces change	May serve as an executor; not a digital employee by itself
Chatbot	Easy to deploy and useful for FAQs	Usually replies but does not own business-state changes	May serve as an interaction entry point; not a role system
Copilot	Improves productivity under human control	Does not hold independent closed-loop authority; accountability remains with the human	May represent L1/L2 capability; not a complete digital employee
Virtual digital human	Human-like interaction and front-end presentation	Avatar quality does not prove role capability; may hide weak back-end capability	May serve as a presentation layer; not the standard object
Point AI tool	Solves local generation, retrieval, or classification tasks	Relies on humans to route context and execute actions	May serve as a component; not a digital employee

An e-commerce digital employee shall cross these gaps: from answering to being accountable, from single tasks to role domains, from demo capability to evaluable capability, and from raw capability to governed capability.

---

2. Terms, Definitions, and Scope

2.1 Core Definitions

E-commerce Digital Employee

A software labor system with explicit e-commerce role responsibilities, KPI/SLA obligations, authorization

boundaries, and governance constraints. It can perceive events, interpret rules, make decisions, invoke tools, and complete cross-system execution in an e-commerce environment, while remaining continuously measurable, auditable, interruptible, and improvable.

E-commerce Digital Human

In market usage, this term often refers to a front-end system with a virtual avatar, voice, video, or human-like interaction. In this standard, it is treated more strictly: it qualifies as an e-commerce digital employee only when the system behind the interface has role responsibilities, authorized execution, evaluation, governance, and business-result accountability. Otherwise, it is only a digital human interface or content persona.

AI Agent

A software entity that understands goals and environmental states, uses models, tools, memory, and external interfaces, and performs tasks with a degree of autonomy. Google Cloud describes agents as systems that combine advanced AI model intelligence with tool access so they can act on a user's behalf under control [R3]. Visa also emphasizes agent observation, decision-making, action loops, and use of tools and APIs [R8].

Agentic System

An engineering system composed of one or more AI agents, memory and knowledge components, policy engines, tool interfaces, monitoring and audit components, and human oversight mechanisms. A digital employee is usually not a single model. It is an agentic system oriented around a role objective.

RPA (Robotic Process Automation)

Technology that executes rule-based tasks by simulating human UI operations or invoking fixed interfaces. RPA may serve as an executor within a digital employee, but without cognition and dynamic decision-making it does not constitute a digital employee.

Copilot

An AI assistant that provides generation, retrieval, recommendations, or local automation while a human remains in control of the workflow. A copilot typically does not hold independent closed-loop execution authority and should not be conflated with a digital employee.

Policy Engine

A component that turns non-negotiable business rules, compliance boundaries, authorization limits, and risk thresholds into machine-enforceable controls. The policy engine constrains model uncertainty.

Executor / Action Tool

A tool component that connects to marketplaces, ERP, OMS, WMS, PIM, CRM, payment, logistics, advertising platforms, or browser interfaces. It is the mechanism through which a digital employee changes business state.

Human Oversight Mechanisms

Human-in-the-loop (HITL): a human must approve critical decisions or actions.

Human-on-the-loop (HOTL): the system runs automatically while humans monitor and may interrupt.

Human-over-the-loop: humans define goals, budgets, rules, and boundaries but do not participate in daily task flow.

Kill Switch

A mechanism that forcibly cuts off part or all execution authority when the system shows unauthorized behavior, abnormal loss, error spikes, attack indicators, or unexplainable behavior.

2.2 Conformance Criteria

A system may be called an "e-commerce digital employee" under this standard only if it meets all of the following clauses:

Clause	Criterion	Explanation
DEF-001	The system shall have a clear role name, role objective, responsibility boundary, and accountable owner	A generic AI assistant shall not be directly declared a digital employee
DEF-002	The system shall have KPI/SLA obligations and continuously collect runtime metrics	Metrics should cover efficiency, quality, stability, safety, experience, and business outcomes
DEF-003	The system shall receive real business events	It should not rely only on manual prompts or one-off user questions
DEF-004	The system shall invoke tools or interfaces to complete authorized business actions	Recommendation-only systems do not constitute complete digital employees
DEF-005	The system shall include a policy engine or equivalent rule control	High-risk actions shall not depend solely on model judgment
DEF-006	The system shall maintain end-to-end audit logs	Audit records should replay input, rationale, tool calls, outputs, and human intervention
DEF-007	The system shall include human oversight, handoff, rollback, and shutdown mechanisms	L3 and above shall not operate without human oversight and shutdown controls
DEF-008	The system shall complete pre-launch evaluation and post-launch continuous evaluation	Demos, proof-of-concepts, or one-off tests shall not replace production evaluation

Systems that do not satisfy these conditions should be described as AI assistants, automation tools, digital human interfaces, RPA workflows, or copilots, not as digital employees.

2.3 Scope

Recommended use cases:

- Multi-channel customer inquiries, logistics queries, and order status explanations.
- Return and exchange review, after-sales ticket triage, and low-risk refund workflows.
- Product information governance, multi-marketplace listing, title and description generation, and multilingual localization.

- Marketing content generation, campaign monitoring, and budget anomaly alerts.
- Competitor price monitoring, inventory alerts, and replenishment recommendations.
- Cross-border multilingual front-line service and back-office ticket classification.

Use with caution:

- High-value refunds, batch price changes, batch delisting, and account permission changes.
- Disputed customer service responses, crisis communication, and public opinion handling.
- Tax, finance, legal, and regulatory interpretation.
- Marketplace penalty appeals, intellectual property disputes, and major consumer complaints.

Out of scope:

- Physical warehouse machinery control without sufficient safety validation.
- Services requiring deep human empathy or psychological intervention.
- Scenarios that cannot define SOPs, cannot be audited, or cannot assign an accountable fallback owner.

---

3. Overall Standard Framework

This standard establishes an eight-layer framework for e-commerce digital employees. The standard object is not a single model. It is a role-level agentic system.

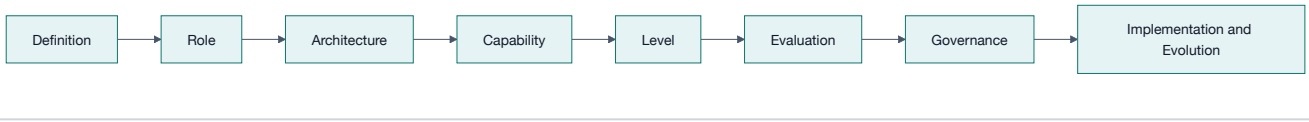
Table 1: Eight-layer standard framework for e-commerce digital employees

Layer	Objective	Key question	Standard output
1. Definition	Align industry language	What is a digital employee, and what is not	Terms, criteria, scope
2. Role	Define business need and accountability	Which role does it perform, who owns it, and how is it measured	Role description, SOP, KPI/SLA, authorization boundary
3. Architecture	Establish an engineering pattern	Which components are required and how they are decoupled	Reference architecture, interfaces, component requirements
4. Capability	Determine role fitness	Can it perceive, reason, decide, and execute	Six capability standards, observations, metrics
5. Level	Manage autonomy	What degree of authority may be granted, and how humans participate	L1-L5 levels, upgrade conditions, prohibited actions
6. Evaluation	Provide evidence for release	How capability is proven, reproduced, and accepted	Offline, sandbox, shadow, gray, and production evaluation

Layer	Objective	Key question	Standard output
7. Governance	Control risk	How to prevent unauthorized action, loss, leakage, and loss of control	Data, permission, audit, shutdown, rollback, appeal
8. Implementation and evolution	Guide scaled adoption	How to move from pilot to digital labor operations	90-day path, maturity model, continuous improvement

This standard aligns governance controls with the Govern, Map, Measure, and Manage risk management loop in NIST AI RMF [R10]. It incorporates the WEF approach of classifying AI agents by role, autonomy, predictability, and context, with progressive governance [R4]. It also uses Anthropic's engineering guidance for agent architecture and tool use [R2].

Figure 1: E-commerce digital employee standard framework



3.1 Clause Numbering System

This standard uses the following prefixes to identify standard requirements:

Prefix	Standard domain	Applicable sections
DEF	Definition and boundary	Terms, definitions, and scope
ROLE	Role modeling	Role objectives, responsibilities, SOP, KPI/SLA, accountable owners
ARCH	Reference architecture	System components, interfaces, tools, audit, oversight
CAP	Capability standards	Perception, cognition, decision-making, execution, learning, business outcomes
LEVEL	Capability levels	L1-L5 autonomy levels, upgrade conditions, prohibited actions
EVAL	Evaluation methods	Test sets, offline tests, sandbox tests, shadow runs, gray releases, production evaluation
GOV	Governance controls	Data, permission, model, execution, audit, shutdown, consumer protection
IMPL	Implementation and evolution	Implementation path, organizational roles, maturity, continuous improvement
CONF	Conformance statement	Adoption statement, level claim, evidence, exemptions

3.2 Standardization Principles

1. **Role before technology**: Define the role and business outcome before choosing models, tools, and architecture.
2. **Authorization before execution**: No write action shall occur without clear permission boundaries, approval rules, and revocation mechanisms.
3. **Evaluation before release**: Higher autonomy levels shall not be granted without offline, sandbox, and gray-release evidence.
4. **Governance before scale**: Systems shall not be scaled into core production flows without audit, shutdown, and human handoff.
5. **Experience and risk carry equal weight**: Automation rate shall not override customer experience, loss control, or brand reputation.

---

4. Reference Architecture and System Components

An e-commerce digital employee shall be designed as a decoupled, multi-component engineering system. Large models shall not be connected directly to production write interfaces.

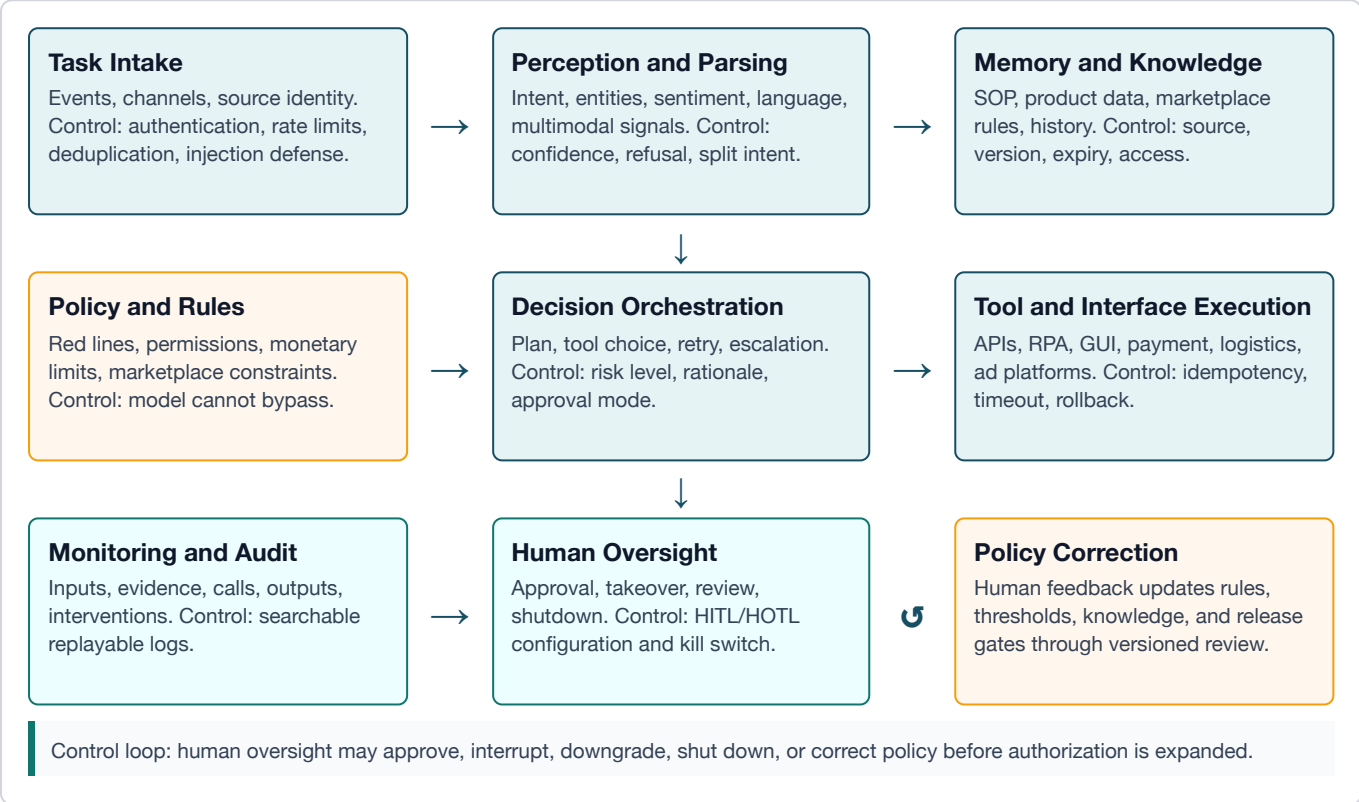
4.1 Standard Reference Architecture

Table 2: Standard reference architecture for e-commerce digital employees

Module	Function	Minimum control requirement
Task intake layer	Receives IM, email, tickets, orders, inventory, ads, logistics, and marketplace events	Authentication, rate limiting, deduplication, injection defense, source identity
Perception and parsing layer	Identifies intent, entities, sentiment, language, image/video information	Confidence score, low-confidence refusal, multi-intent splitting
Memory and knowledge layer	Stores short-term context, enterprise SOP, product knowledge, marketplace rules, historical tickets	Knowledge version, source label, expiration policy, access control
Policy and rule layer	Enforces hard red lines, permissions, monetary thresholds, price thresholds, sensitive terms, marketplace policies	Shall not be bypassed by the model; rule changes shall be auditable
Decision orchestration layer	Decomposes tasks, plans paths, selects tools, retries, and escalates	Outputs plan, risk level, and explanation
Tool and interface execution layer	Invokes APIs, RPA, GUI, browser, payment, logistics, and ad platforms	Idempotency, timeout, retry, duplicate prevention, rollback

Module	Function	Minimum control requirement
Monitoring and audit layer	Records inputs, retrieval, reasoning, tool calls, outputs, and human intervention	Tamper-resistant logs, searchable records, defined retention period
Human oversight layer	Provides approval, takeover, review, shutdown, strategy updates, and feedback	HITL/HOTL configuration and global kill switch

Figure 2: End-to-end reference architecture for e-commerce digital employees



Architecture clauses:

- ARCH-001: An e-commerce digital employee shall use an auditable and layered governance architecture. Large models shall not be connected directly to production write interfaces.
- ARCH-002: The policy and rule layer shall be independent from model reasoning. High-risk red lines shall not be determined by model judgment alone.
- ARCH-003: All external tools and business-system interfaces shall be governed by unified authorization, logging, and exception handling.
- ARCH-004: The monitoring and audit layer shall independently record key inputs, reasoning evidence, tool calls, outputs, and human intervention.
- ARCH-005: The human oversight layer shall support approval, takeover, downgrade, shutdown, and post-incident review.

4.2 End-to-End Closed-Loop Example

Scenario: A buyer requests cancellation of a paid order that has not yet shipped.

1. The task intake layer receives the marketplace after-sales event.
2. The perception layer extracts order ID, buyer identity, payment status, and request reason.
3. The memory and knowledge layer retrieves store cancellation policy, marketplace rules, and historical exception records.
4. The policy layer checks shipment status, high-risk buyer flags, and monetary thresholds.
5. The decision layer plans the workflow: hold shipment, initiate refund, notify warehouse, and respond to buyer.
6. The execution layer invokes OMS, ERP, WMS, and payment interfaces.
7. The audit layer records evidence, interface responses, and final state.
8. The oversight layer triggers human handoff if the amount is abnormal, inventory is locked, the buyer is high risk, or an interface fails.

4.3 Architecture Boundary

Anthropic emphasizes that agents differ from traditional workflows because they can dynamically control decision processes, select tools, and adapt based on feedback [R2]. This does not mean all processes should be delegated to agents. In e-commerce systems, deterministic, low-risk, high-frequency processes may be implemented using workflows or RPA. Agentic systems are better suited for semantically complex, stateful, cross-system tasks that require reasoning and dynamic replanning.

The recommended architecture is therefore:

- **Rules first**: Red lines, monetary thresholds, permissions, and marketplace policies should be enforced by deterministic controls.
- **Models handle uncertainty**: Intent understanding, semantic reasoning, multilingual work, exception diagnosis, and content generation may be handled by models.
- **Tool execution is policy-constrained**: The model shall not bypass the policy layer to invoke write interfaces.
- **Human oversight covers high-risk actions**: Money, price, accounts, public commitments, and consumer rights shall default to stronger oversight.

---

5. Six Core Capability Standards

The six core capabilities determine whether an e-commerce digital employee is fit for its role. Each capability includes a definition, observations, metrics, handoff conditions, and standard clauses. Metric thresholds are recommended baselines and should be calibrated by role risk and business baseline.

5.1 Perception

Definition: The ability to capture business triggers from multi-channel, multimodal, multilingual, and heterogeneous inputs, and to extract structured business entities.

E-commerce observations:

- Can the system identify order IDs, SKUs, tracking numbers, marketplaces, currencies, languages, campaigns, and after-sales reasons?
- Can it handle typos, abbreviations, screenshots, voice, long text, emojis, and marketplace slang?
- Can it split multiple intents within a single user message?

Metric	Recommended measurement	Recommended threshold	Handoff condition
Intent accuracy	Percentage of samples mapped to the correct standard business scenario	>= 90%	Low-confidence or unsupported intent goes to human queue
Key entity recall	Percentage of required fields such as order, SKU, amount, and time correctly extracted	>= 95%	Missing key fields shall block write actions
Language identification accuracy	Correct language detection and knowledge/template selection	>= 95%	Unsupported languages go to human queue
Adversarial input blocking rate	Correct detection of prompt injection, unauthorized requests, and sensitive terms	>= 99%	Red-line hits shall be refused or escalated

CAP-001: The perception layer shall output confidence scores, parsed results, and evidence locations. Low-confidence results shall not enter high-risk execution flows.

5.2 Cognition

Definition: The ability to interpret business state by combining context, enterprise knowledge, product knowledge, order status, marketplace rules, and commercial logic.

E-commerce observations:

- Can the system determine membership benefits, free-shipping rules, return windows, and marketplace penalty rules?
- Can it detect knowledge conflicts, such as a legacy SOP conflicting with a new marketplace policy?
- Can it map a customer issue to an executable SOP rather than produce a generic answer?

Metric	Recommended measurement	Recommended threshold	Handoff condition
Knowledge retrieval hit rate	Percentage of samples with correct high-relevance knowledge retrieval	>= 85%	No evidence means no deterministic commitment
SOP consistency rate	Percentage of outputs consistent with enterprise SOP and marketplace rules	>= 95%	Rule conflicts go to human or policy expert
Red-line refusal rate	Correct refusal rate for prohibited, unauthorized, or out-of-scope requests	100%	Red-line error is a severe incident
Factual hallucination rate	Rate of unsupported claims about policies, inventory, logistics, or price	Approaches 0	Factual information shall use system data

CAP-002: Cognitive results should include evidence sources. Facts involving price, inventory, logistics, refunds, or marketplace policy shall come from trusted business systems or versioned knowledge bases, not from model memory alone.

5.3 Decision-Making

Definition: The ability to decompose non-linear tasks, select tools, sequence actions, and replan or escalate when a path fails.

E-commerce observations:

- Can the system check order status, then logistics, then after-sales policy before initiating a refund?
- Can it choose retry, fallback interface, or human handoff when an API fails?
- Can it identify task risk level and select the correct oversight mode?

Metric	Recommended measurement	Recommended threshold	Handoff condition
Plan correctness	Human blind review of whether the action plan is reasonable	>= 90%	Incomplete high-risk plans shall not execute
Tool selection accuracy	Correct API, system, or template selection	>= 95%	Consecutive wrong tool calls shall stop the task
Failure recovery success rate	Successful completion through an alternate path after first tool failure	>= 80%	More than three retries in one task triggers handoff
Risk classification accuracy	Correct low/medium/high risk classification	>= 95%	Uncertain risk shall be treated as higher risk

CAP-003: The decision layer shall output the action plan, risk level, required authorization, human oversight mode, and failure-handling path.

5.4 Execution

Definition: The ability to complete real business-state changes through external interfaces, business systems, RPA, or GUI tools.

E-commerce observations:

- Can the system query and update orders, inventory, logistics, refunds, ads, and product information?
- Can it ensure idempotency to avoid duplicate refunds, duplicate coupons, or duplicate price changes?
- Can it record state and safely roll back or compensate when an interface fails?

Metric	Recommended measurement	Recommended threshold	Handoff condition
API execution success rate	Successful tool return rate after excluding external downtime	>= 99%	Write failures shall be traceable
End-to-end closure rate	Share of automatable tickets completed without human intervention	Role-specific; customer service may target >= 60%	High-risk scenarios should not optimize for closure rate
Idempotency coverage	Share of write operations protected by unique request and duplicate-submission controls	100%	No idempotency means no fund or price action
Rollback/compensation success rate	Ability to restore a safe state after exception	>= 99%	No compensation path requires human approval

CAP-004: Any write action affecting funds, inventory, price, accounts, public commitments, or consumer rights shall include authorization records, idempotency, audit logs, and exception compensation.

5.5 Learning and Evolution

Definition: The ability to improve policies, knowledge, and work performance based on historical interactions, human corrections, runtime logs, bad cases, and business changes.

E-commerce observations:

- Can human handoff reasons be converted into knowledge updates, rules, or prompt strategies?
- Can the system detect whether similar errors recur?
- Can it update strategy quickly after marketplace rule changes?

Metric	Recommended measurement	Recommended threshold	Handoff condition
Bad-case recurrence rate	Rate at which fixed cases reappear in similar scenarios	Continuous decline	Recurrence increase triggers focused review
Human correction decay rate	Reduction in similar handoff rate within two weeks after correction	>= 50% to 85%	Lack of convergence blocks level upgrade
Knowledge update latency	Time from rule change to system effect	Critical rules should be updated within hours	Downgrade during pending update
Drift detection coverage	Coverage of indicators monitoring model, knowledge, or interface degradation	100% for core flows	Significant drift triggers rollback

CAP-005: Learning mechanisms shall not directly change high-risk policies without approval. Automatically generated rules, prompts, and knowledge updates should enter review, testing, and versioned release workflows.

5.6 Business Outcome Orientation

Definition: The ability to produce role-level business outcomes across efficiency, quality, experience, revenue, cost, and risk, beyond technical test scores.

E-commerce observations:

- Does customer service improve first-contact resolution and response speed, rather than only increasing automated replies?
- Does after-sales automation shorten refund cycles while maintaining zero loss?
- Does product operations improve listing efficiency while reducing incorrect information and marketplace rejection?

Metric	Recommended measurement	Notes
FTE equivalent release	Automated workload converted into full-time-equivalent labor	Enterprises should state what higher-value work the released labor moves to
First contact resolution (FCR)	Tickets resolved without repeated contact or rework	Should be assessed with CSAT and complaint rate
Integrated ROI/TCO	Incremental benefit plus replacement cost saving minus total cost of ownership	Do not judge only by first-three-month ROI
Experience guardrail	CSAT, NPS, complaint rate, marketplace penalty rate	Experience may be a veto metric

Metric	Recommended measurement	Notes
Risk loss rate	Loss caused by wrong refund, wrong compensation, wrong price, or non-compliant commitment	High-risk roles should treat this as zero-tolerance

CAP-006: A digital employee shall not trade customer experience, compliance, or loss control for automation rate.

---

6. Capability Level System: L1-L5

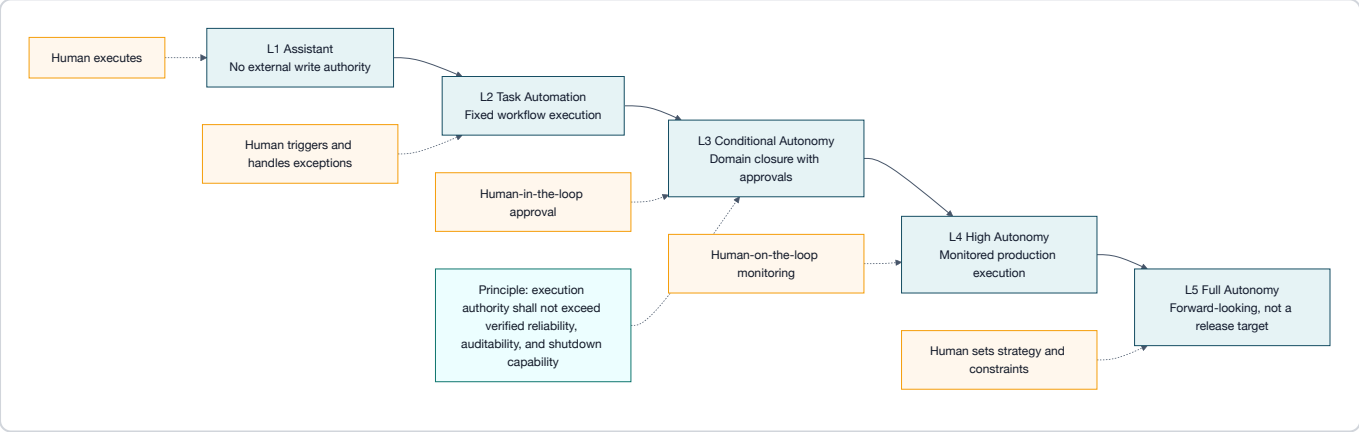
Capability levels determine a digital employee's autonomy and human oversight model. Higher is not always better. The correct level depends on role risk, evidence, and governance maturity.

Table 3: L1-L5 capability levels for e-commerce digital employees

Level	Name	System characteristics	Human role	Recommended e-commerce scenarios	Upgrade condition	Prohibited action
L1	Assistant	Generates scripts, summaries, retrieval results, and drafts; no external write authority	Human executor	Customer reply drafts, product title suggestions, campaign review summaries	Stable suggestion adoption and low factual hallucination	Shall not directly commit refunds, price changes, shipment actions, or marketplace appeals
L2	Task automation	Executes deterministic single tasks or fixed workflows; stops on exception	Human trigger and exception handler	Order lookup, product-field transfer, logistics status sync	Stable interfaces, controlled error rate, idempotency coverage	Shall not dynamically change workflow or handle undefined exceptions
L3	Conditional autonomy	Plans and closes tasks within a defined domain; sensitive actions require approval	Human-in-the-loop approver	After-sales pre-review, standard refunds, customer-service triage, low-risk price recommendations	Declining human rejection rate and reliable red-line blocking	Shall not bypass approval for funds, price, account, or public commitment actions

Level	Name	System characteristics	Human role	Recommended e-commerce scenarios	Upgrade condition	Prohibited action
L4	High autonomy	Runs independently under normal conditions; self-recovers and escalates exceptions	Human-on-the-loop monitor	Standard customer service, low-risk after-sales, mature product operations	Stable production metrics, complete audit, verified shutdown	Shall not operate without monitoring dashboard and kill switch
L5	Full autonomy	Discovers opportunities, restructures workflows, and creates value under strategic goals	Human strategic setter	Forward-looking level only	Requires industry-level evaluation, third-party audit, and mature governance	Not recommended as a current production release target

Figure 3: L1-L5 authorization levels and human oversight model



Higher is not better. A lower level with reliable evidence and clear accountability is preferable to a higher level whose execution authority exceeds verified capability, monitoring coverage, and rollback readiness.

6.1 Relationship Between Level and E-commerce Risk

The same digital employee may operate at different levels for different actions. A customer-service digital employee may reach L4 for logistics lookup, while compensation commitments should remain L1 or L3. Level claims shall be attached to actions, not only to system names.

Recommended default approach:

- Information query, content summary, and reply draft: L1-L2.
- Low-risk deterministic task execution: L2.
- Domain-specific closed-loop work under rules: L3.
- Stable, low-risk, shutdown-protected high-frequency flows: L4.

- Cross-domain proactive business decisioning: not recommended for production L5 release.

6.2 Upgrade Admission Requirements

A digital employee moving from a lower level to a higher level should satisfy the following evidence:

1. It has operated at the previous level for at least one complete business cycle.
2. Core indicators meet the role SLA.
3. Severe compliance violations, financial loss, and PII leakage are zero.
4. Human handoff reasons are explainable, and high-frequency bad cases are closed.
5. Audit log coverage is 100%.
6. Shutdown, rollback, and permission revocation have been exercised.
7. The business owner, compliance/risk owner, and system owner have jointly approved the upgrade.

Level clauses:

- LEVEL-001: Level claims shall be tied to specific roles and actions. A broad system-level claim alone is not sufficient.
- LEVEL-002: Actions involving funds, price, inventory, accounts, public commitments, and consumer rights should be governed at a higher risk level.
- LEVEL-003: L3 and above shall include a policy engine, audit logs, human oversight, and shutdown controls.
- LEVEL-004: L4 shall include production-grade monitoring dashboards, exception alerts, role-level shutdown, and periodic audit.
- LEVEL-005: L5 is a forward-looking level in this version and is not a recommended production release level.

---

7. Evaluation System and Methodology

The evaluation system is not designed to prove that the system is "smart." It is designed to prove that the system is usable, controllable, reproducible, and accountable under real e-commerce constraints.

7.1 Dual-Track Evaluation Principle

Digital employee evaluation shall run on two linked tracks:

Evaluation track	What it proves	Typical evidence	Release implication
Capability track	Whether the system can perceive, reason, decide, use tools, refuse, and recover within the role domain	Offline test sets, sandbox APIs, adversarial tests, regression cases, human blind review	Determines whether the system may enter shadow or gray release

Evaluation track	What it proves	Typical evidence	Release implication
Production track	Whether the system remains stable, valuable, and governable under real traffic	Shadow comparison, gray release metrics, audit sampling, incident records, business KPI/SLA	Determines authorization level, traffic share, and continued operation

The capability track alone is insufficient because benchmark or test-set success does not prove operational reliability. The production track alone is insufficient because live metrics without controlled test evidence may hide untested edge cases. L3 and above shall connect both tracks before authorization is expanded.

7.2 Five-Stage Evaluation

Table 4: Five-stage evaluation system for e-commerce digital employees

Stage	Objective	Test content	Passing evidence
1. Offline test	Validate perception, cognition, and initial decisioning	Standard test set, historical tickets, edge cases, adversarial samples	Accuracy, refusal rate, red-line blocking, hallucination rate
2. Sandbox test	Validate tool calling and exception handling	Simulated APIs, test accounts, simulated orders, simulated inventory	Parameter correctness, idempotency, retry, compensation
3. Shadow run	Validate judgment quality on real traffic	Side-by-side production traffic without execution	Human-machine comparison, adoption rate, risk-classification accuracy
4. Gray release	Validate execution stability on small traffic	Low-risk production actions, limited traffic, HITL	Closure rate, error rate, handoff rate, user experience
5. Production evaluation	Validate long-term business value and governance effectiveness	Stable production traffic, audit sampling, metric dashboard	SLA, FCR, CSAT, ROI/TCO, incident rate

WEF proposes that AI agent adoption connect classification, evaluation, risk assessment, and progressive governance [R4]. Under this standard, evaluation results shall determine authorization level. Evaluation shall not be treated as an after-launch add-on.

Figure 4: Evaluation evidence chain and authorization release process

7.3 Test Set Composition

Recommended test set structure:

Sample type	Share	Purpose	Examples
Routine samples	60%-70%	Validate main-flow efficiency and accuracy	Logistics query, size question, return rule, order cancellation
Edge samples	20%-30%	Validate complex state and exception handling	Partial shipment, multi-marketplace orders, cross-border tax, stacked discounts
Adversarial/red-line samples	10%	Validate refusal, blocking, and escalation	Unauthorized refund, privacy request, induced rule bypass
Regression samples	Continuous	Validate that updates do not reintroduce old failures	Historical bad cases, major complaint samples

Test sets shall have a version number, source record, annotator record, annotation rules, and validity period. User data in test samples shall be desensitized.

7.4 Runtime Metrics

Metric family	Metrics	Purpose	Risk note
Efficiency	First response time, handling time, throughput, concurrency	Measures speed and scale	Speed does not replace quality
Quality	FCR, factual accuracy, SOP consistency, rollback/rework rate	Measures business correctness	Requires human sampling
Stability	API success rate, recovery time, retry success rate, shutdown frequency	Measures engineering reliability	Zero shutdowns may indicate failed risk detection
Safety	Unauthorized action rate, PII leakage rate, red-line blocking rate, audit completeness	Measures baseline controls	High-risk roles should be zero-tolerance
Experience	CSAT, complaint rate, repeated contact rate, human escalation satisfaction	Measures user experience	Bots shall not be used to suppress complaints
Business	FTE release, conversion, refund turnaround, ad ROI, integrated TCO	Measures business value	Requires baseline comparison and attribution limits

7.5 Acceptance Report Requirements

Before launch, a digital employee shall have an acceptance report covering at least:

1. Role description and authorization boundary.
2. Test set version and sample structure.
3. Offline, sandbox, shadow, and gray-stage results.
4. Failed items, risk exemptions, and compensating controls.
5. Human oversight mode and handoff mechanism.
6. Audit log samples.
7. Shutdown and rollback exercise records.
8. 30/60/90-day post-launch review plan.

Evaluation clauses:

- EVAL-001: Any digital employee shall complete offline and sandbox tests before production launch.
- EVAL-002: L3 and above shall complete shadow runs or equivalent real-traffic side-by-side validation.
- EVAL-003: L3 and above shall complete gray release before full production release.
- EVAL-004: The evaluation report shall record test set version, sample composition, metric results, failed items, risk exemptions, and compensating controls.
- EVAL-005: After changes to model, prompt, tool, knowledge base, or policy, regression testing shall be performed according to the risk of the change.

---

8. Role Reference Models

Digital employees shall be managed as roles. Each role shall have objectives, responsibilities, inputs, outputs, interfaces, KPI, authorization level, and governance requirements.

8.1 Role Modeling Template

Table 7: Role modeling template for e-commerce digital employees

Item	Content
Role name	Example: After-sales refund digital employee
Role objective	Business outcome the role is designed to improve
Scope	Supported scenarios, channels, marketplaces, and languages
Out-of-scope items	Scenarios that shall be escalated to humans or prohibited
Inputs	Events, tickets, messages, and system data
Outputs	Replies, ticket labels, interface actions, reports, alerts

Item	Content
Key interfaces	OMS, ERP, WMS, CRM, payment, logistics, ad platforms, and others
KPI/SLA	Response, quality, experience, risk, and business metrics
Recommended level	L1-L4, differentiated by action
Governance requirements	Permission, audit, approval, shutdown, rollback
Accountable owners	Business owner, system owner, human supervisor

Role clauses:

- ROLE-001: A digital employee shall define the role before defining features. A feature list shall not replace role modeling.
- ROLE-002: Each role shall define role objective, responsibility boundary, inputs and outputs, key interfaces, KPI/SLA, authorization level, and accountable owners.
- ROLE-003: Each role shall define out-of-scope scenarios and human handoff conditions.
- ROLE-004: Role KPI shall not include efficiency only. It shall also cover quality, experience, and risk.
- ROLE-005: The same digital employee shall declare authorization levels separately for different actions.

8.2 Customer Service Digital Employee

Role objective: Handle front-line inquiries, order lookups, logistics explanations, standard responses, and multilingual service to improve response speed and first-contact resolution.

Item	Recommended standard
Core responsibilities	FAQ, product inquiries, order/logistics lookup, pre-sales guidance, ticket triage, multilingual replies
Key interfaces	IM platform, CRM, OMS, logistics tracking, product knowledge base
Recommended level	FAQ/query L3-L4; compensation commitments L1-L3
KPI	3-minute response rate, FCR, CSAT, repeated contact rate, human handoff rate
Handoff conditions	Angry sentiment, legal threat, marketplace complaint, monetary dispute, abnormal identity, out-of-scope issue
Controls	Evidence-backed replies; no fabricated logistics, inventory, or promotion; human channel required

8.3 After-sales and Refund Digital Employee

Role objective: Improve return and refund review efficiency while controlling financial loss and consumer-experience risk.

Item	Recommended standard
Core responsibilities	Return eligibility, logistics verification, material completeness check, shipping insurance matching, refund workflow
Key interfaces	After-sales tickets, OMS, WMS, logistics, payment, finance system
Recommended level	Pre-review L3; low-value refund L3/L4; high-value refund L1/L3
KPI	Refund turnaround, incorrect refund rate, loss rate, complaint rate, human rejection rate
Handoff conditions	High-value order, fraud risk, repeated refund, logistics dispute, marketplace arbitration
Controls	Monetary threshold, idempotency, duplicate-refund prevention, human approval, 100% audit logs

8.4 Product Operations Digital Employee

Role objective: Improve product information maintenance, listing, marketplace adaptation, and product-data quality.

Item	Recommended standard
Core responsibilities	Product title/description generation, category mapping, attribute completion, multilingual translation, image quality check, listing readiness
Key interfaces	PIM, ERP, marketplace product APIs, image/video asset library
Recommended level	Draft L1; field transfer L2; low-risk listing L3
KPI	Listing time, field completeness, marketplace rejection rate, information error rate
Handoff conditions	Sensitive categories, regulated terms, intellectual property, medical/performance claims, uncertain marketplace rules
Controls	Versioned marketplace rule base, high-risk term blocking, gray release for batch actions

8.5 Marketing Operations Digital Employee

Role objective: Assist marketing teams in monitoring ads, budgets, creatives, and campaign performance, and execute low-risk optimization actions.

Item	Recommended standard
Core responsibilities	Data monitoring, anomaly alerts, creative drafts, keyword expansion, budget recommendations, campaign review

Item	Recommended standard
Key interfaces	Ad platforms, BI, asset library, product database, CRM
Recommended level	Analysis recommendation L1; budget alert L2; small optimization L3
KPI	Anomaly detection speed, recommendation adoption, creative approval rate, assisted ad ROI improvement
Handoff conditions	Large budget change, brand-sensitive creative, marketplace penalty risk
Controls	Budget thresholds, creative review, brand term list, change audit

8.6 Content and Video Digital Employee

Role objective: Increase output capacity for product content, short-video scripts, livestream clips, multilingual localization, and A/B testing assets.

Item	Recommended standard
Core responsibilities	Selling-point extraction, short-video scripts, voiceover copy, multilingual rewriting, asset tagging
Key interfaces	PIM, DAM, video tools, translation engines, marketplace content review
Recommended level	Content draft L1; low-risk batch rewrite L2; auto-publishing requires L3 and approval
KPI	Asset production time, review pass rate, content compliance rate, assisted conversion metrics
Handoff conditions	Medical, finance, children, performance claims, brand crisis, infringement risk
Controls	Fact sources, prohibited terms, copyright check, publishing approval

8.7 Fulfillment Coordination Digital Employee

Role objective: Detect fulfillment exceptions and coordinate inventory, warehouse, logistics, and customer-service actions.

Item	Recommended standard
Core responsibilities	Stockout alerts, delayed-shipment detection, split-shipment suggestions, logistics exception notification, customer-service preparation
Key interfaces	WMS, OMS, logistics, inventory, customer-service system

Item	Recommended standard
Recommended level	Alert L2; low-risk notification L3; reshipment/replacement L3 with approval
KPI	Exception detection speed, fulfillment delay rate, proactive notification coverage, complaint reduction
Handoff conditions	Large warehouse disruption, cross-border customs, compensation, marketplace penalty
Controls	Reshipment/replacement approval, inventory-lock idempotency, customer notice template review

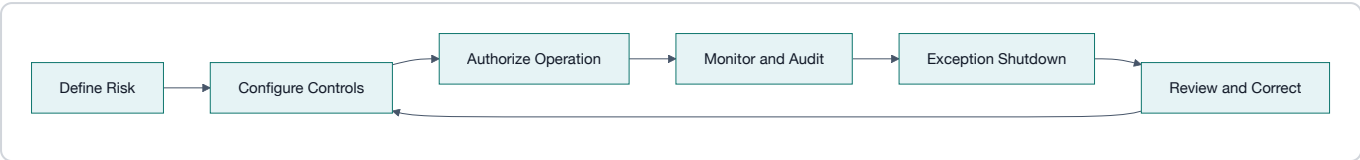
9. Governance Requirements and Control Baselines

Governance is not an add-on. It is the precondition for authorization. NIST AI RMF emphasizes governance, context mapping, measurement, and management across AI risk [R10]. WEF also notes that real AI agent deployment must connect architecture, role classification, evaluation, and progressive governance [R4].

Table 5: Governance control domains for e-commerce digital employees

Control domain	Control objective	Minimum requirement
Data governance	Prevent unauthorized access, leakage, and contamination	Data minimization, PII desensitization, knowledge versioning
Permission governance	Prevent unauthorized execution and shared credentials	RBAC/ABAC, least privilege, temporary authorization, immediate revocation
Model and knowledge governance	Prevent version loss of control and policy drift	Versioning, regression testing, approved release, rollback
Execution governance	Prevent duplicate submissions, incorrect writes, and financial loss	Idempotency, thresholds, state recheck, compensation path
Auditability	Support review, accountability, and external assessment	Full records of input, evidence, calls, outputs, and human intervention
Exception shutdown	Control incident spread	P0-P3 classification, role-level shutdown, global kill switch
Consumer protection	Protect user awareness, appeal, and correction rights	Human channel, explainability, review path, anti-misleading controls

Figure 5: Governance feedback loop



9.1 Data Governance

Requirements:

1. The system shall follow data minimization and access only fields required for the task.
2. Personal identity information, payment information, addresses, phone numbers, IDs, and sensitive notes shall be desensitized or permission-isolated.
3. Production user data shall not be used directly for model training unless legally authorized and desensitized.
4. Cross-border e-commerce scenarios should record data location, model invocation location, processing purpose, and retention period.
5. Knowledge bases shall have source, version, owner, and validity period.

9.2 Permission Governance

Requirements:

1. Every tool call shall be linked to a service account, role, permission scope, and validity period.
2. Write operations shall use least privilege and shall not use personal administrator accounts or shared keys.
3. High-risk actions shall have limits for amount, count, frequency, category, marketplace, and time window.
4. Immediate revocation, role-level suspension, and global kill switch shall be supported.
5. Permission changes shall be auditable.

9.3 Model, Knowledge, and Policy Governance

Requirements:

1. Base models, prompts, tool schemas, knowledge bases, and rule engines shall be versioned.
2. Any policy change affecting funds, price, consumer commitments, or marketplace rules shall be tested and approved before release.
3. Model upgrades shall undergo regression testing that covers historical bad cases.
4. Rollback to a stable version shall be available within minutes or within an acceptable operational time window.
5. The model shall not be allowed to modify red-line policies and make them effective without approval.

9.4 Execution Governance

Requirements:

1. All write interfaces shall have idempotency keys, duplicate-submission prevention, timeout handling, and returned-state records.
2. Actions involving funds, price, inventory, accounts, or public content shall have a risk level.
3. High-risk actions shall enter HITL approval.
4. Repeated failures or repeated calls beyond a threshold shall stop the task and trigger human handoff.
5. Post-execution state shall be checked to confirm that business-system results match the expected state.

9.5 Audit Trail

Audit logs shall record at least:

- Original input and source channel.
- Parsed intent, entities, and confidence score.
- Retrieved knowledge and system data.
- Decision plan, risk level, and oversight mode.
- Tool-call parameters, return values, latency, and errors.
- Final output, business-state changes, and user notification.
- Human approval, takeover, rejection, and correction records.
- Model, prompt, knowledge-base, and rule versions.

The recommended retention period is at least 90 days. Records involving transactions, funds, complaints, marketplace penalties, or compliance obligations should be retained longer according to enterprise and regulatory requirements.

9.6 Exception Classification and Shutdown

Severity	Example	Response requirement
P0 critical incident	PII leakage, batch incorrect refund, unauthorized write, unexplainable fund action	Immediate global shutdown, owner notification, incident response
P1 high-risk exception	Incorrect price change, consecutive high-value refunds, marketplace-rule violation, major complaint	Role-level shutdown, freeze related interfaces, human review
P2 medium-risk exception	Error-rate increase, interface failure, repeated reply, knowledge conflict	Downgrade operation, review trigger, limit high-risk actions
P3 low-risk exception	Individual poor reply, low-confidence refusal, minor delay	Enter bad-case queue and knowledge optimization

Example shutdown triggers:

- Ten consecutive authentication failures within one minute.
- Abnormal concentration of high-value refund approvals within five minutes.
- SKU price changes exceeding the configured threshold.
- PII detector finds sensitive information in an external model request.
- The same tool call fails more than three times.
- Complaint rate, human handoff rate, or factual error rate exceeds the dynamic baseline.

9.7 Consumer Protection

Requirements:

1. Users should have access to human appeal or human handoff.
2. Digital employees shall not fabricate inventory, logistics status, compensation policy, or marketplace rules.
3. Decisions with material impact on consumer rights should be explainable, reviewable, and correctable.
4. A human-like front-end digital human shall not mislead users into believing it is human or has unlimited authorization.
5. Customer-service digital employees should detect emotion, complaints, legal threats, and vulnerable-consumer contexts and escalate to humans.

9.8 Governance Framework Alignment

Digital employee deployments shall maintain a governance mapping to applicable AI risk-management frameworks, standards, laws, regulations, and marketplace rules.

External framework	Relevant position	Standard mapping	Adoption requirement
NIST AI RMF 1.0	AI risk should be governed, mapped, measured, and managed [R10]	Eight-layer framework, role modeling, evaluation stages, governance controls, incident review	Governance owners should map internal controls to Govern, Map, Measure, and Manage functions
ISO/IEC 42001:2023	Specifies requirements for establishing, implementing, maintaining, and continually improving an AI management system [R18]	Owner assignment, policy/version control, audit, review cadence, continuous improvement	Organizations operating digital employees should maintain AI management-system evidence
EU AI Act, Regulation (EU) 2024/1689	Establishes risk-based obligations for AI systems, with phased application timelines [R19]	Risk classification, human oversight, transparency, auditability, post-launch monitoring, incident response	EU-facing deployments shall maintain a current legal mapping by system category, jurisdiction, role, and use case

External framework	Relevant position	Standard mapping	Adoption requirement
WEF AI agent governance work	Agent adoption should connect role classification, evaluation, risk assessment, authorization, and scaling [R4][R5]	Role-level conformance, L1-L5 authorization, dual-track evaluation, progressive release	Cross-organization agent deployments should map roles, authorization, and escalation responsibilities

Governance alignment shall be maintained as a living mapping. When laws, standards, or marketplace rules change, enterprises should update policy rules, test sets, audit criteria, and conformance statements before expanding authorization.

9.9 Minimum Governance Control Checklist

Before launch, the following controls shall be completed:

- [ ] PII desensitization and sensitive-field permission isolation.
- [ ] Separation of read and write permissions, with least privilege for write operations.
- [ ] Idempotency, duplicate prevention, and audit for all write operations.
- [ ] Risk levels for actions affecting funds, price, inventory, accounts, and public content.
- [ ] HITL/HOTL oversight configuration.
- [ ] Global kill switch and role-level shutdown exercise.
- [ ] Rollback capability for model, prompt, knowledge base, and rule versions.
- [ ] User human-handoff and appeal path.
- [ ] Launch acceptance report jointly approved by business, engineering, and risk/compliance.

Governance clauses:

- GOV-001: L3 and above shall complete minimum controls for data, permission, execution, audit, and shutdown.
- GOV-002: Production user data shall not enter training or external model invocation chains without authorization, desensitization, and documented processing purpose.
- GOV-003: All write operations shall use role-level service accounts or equivalent mechanisms. Personal administrator accounts or shared keys shall not be used.
- GOV-004: Actions affecting funds, price, inventory, accounts, public content, or consumer rights shall include auditability and rollback or compensation paths.
- GOV-005: The system shall provide human handoff, appeal, review, and correction paths.
- GOV-006: The system shall support role-level and global shutdown. Shutdown capability shall be exercised before launch.

---

10. Implementation Path and Organizational Adoption

Digital employee adoption is not a single technology project. It is a redesign of roles, processes, permissions, and governance.

10.1 Preconditions

Enterprises should not launch digital employees directly into scenarios with chaotic processes, missing SOPs, unstable interfaces, unclear permissions, or no metric baseline. Preconditions include:

1. The target role has a clear SOP.
2. Key systems have stable APIs or controlled execution methods.
3. Business baselines are measurable, including manual handling time, cost, complaint rate, and refund turnaround.
4. Data and knowledge can be desensitized, versioned, and maintained.
5. The business team is willing to assign supervisors and run bad-case review.
6. Management accepts a 60- to 90-day period for trial, gray release, and downgrade.

10.2 90-Day Implementation Blueprint

Stage	Time	Objective	Key actions	Exit condition
Stage 0	Before launch	Establish baseline and boundaries	Select role, map SOP, review interfaces, define permissions, collect manual baseline	Role description and authorization boundary confirmed
Stage 1	Days 1-30	System preparation and offline validation	Knowledge-base initialization, tool integration, test set annotation, offline and sandbox tests	Core capability reaches L1/L2 requirements
Stage 2	Days 31-60	Shadow run and gray release	Side-by-side production traffic, human-machine comparison, small-traffic execution, bad-case fixes	Adoption, error, and handoff metrics meet threshold
Stage 3	Days 61-90	Staged authorization and performance review	L2/L3 actions launched, oversight dashboard, shutdown exercise, KPI/SLA tracking	Role acceptance metrics met

Stage	Time	Objective	Key actions	Exit condition
Stage 4	After day 90	Scaled operation	Role replication, continuous audit, organizational role adjustment, ROI/TCO review	Enter next maturity stage

10.3 Organizational Roles

Role	Responsibility
Executive sponsor	Defines business goals, budget, risk appetite, and cross-functional coordination
Business owner	Defines role, SOP, KPI, acceptance criteria, and human handoff rules
AI product owner	Owns feature design, runtime metrics, and bad-case mechanism
System architecture/engineering owner	Owns interfaces, tools, logs, reliability, and rollback
Data/knowledge-base administrator	Owns knowledge source, versioning, desensitization, expiration, and update
Risk/compliance owner	Approves permissions, red lines, audit, shutdown, and incident response
Human supervisor	Approves high-risk actions, handles exceptions, and provides correction feedback
Department AI ambassador	Drives frontline adoption, training, feedback, and process improvement

Prosus emphasizes that non-technical employees using AI agents need department ambassadors and technical support roles [R7]. This aligns with e-commerce digital employee adoption: the hard part is not only model capability, but whether business teams can continuously feed knowledge, exceptions, and feedback into the system.

Workforce planning should be part of the implementation charter rather than a post-launch reaction. The World Economic Forum Future of Jobs Report 2025 projects 170 million roles created and 92 million displaced by 2030, and reports that 86% of employers expect AI and information processing technologies to transform their business by 2030 [R17]. For e-commerce operators, this means digital employee deployment should define how human roles move from transaction handling toward exception management, policy ownership, knowledge maintenance, customer escalation, and AI supervision.

Implementation clauses:

- IMPL-001: Enterprises should establish manual business baselines before implementation, including handling time, cost, quality, experience, and risk.

- IMPL-002: Enterprises should pilot high-frequency, rule-clear, risk-controllable, interface-stable scenarios first.
- IMPL-003: Enterprises shall not launch L3 or above digital employees directly into scenarios where SOPs are missing, accountability is unclear, or interfaces are not auditable.
- IMPL-004: Enterprises should assign a business owner, system owner, risk/compliance owner, and human supervisor.
- IMPL-005: Enterprises should establish 30/60/90-day post-launch reviews and adjust authorization levels based on review results.

---

11. Maturity Model and Continuous Improvement

Stage	Characteristics	Key metrics	Governance focus	Exit condition
1. Pilot validation	Single role, low risk, small traffic	Accuracy, adoption, interface connectivity	Data desensitization, least privilege, HITL	Process feasibility proven
2. Controlled expansion	Multi-scenario replication, beginning write actions	Closure rate, handoff rate, error rate	Audit, idempotency, rollback, policy engine	L2/L3 stable operation
3. Scaled operation	Digital employees become daily production roles	SLA, FCR, CSAT, ROI/TCO	Shutdown, continuous monitoring, incident response	Core metrics stable
4. Continuous optimization	Multi-role collaboration, knowledge and policy evolution	Bad-case recurrence, correction decay	Periodic audit, drift detection, third-party assessment	Ready for broader conformance alignment

Continuous improvement mechanisms:

1. Daily monitoring: errors, handoffs, shutdowns, interface exceptions, complaints.
2. Weekly review: frequent bad cases, knowledge gaps, rule conflicts.
3. Monthly evaluation: KPI/SLA, ROI/TCO, user experience, risk events.
4. Quarterly audit: permissions, logs, model versions, knowledge base, compliance controls.
5. Every version upgrade: regression test, gray release, rollback plan.

---

12. Case Method Template

Cases are used to show how the standard is applied. They shall not be used for exaggerated marketing. Each case shall disclose boundary conditions, observation period, sample size, baseline, authorization level,

handoff conditions, and attribution limits. Negative and null results are publishable under this template. A case that shows no statistically meaningful improvement may still be valuable if it identifies unsupported scenarios, weak data foundations, or governance controls that prevented loss.

Published case metrics should be tied to a defined observation window and sample base. For high-frequency customer-service or after-sales cases, a recommended minimum is one complete business cycle and, where traffic allows, no fewer than 10,000 tickets or equivalent events. Smaller samples may be disclosed, but the limitation shall be stated. Where comparisons are made against a baseline, confidence intervals or an equivalent uncertainty statement should be included when applicable.

12.1 Case Disclosure Template

Item	Requirement
Business background	Business scale, channels, languages, pain points
Role scope	What the digital employee handles and does not handle
System scope	Connected systems, knowledge bases, tools, and models
Authorization level	L1-L4, differentiated by action
Baseline metrics	Pre-launch manual efficiency, cost, quality, and experience
Evaluation process	Offline, sandbox, shadow, gray, and production evaluation
Observation period and sample size	Calendar window, traffic share, total events, exclusions, and minimum sample rationale
Outcome metrics	Efficiency, quality, experience, risk, ROI/TCO, with confidence intervals or uncertainty notes where applicable
Risk events	Errors, handoffs, shutdowns, complaints, and resolutions
Negative/null findings	Scenarios with no improvement, degraded performance, or blocked release, including root cause
Attribution limits	Which results cannot be fully attributed to the digital employee

12.2 Example: Cross-Border Customer Service Digital Employee

Business background: A cross-border brand sells to North America and Europe. Tickets concentrate on logistics tracking, size questions, return policies, and promotion explanations. The human team is constrained by time zones and language coverage, resulting in long first-response time.

System scope: IM platform, CRM, OMS, logistics tracking, product knowledge base, multilingual template library.

Authorization levels:

- Logistics lookup and order status explanation: L4.

- Return and exchange policy explanation: L3.
- Compensation, discount commitment, and marketplace complaint response: L1/L3.

Example baseline metrics:

- Average human first response: 4 hours.
- Basic inquiry share handled by humans: 70%.
- Supported languages: English and Spanish.

Example result presentation:

- First response reduced to minutes.
- Basic inquiry auto-closure reaches an acceptable operating range.
- Additional German, French, Italian, and other language support is added.
- High-risk complaints remain under human handoff.

Attribution limits: Multilingual quality, logistics-system stability, and clarity of promotion rules materially affect results. Improvements should not be fully attributed to the model alone.

12.3 Example: After-sales Refund Digital Employee

Business background: Order volume growth causes return review backlog. Human agents switch among marketplaces, OMS, WMS, logistics, and payment systems.

System scope: After-sales tickets, OMS, WMS, logistics tracking, payment refund interface, refund policy engine.

Authorization levels:

- Return-material completeness check: L3.
- Low-value standard refunds: L3/L4.
- High-value, abnormal-account, and disputed refunds: L1/L3.

Key controls:

- Monetary threshold.
- Duplicate-refund prevention through idempotency key.
- Logistics receipt verification.
- High-risk account flag.
- Post-refund state recheck.

Example result presentation:

- Standard refund cycle is shortened.
- Incorrect refund rate and loss rate remain within zero-tolerance range.
- Human staff focus on fraud, disputes, and marketplace arbitration.

Attribution limits: Warehouse scan timeliness, logistics data quality, and marketplace after-sales rule changes affect performance.

12.4 Example: Product Operations Digital Employee

Business background: Multi-marketplace listing requires repeated rewriting of titles, attributes, descriptions, image specifications, and multilingual content. Manual work is time-consuming and error-prone.

System scope: PIM, ERP, asset library, marketplace product APIs, marketplace rule base, prohibited term list.

Authorization levels:

- Title and description draft: L1.
- Field-format conversion: L2.
- Low-risk product listing: L3.
- Sensitive categories and performance claims: human approval.

Example result presentation:

- Product listing preparation time is reduced.
- Field completeness improves.
- Marketplace rejection rate declines.

Attribution limits: Product master data quality determines the ceiling. If the original product data is disordered, a digital employee will amplify governance defects.

---

13. Conformance Statement and Adoption Model

Conformance claims shall disclose the applicable role, scenario, capability level, evaluation evidence, governance controls, and out-of-scope items. Third-party certification or audit conclusions shall be issued only by qualified independent assessment bodies.

13.1 Conformance Classes

Table 6: Conformance classes for e-commerce digital employees

Class	Applicable object	Minimum requirement	Claim boundary
C0 component conformance	A single model, tool, knowledge base, RPA, or interface component	Explains its module role in the digital employee architecture	Shall not claim to be a complete digital employee

Class	Applicable object	Minimum requirement	Claim boundary
C1 basic digital employee	L1-L2 role assistance or task automation system	Meets basic DEF, ROLE, ARCH clauses and pre-launch evaluation	May claim assistant or task-automation level only
C2 conditional autonomy digital employee	L3 domain closed-loop system	Meets DEF, ROLE, ARCH, CAP, LEVEL, EVAL, GOV clauses; includes HITL	May claim conditional autonomy
C3 production high-autonomy digital employee	L4 stable production system	Meets C2 requirements and includes production monitoring, role-level shutdown, continuous audit, periodic review	May claim high autonomy
C4 forward-looking autonomy capability	L5 research or pilot system	Discloses experimental boundary, non-production nature, and risk controls	Not recommended as production conformance claim

13.2 Conformance Statement Materials

A conformance statement should include at least:

1. System name, version, provider, and user.
2. Applicable role, business scenario, channel, marketplace, and language.
3. Claimed class: C0-C4, plus action-level L1-L5 mapping.
4. Satisfied clause list: DEF, ROLE, ARCH, CAP, LEVEL, EVAL, GOV, IMPL.
5. Evaluation evidence: test set version, sample composition, stage results, production observation period.
6. Governance evidence: permission configuration, audit samples, shutdown exercise, rollback mechanism, human oversight.
7. Exemptions and non-applicable items: uncovered clauses, rationale, and compensating controls.
8. Accountable owners: business owner, system owner, risk/compliance owner, human supervision owner.
9. Validity period and review cycle.

13.3 Conformance Clauses

- CONF-001: A conformance statement shall be tied to a specific system version, role, scenario, and authorization level.
- CONF-002: C0 component capability shall not be packaged as a complete digital employee.
- CONF-003: A single demo, proof-of-concept, or model capability test shall not replace a conformance statement.
- CONF-004: L3 and above claims shall provide shadow-run, gray-release, or equivalent pre-production validation evidence.
- CONF-005: A conformance statement shall disclose out-of-scope items, exemptions, and compensating controls.

- CONF-006: After major changes to model, knowledge base, tools, permissions, or policies, the original conformance statement should be reassessed.

13.4 Adoption Guidance

Enterprises adopting this standard should proceed in the following order:

1. Use DEF clauses to determine whether a supplier solution qualifies as a digital employee.
2. Use ROLE clauses to establish role descriptions and accountability boundaries.
3. Use ARCH and CAP clauses to evaluate architecture and capability.
4. Use LEVEL clauses to determine authorization levels.
5. Use EVAL clauses to design testing and acceptance.
6. Use GOV clauses to establish launch governance baselines.
7. Use CONF clauses to form internal or external conformance statements.

---

14. Future Standardization Topics

This standard identifies the following areas for further standardization:

1. **Agent-ready product data standards**: Product titles, attributes, price, inventory, reviews, and return policies need to be readable, trustworthy, and real-time accessible to AI agents.
2. **Know Your Agent (KYA)**: Payment providers, platforms, and merchants need to identify agent identity, authorization scope, and behavioral trust.
3. **Agentic commerce protocol adaptation**: ACP, AP2, A2A, MCP, and related protocols are shaping agent transaction infrastructure [R9][W1]. Merchant systems need an adoption strategy.
4. **Interaction rules between buyer agents and seller digital employees**: Consumer agents may directly negotiate inventory, price, fulfillment, and after-sales with merchant-side digital employees.
5. **Agent preference audit**: Research on AI shopping agents shows that agents may exhibit position effects, model-dependent behavior, and demand concentration that differ from human-centric commerce, raising platform design, seller strategy, and regulatory questions [R14].
6. **Public e-commerce benchmark datasets**: The industry needs public test sets covering long-tail after-sales, marketplace rules, cross-border scenarios, multilingual cases, and adversarial samples.
7. **Third-party evaluation and certification**: A balanced assessment system should separate system designers, users, and auditors.
8. **Consumer transparency and appeal rights**: When consumers are served by digital employees or represented by buyer agents, platforms should clarify notice, authorization, appeal, and correction paths.

---

15. Conclusion

The value of an e-commerce digital employee is not to wrap a large model in the image of a "talking employee." Its value is to convert high-frequency, complex, cross-system, verifiable e-commerce work into digital labor that can be defined, authorized, evaluated, and governed.

The industry does not need more vague terminology. It needs a shared standard language: what a digital employee is and is not; what it may and may not do; how capability is proven; how risk is controlled; and how enterprises move from pilot to scaled operations. Only when a digital employee combines role accountability, engineering closure, capability evidence, and governance controls can it become a trusted production unit for e-commerce enterprises.

Yepai Team publishes this standard to establish a common baseline for e-commerce digital employee definition, authorization, evaluation, governance, implementation, and conformance.

Appendix A: Core Glossary

Term	Definition
E-commerce digital employee	A software labor system with e-commerce role responsibilities, KPI/SLA, authorization boundaries, and governance constraints, capable of perception, cognition, decision-making, execution, and continuous evaluation
E-commerce digital human	A front-end human-like interface; it qualifies as a digital employee only when backed by a role-based agentic system
AI Agent	A software entity that understands goals, uses tools, acts, and completes tasks with a degree of autonomy
Agentic System	A system composed of agents, memory, tools, policies, audit, and oversight
RPA	Automation script or workflow for rule-based operations
Copilot	AI assistant that supports human-led work through recommendation, generation, retrieval, or local automation
Policy engine	Component that enforces permissions, compliance, monetary limits, price thresholds, and marketplace policies
Executor	Component that invokes APIs, RPA, GUI, or protocol interfaces to perform business actions
HITL	Human-in-the-loop, where critical actions require human approval
HOTL	Human-on-the-loop, where the system runs automatically under human monitoring

Term	Definition
Kill switch	Mechanism that forcibly removes execution authority under abnormal conditions
FTE	Full-time equivalent labor
TCO	Total cost of ownership

Appendix B: Metric Definitions

Metric	Calculation	Notes
Intent accuracy	Correctly identified intent samples / total samples	Unsupported scenarios should be reported separately
Entity recall	Correctly extracted key entities / required entities	E-commerce entities include order, SKU, amount, tracking number, marketplace
SOP consistency	Outputs consistent with latest SOP / sampled outputs	SOPs shall be versioned
Red-line blocking rate	Correctly blocked red-line requests / red-line requests	High-risk scenarios require 100%
End-to-end closure rate	Completed without human intervention / automatable tickets	Human handoff for high-risk cases should not be counted as failure
First contact resolution	Tickets resolved without repeated contact or rework / total tickets	Observation window shall be defined
Human handoff rate	Human-handoff tasks / total tasks	Very low handoff may indicate weak risk detection
Shutdown frequency	Number of shutdown triggers in a period	Zero shutdowns are not automatically good
Incorrect refund rate	Incorrect refund count / refund count	Track both count and amount
Integrated ROI	(Replacement cost saving + incremental benefit - TCO) / TCO	Attribution limits shall be disclosed

Appendix C: Go-Live Checklist

- [ ] Role description completed.
- [ ] Role SOP finalized and versioned.

- [ ] KPI/SLA defined.
- [ ] Authorization boundary and prohibited actions defined.
- [ ] Key interfaces configured with least privilege.
- [ ] Test set covers routine, edge, adversarial, and regression samples.
- [ ] Offline test passed.
- [ ] Sandbox test passed.
- [ ] Shadow run completed.
- [ ] Gray release completed.
- [ ] Human oversight mechanism configured.
- [ ] Audit logs can be replayed.
- [ ] Shutdown and rollback exercises completed.
- [ ] User human-handoff path available.
- [ ] Business, engineering, and risk teams jointly approve launch.

Appendix D: Conformance Statement Template

Item	Requirement
Declaring party	Enterprise, supplier, or joint declaring parties
System name and version	Product name, deployment version, model version, knowledge-base version, policy version
Applicable role	Customer service, after-sales refund, product operations, marketing operations, content/video, fulfillment coordination, etc.
Scope	Channels, marketplaces, languages, business flows, user types, and order types
Out of scope	Scenarios that shall be escalated to humans or prohibited
Claimed class	C0-C4, plus action-level L1-L5
Satisfied clauses	DEF, ROLE, ARCH, CAP, LEVEL, EVAL, GOV, IMPL, CONF clause list
Evaluation evidence	Test set version, sample size, test stages, metric results, production observation period
Governance evidence	Permission configuration, audit samples, shutdown exercise, rollback mechanism, human oversight configuration
Exemptions	Unsatisfied or non-applicable clauses, rationale, compensating controls

Item	Requirement
Accountable owners	Business owner, system owner, risk/compliance owner, human supervision owner
Validity	Statement validity period and review date

Example statement:

> **This system claims conformance to Class C2 under the Yepai E-commerce Digital Employee Standard White Paper YEP-DE-EC-2026-001. The claim is limited to the cross-border customer service role for logistics lookup, order status explanation, and standard return-policy explanation. Logistics lookup is L4, return-policy explanation is L3, and compensation commitments and marketplace complaint responses are out of scope.**

Appendix E: External Source Register

Table 8: External source register

Claim area	Figure or source position	Source basis	Standard use
Organizational AI adoption	88% organizational AI adoption in 2025; 70% of organizations used generative AI in at least one business function	Stanford HAI AI Index Report 2026, Economy chapter [R15]	Industry background
AI agent deployment	Agent deployment remained in the single digits across nearly all business functions	Stanford HAI AI Index Report 2026, Economy chapter [R15]	Adoption maturity baseline
Agentic capability	OSWorld task success rose from roughly 12% to 66.3%	Stanford HAI AI Index Report 2026, Technical Performance chapter [R15]	Level and authorization rationale
Productivity effects	Gains are strongest in structured, measurable work; examples include customer support, software development, and marketing output	Stanford HAI AI Index Report 2026, Economy chapter [R15]	Role selection and evaluation rationale
AI incidents	Documented incidents rose from 233 in 2024 to 362 in 2025	Stanford HAI AI Index Report 2026, Responsible AI chapter [R15]	Governance and shutdown rationale

Claim area	Figure or source position	Source basis	Standard use
Generative AI economic potential	USD 2.6 trillion to USD 4.4 trillion annually; about 75% of value in customer operations, marketing and sales, software engineering, and R&D	McKinsey Global Institute, 2023 [R16]	Industry background
Workforce transition	170 million roles created, 92 million displaced, net 78 million by 2030; 86% of employers expect AI and information processing to transform business	World Economic Forum Future of Jobs Report 2025 [R17]	Organizational adoption requirements
AI governance frameworks	NIST AI RMF 1.0, ISO/IEC 42001:2023, and EU AI Act mapping	NIST [R10], ISO [R18], European Commission [R19]	Governance mapping

References

Research Reports, Standards, and Papers

[R1] Microsoft. Work Trend Index Annual Report 2026.

[R2] Anthropic. Building Effective AI Agents: Architecture Patterns and Implementation Frameworks.

[R3] Google Cloud. AI Agent Trends 2026.

[R4] World Economic Forum. AI Agents in Action: Foundations for Evaluation and Governance.

[R5] World Economic Forum. AI Agents in Action: A Playbook for Trusted Adoption, Authorization and Scaling.

[R6] Prosus. The Rise of the Agentic Workforce.

[R7] Prosus. The Coming Age of AI Colleagues.

[R8] Visa. The Rise of Agentic Commerce.

[R9] Agentic Commerce Protocol. Agentic Commerce for Developers: Technical Whitepaper.

[R10] NIST. Artificial Intelligence Risk Management Framework 1.0.

[R11] Salesforce/IDC. The Tipping Point: How Agentic AI Is Redefining the Future of Work.

[R12] Agashe et al. Agent S: An Open Agentic Framework that Uses Computers Like a Human. ICLR 2025.

[R13] Hu, Lu, Clune. Automated Design of Agentic Systems. ICLR 2025.

[R14] Allouah et al. What Is Your AI Agent Buying? Evaluation, Biases, Model Dependence, and Emerging Implications for Agentic E-Commerce.

[R15] Stanford Institute for Human-Centered Artificial Intelligence. AI Index Report 2026. Economy, Technical Performance, and Responsible AI chapters.

[R16] McKinsey Global Institute. The Economic Potential of Generative AI: The Next Productivity Frontier. 2023.

[R17] World Economic Forum. The Future of Jobs Report 2025.

[R18] ISO/IEC 42001:2023. Information technology - Artificial intelligence - Management system.
[R19] European Commission. AI Act implementation and regulatory framework materials, including the official application timeline.

Public Web References

[W1] McKinsey. The agentic commerce opportunity: How AI agents are ushering in a new era for consumers and merchants. 2025.
[W2] Bain & Company. 2030 Forecast: How Agentic AI Will Reshape US Retail. 2025.
[W3] BCG. AI Agents: What They Are and Their Business Impact.
[W4] IBM. Digital Worker and Agentic Commerce related research.
[W5] OpenAI and Stripe / Agentic Commerce Protocol public resources.