WHITE PAPER

Agentic orchestration for ANZ professional services.

Trust, guardrails, and progressive autonomy in accounting, tax, audit, and compliance advisory.

PUBLISHED BY
Zato Origin Limited
Auckland, NZ

EDITION
Q2 2026 · Version 2.1

A Dual-Zone Architecture for deploying autonomous sub-agents and governed decision agents across regulated workflows in New Zealand and Australia.

PUBLISHED BY

Zato Origin Limited
Level 2, 125 The Strand
Parnell, Auckland, New Zealand
zatohq.com

AUTHORS

Saurav J Bansal
Dr Srinivas Kishan Anapu
Sanjeev Reddy Bora

ENQUIRIES

All enquiries and questions to team@zatohq.com.

EDITION

Q2 2026 · Version 2.1 · Classification: Public

CITATION

Zato Origin Limited. *Agentic Orchestration for ANZ Professional Services: Trust, Guardrails, and Progressive Autonomy in Accounting, Tax, Audit, and Compliance Advisory.* Q2 2026, Version 2.1. Auckland, NZ.

CONTENTS

What's inside.

Ten sections covering the Trust Imperative, the Dual-Zone Architecture, the four-stage Progressive Autonomy Model, the Job Orchestrator, the Trust & Governance framework, security posture, alignment with international standards, applied use cases, future directions, and conclusion.

10	7	3	19
SECTIONS	DIAGRAMS	DATA TABLES	REFERENCES

EXECUTIVE SUMMARY

Trust is the precondition. Architecture is the answer.

Trust is not a feature you market. It is the precondition for deploying AI agents in accounting, tax, audit, and compliance advisory work. Firms in New Zealand and Australia carry statutory duties. Their AI must carry the same evidentiary burden as a human professional.

This paper sets out Zato's framework for agentic orchestration in ANZ professional services. It is built on a Dual-Zone Architecture that splits all agent work into two operating domains, governed differently, staffed by different agent types, and bridged by an explicit handover protocol.

The **Autonomous Execution Zone** runs preparatory, analytical, and classification work. Sub-agents gather data, normalise documents, classify transactions, draft workpapers, and detect anomalies. They operate at speed under sampling-based oversight. The **Governed Decision Zone** runs judgment work: opinions, regulatory submissions, client-facing advice. Decision agents operate under progressive autonomy, execution guardrails, and risk-tiered human review.

A four-stage autonomy model governs how decision agents earn independence: Shadow, Assisted, Supervised, Constrained. Each stage names the agent's scope and the human role. Independence is earned by demonstrated performance. It is never assumed.

The Job Orchestrator sequences work across both zones using version-controlled templates. Each template is a directed acyclic graph of task nodes, zone transitions, quality gates, human-in-the-loop checkpoints, and rollback paths. The Trust and Governance framework wraps every action: execution guardrails, risk-tiered HITL controls, and reasoning traceability that produces a regulator-ready audit trail.

The question is not whether an AI agent can perform the task. The question is whether it should perform the task without human governance, which components are safe to automate, and which require professional judgment.

> The question is not whether an agent can perform the task. It is whether it should, without human governance.

01 SECTION · TRUST IMPERATIVE

From chatbots to autonomous agents: why trust cannot be assumed: and the regulatory landscape in ANZ that makes traceable, verifiable agent output a precondition for deployment.

The Trust Imperative.

1.1

From Chatbots to Autonomous Agents.

AI in professional services has moved through three phases. The first ran on rule-based scripts and robotic process automation. It executed without context. The second introduced large language models as conversational assistants. They answered questions. They did not act. The third phase is operational now: agentic systems that plan, reason, execute multi-step workflows, and interact with external tools and data sources without supervision at every step.

SOURCE *Gartner*, "Multiagent Systems in Enterprise AI," December 2025; *IBM Think*, "AI Agents in 2025: Expectations vs. Reality," November 2025.

This shift changes the risk profile. When an AI agent can prepare a tax filing, trigger a compliance review, or generate an audit workpaper, the question is not whether it can perform the task, but whether it should perform the task without human governance: and if so, which specific components of the task are safe for autonomous execution and which require human judgment.

1.2

Why Trust Cannot Be Assumed.

In accounting, tax, audit, and compliance advisory, trust is not a feature to be marketed: it is a precondition for deployment. Professional services firms operate under statutory obligations, professional standards (ISA, ASA, NZ IFRS, AASB), and fiduciary duties that require demonstrable diligence in every action taken on a client's behalf. An AI agent operating within these environments must meet the same evidentiary burden as a human professional.

Research from the OECD's AI Policy Observatory and NIST's AI Risk Management Framework (AI RMF 1.0) emphasises that trustworthy AI systems must be transparent, accountable, secure, and subject to human oversight: principles that are especially critical when AI systems move from advisory roles to execution roles.

SOURCE *OECD AI Policy Observatory*, "Trustworthy AI Principles," 2024; *NIST AI 100-1*, "AI Risk Management Framework," January 2023.

1.3

The ANZ regulatory landscape.

In New Zealand, Inland Revenue's intensified enforcement posture, backed by NZ$116 million in additional funding over four years from the 2024 Budget, and the provisions of the Taxation (Annual Rates for 2025–26, Compliance Simplification, and Remedial Measures) Bill, demand that AI-assisted compliance output be traceable, verifiable, and defensible.

In Australia, the AASB Climate Reporting and Assurance Standards require regulated entities to lodge climate-related financial disclosures with audit-grade evidence trails from FY2025 onwards. The ATO's expanded compliance programmes raise the documentary burden on tax-effect accounting and transfer pricing positions. Firms deploying AI agents on tax compliance, audit support, or advisory work face one bar across both jurisdictions: every agent action is governed, auditable, and subject to human oversight proportionate to the risk and regulatory sensitivity of the specific task.

SOURCE *CPA Australia*, "Well-funded IRD drives an aggressive compliance approach," 2025. *AASB Climate Reporting and Assurance Standards*, 2025. *NZ Inland Revenue*, Taxation (Annual Rates for 2025–26, Compliance Simplification, and Remedial Measures) Bill Commentary, August 2025.

NZ$116m
IRD ENFORCEMENT FUNDING
NZ BUDGET 2024 · 4-YEAR

2025–26
TAXATION BILL
IN FORCE

AASB
CLIMATE REPORTING
ASSURANCE · AU

> Every agent action must be governed, auditable, and subject to human oversight proportionate to risk.

02 SECTION · DUAL-ZONE ARCHITECTURE

Two operationally distinct domains: one for speed, one for judgment: separated by an explicit boundary and bridged by a structured handover protocol.

The Dual-Zone Architecture.

A single architectural call decides where agentic AI is safe in professional services: where autonomous execution is allowed, and where governed human oversight is essential.

Zato addresses this through a Dual-Zone Architecture that classifies all agent work into two operationally distinct domains, each with its own governance model, oversight requirements, and agent types.

Zone 01

AUTONOMOUS EXECUTION
"UNDERWATER" · SPEED + EFFICIENCY

Zone 02

GOVERNED DECISION
"SURFACE" · JUDGMENT + ACCOUNTABILITY

FIGURE 01 · OVERLEAF

ZATO DUAL-ZONE MASTER ARCHITECTURE.

A single reference diagram showing the Job Orchestrator, the two zones, the Handover Bundle that bridges them, and the governance, security and routing layers that enforce the architecture programmatically.

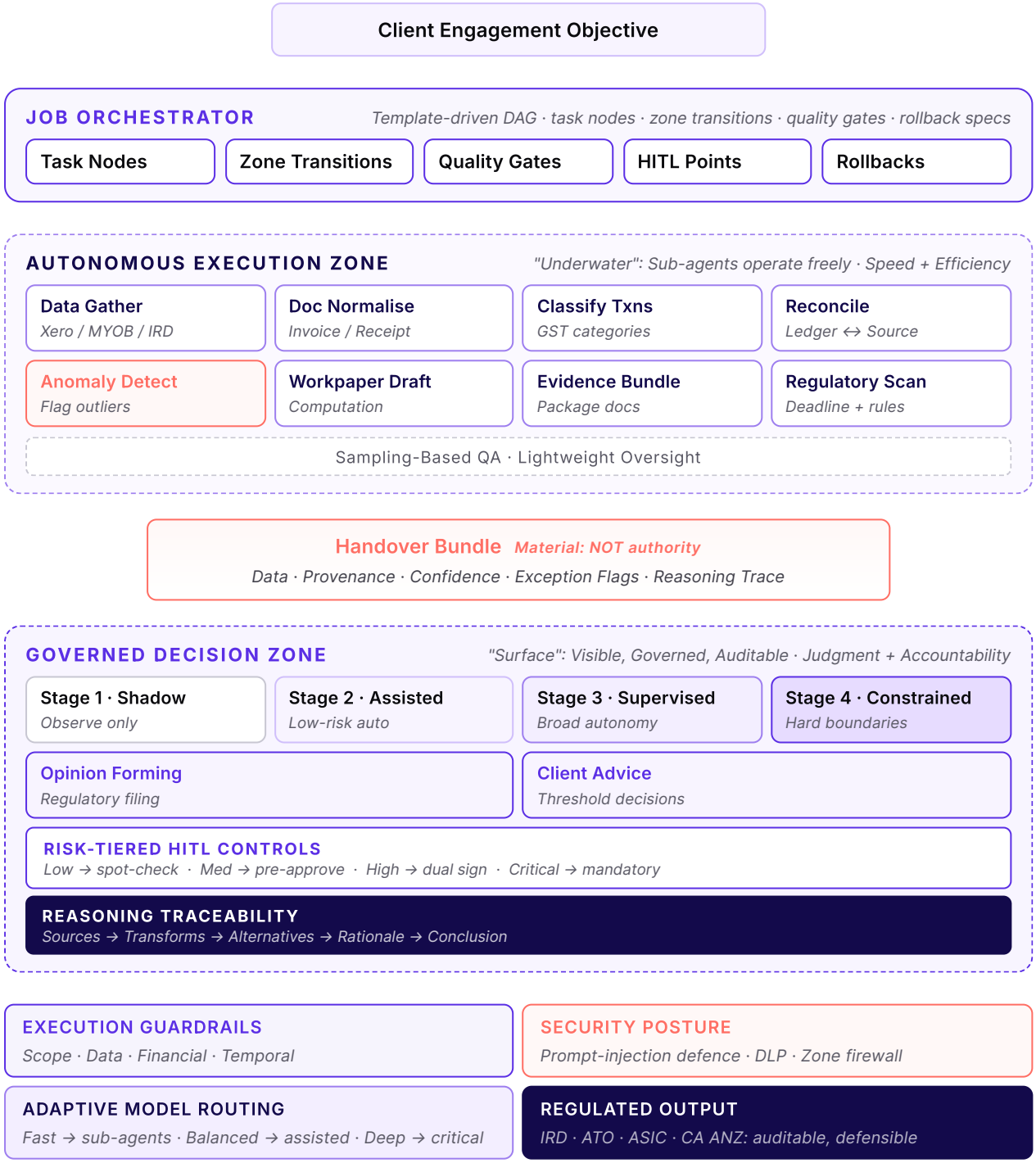
HOW TO READ IT

The Autonomous Execution Zone (left) hosts sub-agents that prepare material. The Governed Decision Zone (right) hosts decision agents that exercise professional judgment. The Handover Bundle in the middle is a transfer of *material*, not *authority*: every decision agent reviews sub-agent work as a senior practitioner reviews a junior's.

FIGURE 01

Zato Dual-Zone Agentic Orchestration.

Autonomous sub-agents + governed decision agents across regulated workflows.



2.1

The Classification Principle.

Not all tasks in accounting, tax, audit, and compliance carry equal regulatory risk or require equal professional judgment. Preparing a bank reconciliation is different work from forming an audit opinion. Normalising invoices for GST classification is different work from advising on a cross-border restructuring. The Dual-Zone Architecture makes this distinction explicit and enforceable.

The classification principle operates on a simple test: does the task require professional judgment, form a legally consequential opinion, or produce an output that will be communicated externally to a client, regulator, or third party without further human review? If the answer is no, the task falls within the Autonomous Execution Zone. If the answer is yes, the task falls within the Governed Decision Zone.

THE CLASSIFICATION TEST

Does the task require professional judgment, form a legally consequential opinion, or produce an output that will be communicated externally without further human review?

No → Autonomous Execution Zone. Yes → Governed Decision Zone.

2.2

The Autonomous Execution Zone.

The Autonomous Execution Zone encompasses preparatory, analytical, and classification tasks. Sub-agents are specialised, lightweight agents optimised for speed and throughput. They run with minimal human oversight. These are tasks in which the agent gathers, organises, classifies, or prepares material for subsequent human or governed-agent review. The outputs of these tasks feed into the Governed Decision Zone but do not, by themselves, constitute professional advice, regulatory submissions, or client-facing opinions.

Zato defines eleven categories of regulated-safe autonomous work within the ANZ professional services context: set out overleaf in Table 1.

TABLE 01

Autonomous Execution Zone: Regulated-safe task categories.

Eleven categories of preparatory, analytical, and classification work that sub-agents perform under sampling-based oversight.

#	TASK CATEGORY	DESCRIPTION
01	**Gathering Data**	Collecting and ingesting structured and unstructured data from connected source systems and API endpoints
02	**Normalising Documents**	Standardising document formats, extracting key fields, and structuring unstructured content into consistent schemas
03	**Reconciling Records**	Matching transactions across systems and identifying discrepancies for review against source ledgers
04	**Classifying Transactions**	Categorising financial transactions against chart of accounts, regulatory taxonomies, or supply type classifications
05	**Drafting Workpapers**	Preparing structured analytical workpapers with supporting schedules, cross-references, and computation details
06	**Preparing Checklists**	Generating regulatory and procedural checklists with completion status tracking and gap identification
07	**Proposing Exceptions**	Identifying items that fall outside normal parameters and flagging them for human evaluation and disposition
08	**Detecting Anomalies**	Applying statistical and pattern-based analysis to identify outliers, irregularities, and unusual transaction patterns
09	**Routing Cases**	Assessing incoming work items and directing them to appropriate specialist agents or human reviewers based on domain and risk
10	**Preparing First-Pass Narratives**	Generating draft analytical commentary and explanatory narratives for human refinement and professional review
11	**Generating Evidence Bundles**	Assembling and packaging supporting documentation into structured bundles for human review and sign-off

These eleven categories share a common characteristic: they produce intermediate outputs that inform professional judgment but do not themselves constitute that judgment. A sub-agent that classifies a transaction as zero-rated for GST purposes *proposes* the classification; a human practitioner (or a governed decision agent operating under HITL controls) *confirms* it. A sub-agent that drafts an audit workpaper *prepares* a document; the engagement team *reviews*, *adjusts*, and *approves* it.

2.3

The Governed Decision Zone.

The Governed Decision Zone encompasses all tasks that involve professional judgment, opinion formation, regulatory submission, or client-facing communication. These tasks are performed by governed decision agents operating under the progressive autonomy model (Section 3), subject to execution guardrails, risk-tiered HITL controls, and full traceability of reasoning.

Examples of Governed Decision Zone activities in ANZ professional services include:

- **Opinion-forming:** Audit opinions, tax position conclusions, compliance assessments, advisory recommendations.
- **Regulatory submissions:** Tax return filing, GST/BAS lodgement, suspicious activity reports, and statutory filings.
- **Client communications:** Advice letters, engagement deliverables, management reports, board presentations.
- **Judgment-dependent analyses:** Transfer pricing determinations, materiality assessments, going-concern evaluations, restructuring advice.
- **Threshold decisions:** Whether to escalate a matter, qualify an opinion, disclose a risk, or recommend a course of action.

2.4

The Handover Mechanism: From Sub-Agent to Decision Agent.

The boundary between the Autonomous Execution Zone and the Governed Decision Zone is enforced through a structured handover protocol managed by the Job Orchestrator. When a sub-agent completes its autonomous task, it packages its outputs into a standardised handover bundle that includes the processed data or document, a provenance record detailing sources and applied transformations, a confidence score for proposed classifications or judgments, and any flags or exceptions identified during processing.

FIGURE 02

Zone Handover Protocol: From Sub-Agent to Decision Agent.

Transfer of prepared material: not transfer of authority.

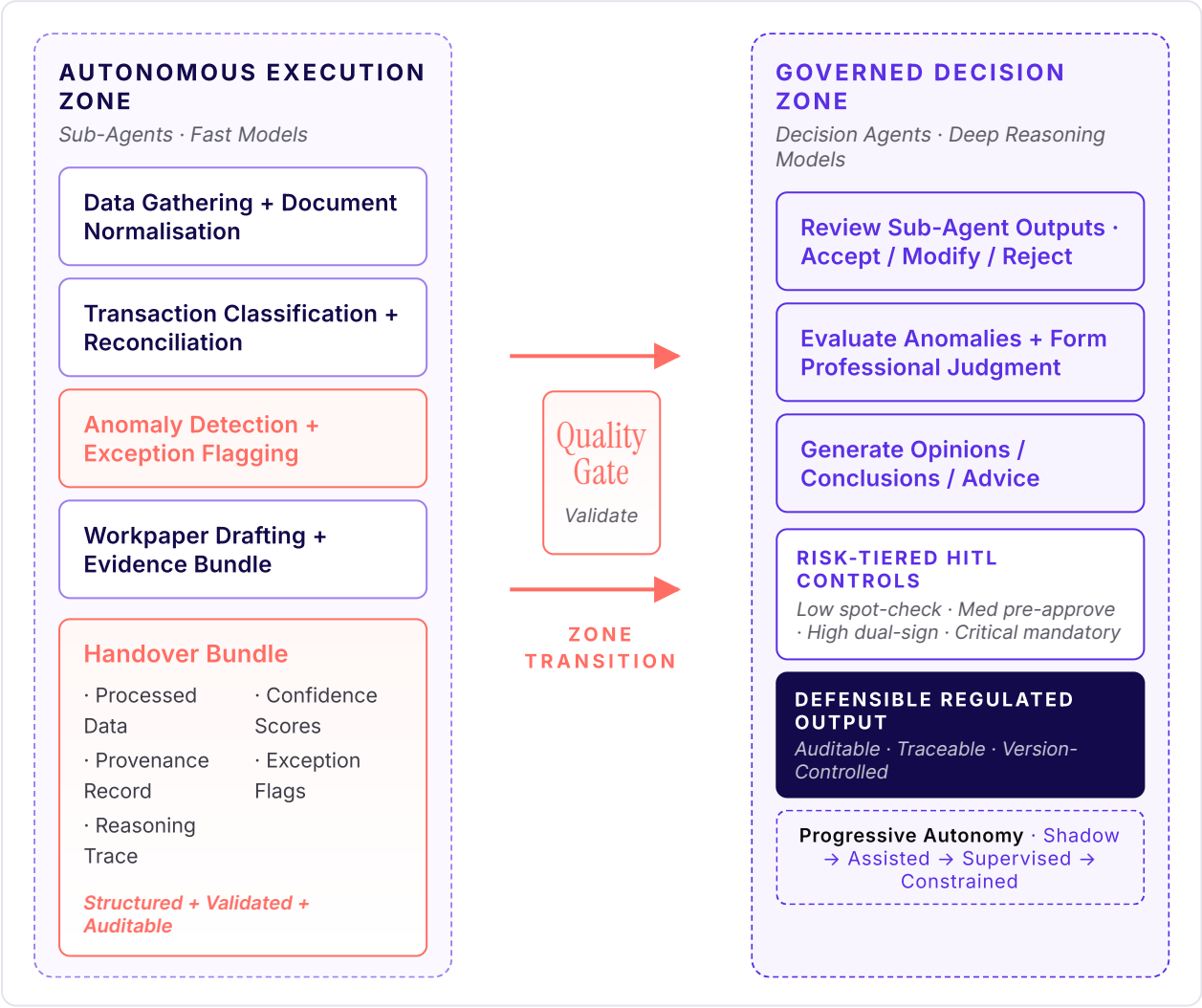


FIGURE 02 *The Handover Bundle is a transfer of prepared material: the decision agent reviews it as a senior practitioner reviews a junior's work.*

The governed decision agent receives this handover bundle as its input, along with the full audit trail of the sub-agent's work. Critically, the decision agent is not bound by the sub-agent's proposed classifications or drafts: it reviews them as a senior practitioner would review a junior's work, accepting, modifying, or rejecting as professional judgment requires. The handover is a transfer of prepared material, not a transfer of authority.

SOURCE *Gartner*, "Multiagent Systems in Enterprise AI," December 2025; *ISACA*, "No Looking Back: Transforming Audit with Artificial Intelligence," June 2025.

03 SECTION · PROGRESSIVE AUTONOMY

A four-stage maturity model: Shadow, Assisted, Supervised, Constrained: through which decision agents earn independence by demonstrated performance.

Progressive Autonomy.

Within the Governed Decision Zone, Zato implements a four-stage maturity model that progressively scales agent autonomy as trust is earned through demonstrated performance, compliance, and governance maturity.

This model applies to decision agents performing judgment-dependent, opinion-forming, or legally consequential work. It does not apply to the sub-agents operating within the Autonomous Execution Zone. Sub-agents are governed by the lighter oversight model described in Section 2.2.

FIGURE 03

Trust is earned through demonstrated performance: not assumed by default.

Four-stage progressive autonomy model.

IMPORTANT: SCOPE OF THIS MODEL

This model applies **only** to the Governed Decision Zone. Sub-agents in the Autonomous Execution Zone operate under lightweight sampling-based oversight from deployment.

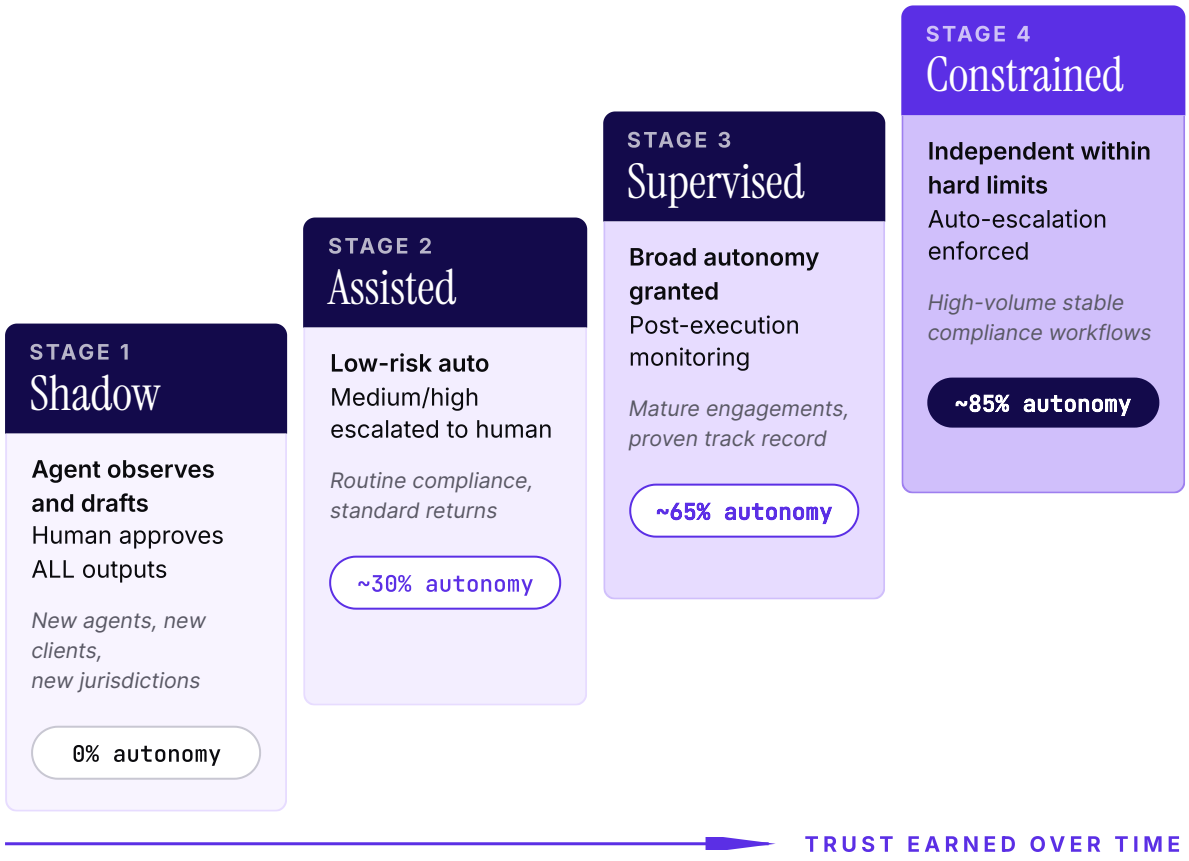


FIGURE 03 *Trust is earned through demonstrated performance: not assumed by default.*

3.1 · STAGE 1

Shadow Mode.

In Shadow Mode, decision agents observe and analyse but do not act on governed tasks. The agent processes incoming handover bundles from sub-agents, generates recommendations, and produces draft outputs, but every deliverable is presented to a human professional for review before any external action is taken. The agent's outputs are logged alongside the human's final decisions, creating a comparative record that enables performance benchmarking over time.

- **Agent capability:** Full analytical processing of sub-agent handover bundles, recommendation generation, draft output creation.
- **Human role:** Mandatory review and approval of every output before execution or external communication.
- **Trust mechanism:** Performance comparison between agent recommendations and human decisions, building a statistical trust baseline.
- **Applicable contexts:** Initial deployment of any new decision agent type; first engagement with a new client; entry into a new regulatory jurisdiction.

3.2 · STAGE 2

Assisted Mode.

In Assisted Mode, the decision agent begins to execute lower-risk governed tasks autonomously while routing higher-risk or ambiguous tasks for human approval. The boundary between autonomous execution and human review is determined by a risk classification engine that evaluates each task against predefined criteria including regulatory sensitivity, financial materiality, client risk profile, and task complexity.

- **Agent capability:** Autonomous execution of lower-risk governed tasks; flagging and routing of medium and high-risk tasks for human review.
- **Human role:** Approval required for tasks above the defined risk threshold; periodic spot-checking of autonomous executions.
- **Trust mechanism:** Risk-tiered escalation policies; error rate monitoring; client feedback integration.
- **Applicable contexts:** Routine compliance conclusions; standard tax return sign-off components; well-precedented advisory responses.

3.3 · STAGE 3

Supervised Mode.

In Supervised Mode, the decision agent operates with broad autonomy across most governed task categories, with human oversight shifting from pre-approval to post-execution review. The human professional monitors agent performance through dashboards, exception reports, and audit trails rather than reviewing every individual output.

- **Agent capability:** Broad autonomous execution across standard and moderate-complexity governed tasks; self-flagging of edge cases.
- **Human role:** Post-execution monitoring; exception-based intervention; periodic quality assurance reviews.
- **Trust mechanism:** Statistical process control; anomaly detection; client outcome tracking; regulatory compliance audits.
- **Applicable contexts:** Mature client engagements with established agent performance history; well-understood regulatory domains.

3.4 · STAGE 4

Constrained Autonomous Mode.

Constrained Autonomous Mode is the highest level of decision agent independence within Zato's framework. It is not unconstrained autonomy. The agent operates independently within explicitly defined boundaries, with hard guardrails that prevent execution outside its sanctioned scope. Actions that exceed defined thresholds are automatically escalated regardless of the agent's confidence level.

- **Agent capability:** Independent operation within defined boundaries; automatic escalation when boundaries are approached or exceeded.
- **Human role:** Governance oversight; boundary definition and adjustment; escalation handling; strategic decision-making.
- **Trust mechanism:** Hard execution boundaries; automatic escalation triggers; full audit trails; regulatory compliance certification.
- **Applicable contexts:** High-volume, well-defined compliance workflows with established performance baselines and stable regulatory requirements.

TABLE 02

Progressive autonomy stages for governed decision agents.

A summary of capability, human role, risk scope and trust basis at each of the four stages.

STAGE	AGENT CAPABILITY	HUMAN ROLE	RISK SCOPE	TRUST BASIS
SHADOW	Observe, analyse, recommend from sub-agent outputs	Review and approve all outputs	No autonomous action	Performance benchmarking
ASSISTED	Execute lower-risk governed tasks autonomously	Approve medium/high-risk tasks	Low-risk governed tasks	Risk-tiered escalation
SUPERVISED	Broad autonomous execution of governed tasks	Post-execution monitoring and exception review	Standard and moderate	Statistical process control
CONSTRAINED	Independent within hard boundaries	Governance, boundary setting, escalation	All within hard limits	Boundary enforcement

> Agent independence is earned through demonstrated performance: never assumed by default.

04 SECTION · JOB ORCHESTRATOR

From ad hoc prompts to structured workflows: version-controlled job templates as DAGs with task nodes, zone transitions, quality gates, HITL points and rollback paths.

The Job Orchestrator.

4.1

What a Job Template is.

A Job Template is the complete specification of one piece of professional work, expressed as a directed acyclic graph (DAG) of agent tasks. It names every sub-agent action, every decision-agent action, every zone transition, every quality gate, every human checkpoint, and every rollback path. It is version-controlled. It is auditable. Every job that runs against the template inherits its governance.

Professional workflows follow defined methodologies with specific sequencing, quality gates, and documentation obligations. Free-form prompting cannot meet that bar. Templates encode the methodology directly into the architecture.

4.2

Template Architecture.

Each Job Template encapsulates a complete workflow as a directed acyclic graph (DAG) of agent tasks. Templates are defined declaratively and include the following components:

- **Task Nodes:** Individual agent actions with specified inputs, outputs, and success criteria. Each node is explicitly tagged as belonging to the Autonomous Execution Zone or the Governed Decision Zone.
- **Zone Transitions:** Explicit handover points where sub-agent outputs are packaged and transferred to governed decision agents, including the validation checks that must pass before handover is accepted.
- **Dependency Edges:** Sequencing constraints that prevent downstream tasks from executing before upstream prerequisites are complete.
- **Quality Gates:** Validation checkpoints that verify output integrity, regulatory compliance, and data consistency before the workflow progresses.
- **HITL Insertion Points:** Predefined positions within the Governed Decision Zone where human review and approval are mandatory, determined by the risk classification of the task and the current autonomy stage.
- **Rollback Specifications:** Defined procedures for reversing completed steps when downstream failures or quality gate rejections require workflow correction.

4.3

Workflow Example: GST Return Preparation.

The following diagram illustrates a complete GST return preparation workflow as a DAG, showing how sub-agent task nodes in the Autonomous Execution Zone feed through a quality gate into decision agent nodes in the Governed Decision Zone, with rollback paths and mandatory HITL checkpoints.

FIGURE 04 *Template-driven · version-controlled*

Job Orchestrator DAG: GST return preparation.

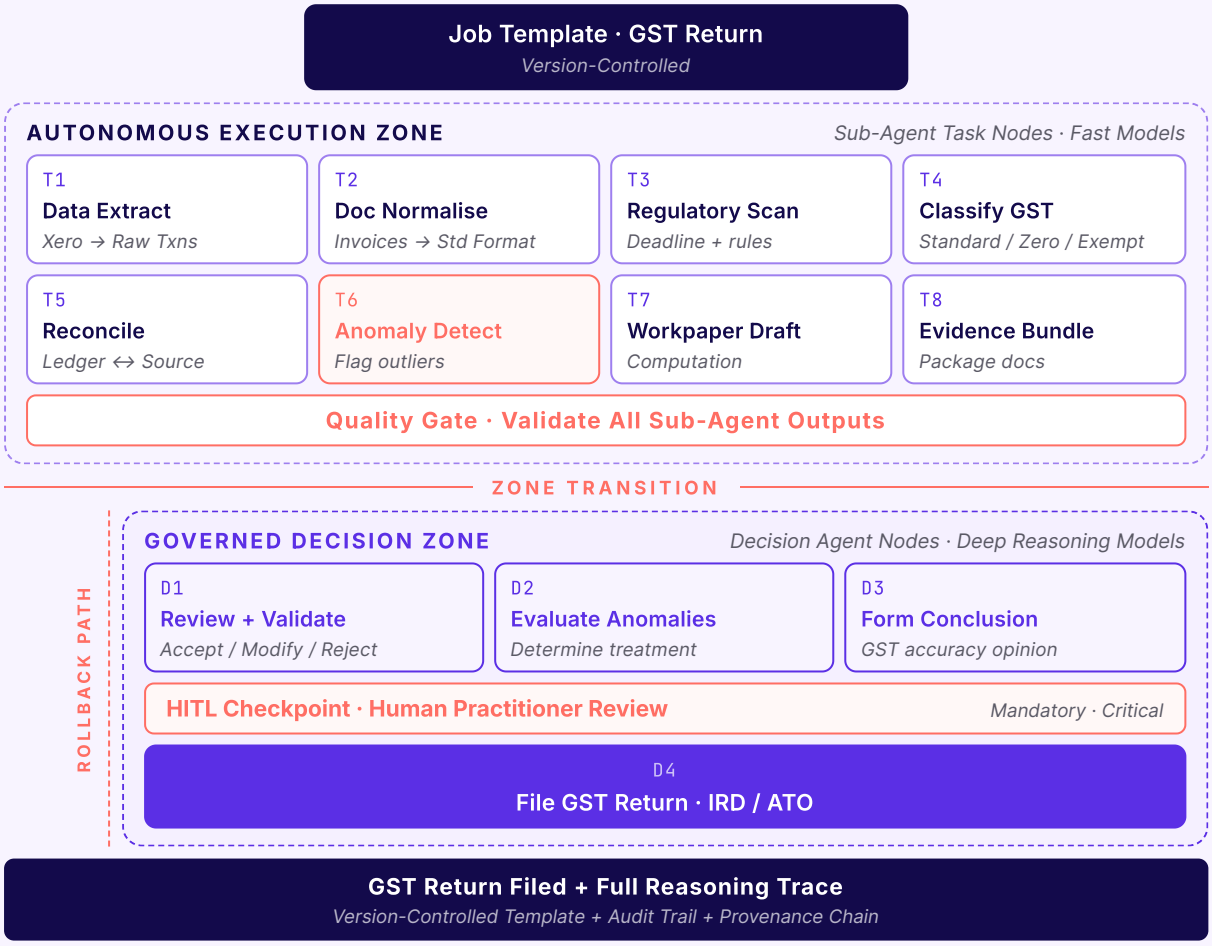


FIGURE 04 *DAG flows from version-controlled template → 8 sub-agent tasks → Quality Gate → 3 decision-agent tasks → HITL checkpoint → filing terminal, with a Rollback Path returning to the autonomous layer.*

4.4

Adaptive Model Routing Within Templates.

Building on Zato's adaptive model routing capability, the Job Orchestrator assigns different AI models to different task nodes based on the cognitive demands of each step. Sub-agents in the Autonomous Execution Zone are powered by fast, cost-efficient models optimised for throughput. Decision agents in the Governed Decision Zone engage balanced-reasoning or deep-reasoning models calibrated to the complexity of the governed task.

This mirrors the staffing model of professional services firms: data extraction and reconciliation are handled by capable but efficient team members, while complex advisory judgments engage senior specialists.

FAST / CHEAP

Sub-agents

Throughput-optimised models for data gathering, normalisation, classification and reconciliation.

BALANCED

Assisted decisions

Balanced-reasoning models for governed tasks at lower-risk tiers and Assisted-stage execution.

DEEP REASONING

Critical judgments

Deep-reasoning models for opinion formation, multi-jurisdictional analysis and Critical-tier decisions.

4.5

Template governance and version control.

Job Templates run under formal governance: version control, change-approval workflows, audit trail documentation. When a regulatory requirement changes, the template updates through a controlled process that keeps every active workflow compliant with current legislation. Historical template versions are retained for audit. Every job that ever ran against the template is traceable back to the exact template version that governed it.

This is the architecture that satisfies the ISO/IEC 42001 AI Management System standard. Section 7.1 returns to the alignment in detail. **Zato is certified to ISO 42001.**

SOURCE *Gartner*, "Emerging Practices in AI Governance for Regulated Industries," 2025; *ISACA*, "No Looking Back: Transforming Audit with Artificial Intelligence," June 2025.

05 SECTION · TRUST & GOVERNANCE

Execution guardrails, risk-tiered HITL controls, reasoning traceability: a programmatic, defence-in-depth framework that constrains every agent action.

Trust and Governance Framework.

FIGURE 05

Programmatic enforcement across both zones.

Trust and Governance Framework: Layered enforcement.

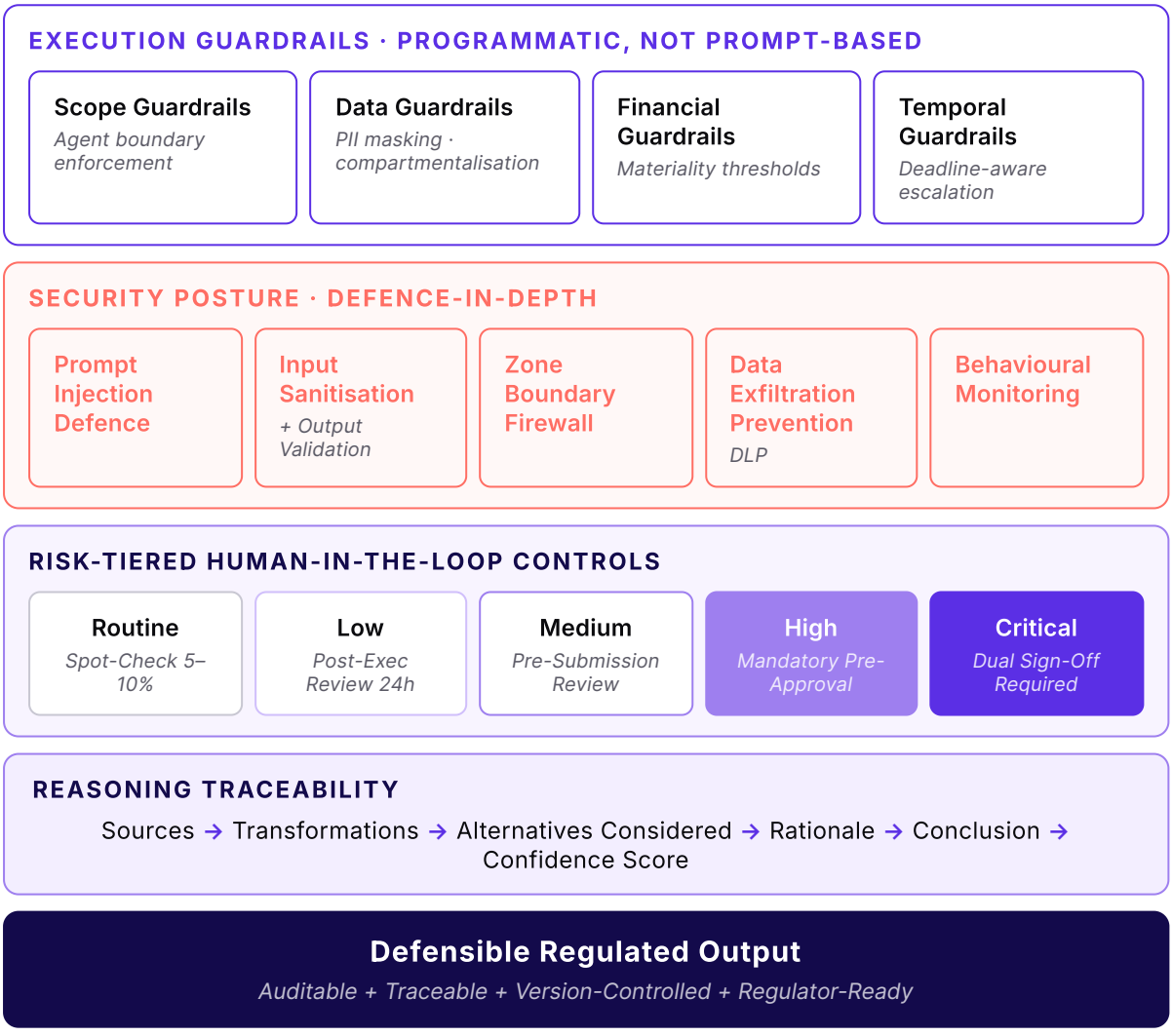


FIGURE 05 *Four enforcement layers converge on a defensible regulated output. Guardrails, security posture, HITL controls, reasoning traceability.*

5.1

Execution Guardrails.

Execution guardrails are hard constraints that govern agent behaviour at runtime across both zones. Unlike guidelines or recommendations, guardrails are enforced programmatically and cannot be overridden by the agent's own reasoning. Zato implements guardrails across four dimensions:

- **Scope Guardrails:** Define the boundaries of what each agent type is permitted to do. A sub-agent tasked with transaction classification cannot autonomously escalate to forming a tax opinion. A decision agent tasked with GST compliance cannot autonomously initiate a transfer pricing analysis.
- **Data Guardrails:** Restrict the data sources an agent can access and the data it can output. Client confidential information is compartmentalised, preventing cross-client data contamination. PII is subject to masking and access controls aligned with the Privacy Act 2020 (NZ) and the Privacy Act 1988 (AU).
- **Financial Guardrails:** Impose materiality thresholds that trigger mandatory human review. Any agent-generated output involving financial figures above defined thresholds requires human validation before external communication or filing.
- **Temporal Guardrails:** Enforce deadline awareness and prevent agents from executing time-sensitive actions outside approved windows. Filing deadline proximity triggers escalation protocols that increase human oversight.

WHY PROGRAMMATIC, NOT PROMPT-BASED

Guardrails are enforced in code at runtime: they cannot be argued with, talked around, or overridden by the agent's own reasoning. Prompts are guidance; guardrails are the wall.

5.2

Risk-Tiered Human-in-the-Loop (HITL) Controls.

Zato's HITL framework implements a risk-tiered model where the intensity and nature of human involvement is calibrated to the risk profile of each specific task. HITL controls apply primarily within the Governed Decision Zone, while the Autonomous Execution Zone operates under sampling-based quality assurance.

TABLE 03

Risk-tiered HITL control matrix.

RISK TIER	ZONE	EXAMPLE TASKS	HITL REQUIREMENT	ESCALATION PATH
ROUTINE	Autonomous Execution	Data gathering, normalisation, reconciliation, classification	Post-execution sampling (5–10%)	Team lead
LOW	Governed Decision	Standard compliance conclusions, routine advisory	Post-execution review within 24h	Senior practitioner
MEDIUM	Governed Decision	Tax return preparation, compliance assessments	Pre-submission review	Manager / Senior Manager
HIGH	Governed Decision	Advisory opinions, multi-jurisdictional analysis	Mandatory pre-approval	Partner / Director
CRITICAL	Governed Decision	Audit sign-off, regulatory submissions, formal advice	Dual review with sign-off	Partner + compliance

> The intensity of human involvement is calibrated to the risk profile of each specific task: proportionate, not uniform.

5.3

Reasoning Traceability.

Every agent action within Zato generates a structured reasoning trace. This applies in both the Autonomous Execution Zone and the Governed Decision Zone. The trace documents the agent's analytical process from input to output.

FIGURE 06

Every agent action generates a structured, auditable reasoning chain.

Anatomy of a reasoning trace.

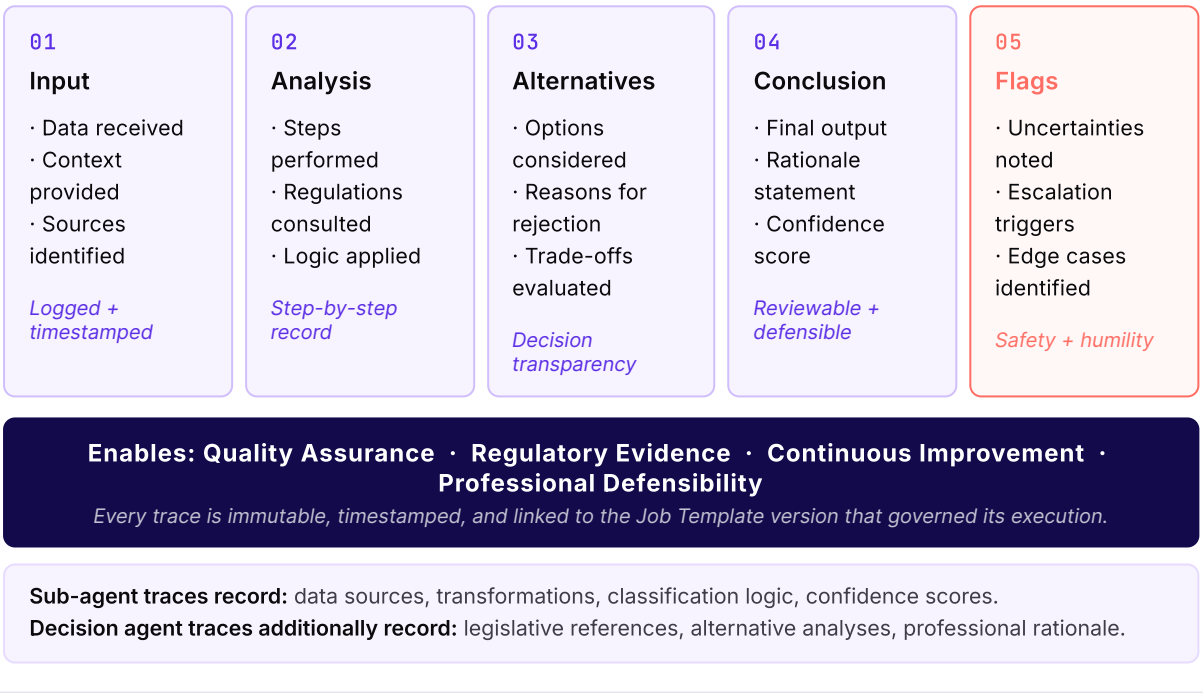


FIGURE 06 *Five-stage chain: Input → Analysis → Alternatives → Conclusion → Flags: immutable and linked to a Job Template version.*

For sub-agents, this trace records data sources accessed, transformations applied, classification logic, and confidence scores. For decision agents, the trace additionally records the legislative and regulatory references consulted, alternative analyses considered and reasons for rejection, and the rationale for the final professional conclusion.

Reasoning traces are not optional artefacts. They are the audit evidence that satisfies **ISO 27001:2022** information security audit requirements, the documentation requirements of **SOC 2**, and the AI lifecycle traceability requirements of **ISO 42001**. A regulator, a professional body, or an external auditor can pull any agent-assisted output and reconstruct exactly how it was produced.

SOURCE *NIST AI 100-1*, "AI Risk Management Framework," January 2023; *OECD*, "Advanced AI Assistants: Governance Challenges," 2024.

06 SECTION · SECURITY POSTURE

Zato runs to commercial-bank security standards. ISO 27001:2022, SOC 2, ISO 9001, AWS data residency, six-monthly penetration tests, no client data on external models.

Security posture.

6.1

The threat surface.

Prompt injection is the highest-risk class of attack against agentic AI in regulated environments. It exploits the natural-language interface through which agents receive instructions. An attacker embeds malicious instructions inside seemingly benign data (a client document, an email, a data feed) and the agent acts on them. Classification of a tax position is altered. A workpaper is poisoned. A filing is misdirected.

In professional services the consequences are direct: confidential client data, legally consequential outputs, regulator-facing submissions. The Dual-Zone Architecture is the first defence layer. Sub-agents in the Autonomous Execution Zone have limited scope and cannot escalate to governed actions. A compromised sub-agent cannot file a tax return, issue professional advice, or invoke a decision agent. The zone boundary is the architectural firewall.

SOURCE *OWASP*, "Top 10 for Large Language Model Applications," Version 1.1, 2024.

6.2

Zato's certifications and operational standards.

Zato runs to commercial-bank security and quality standards. The architecture described in this paper is wrapped in independently audited controls.

- **ISO 27001:2022 · Information Security Management.** Audited annually. Covers asset management, access control, cryptography, operations security, supplier relationships, incident management, business continuity.
- **ISO 42001 · AI Management System.** Zato is certified. ISO 42001 is the international standard for governing AI systems across the lifecycle: governance, risk, lifecycle controls, transparency, human oversight.
- **ISO 9001 · Quality Management.** Process discipline across the engineering, delivery, and support functions.
- **SOC 2 (Type 2).** Attestation on the trust service criteria: security, availability, processing integrity, confidentiality, privacy.
- **GDPR compliant.** Aligned with the EU General Data Protection Regulation for European data subjects.
- **Privacy alignment.** NZ Privacy Act 2020 and AU Privacy Act 1988, with regional data residency.

6.3

Operational facts.

DATA RESIDENCY.

NZ firm data is hosted on AWS NZ (Auckland region). AU firm data is hosted on AWS AU (Sydney region). Client data does not cross national boundaries without explicit, contracted consent. This answers the most-asked due diligence question on day one.

NO EXTERNAL MODEL TRAINING.

Ziffy, Zato's built-in AI assistant, runs against locally hosted inference. No client data is sent to external model providers. No client data is used to train any external model. The entire reasoning loop sits inside the customer's tenancy.

PENETRATION TESTING.

Independent third-party penetration testing runs every six months at minimum, and after every major release. The OWASP Top 10 for LLM Applications is part of the test surface. Reports are made available to firms under non-disclosure on request.

DEFENCE IN DEPTH.

Input sanitisation runs first. The instruction hierarchy is strict: system-level instructions defined in the Job Template cannot be overridden by operator-level or data-level inputs. Zone boundary enforcement means sub-agents cannot invoke decision-agent capabilities or governed action endpoints, regardless of what instructions arrive. Outputs are validated against expected patterns, regulatory constraints, and historical baselines before they execute. Runtime behavioural monitoring runs continuously. Each agent operates inside its own sandboxed environment with defined resource access.

AUDIT-GRADE BY DEFAULT

Every operational control above is documented, version-controlled, and inspectable. The reasoning traces in §5.3 are the live evidence. A regulator, professional body, or insurer can reconstruct any agent-assisted output end-to-end.

6.4

Data exfiltration prevention.

Zato addresses the broader risk of data exfiltration through layered data loss prevention (DLP) controls. Every agent output is screened for sensitive data patterns (tax identification numbers, bank account details, client identifiers) before any external communication. Cross-client isolation blocks information from one engagement leaking into another through shared agent context or model state.

DLP · 01

Pattern screening

Outputs are checked for sensitive data patterns before any external communication.

DLP · 02

Cross-client isolation

Information from one engagement cannot leak into another through shared context or model state.

DLP · 03

Zone boundary firewall

A compromised sub-agent cannot file a tax return or issue professional advice. The architecture is the firewall.

> The zone boundary is a hard architectural firewall. Not a policy. Not a prompt.

07 SECTION · RESPONSIBLE AI

Direct alignment with NIST, OECD, the EU AI Act and ISO/IEC 42001: and what the Dual-Zone Architecture means for ANZ regulators and professional bodies.

Positioning Within the Responsible AI Discourse.

7.1

Alignment with international frameworks.

Zato's Dual-Zone Architecture and progressive autonomy framework are aligned with every principal international framework for trustworthy and responsible AI. The alignment is not aspirational. **Zato is certified to ISO 42001**, the AI Management System standard.

FIGURE 07

Zato framework alignment with international AI governance standards.

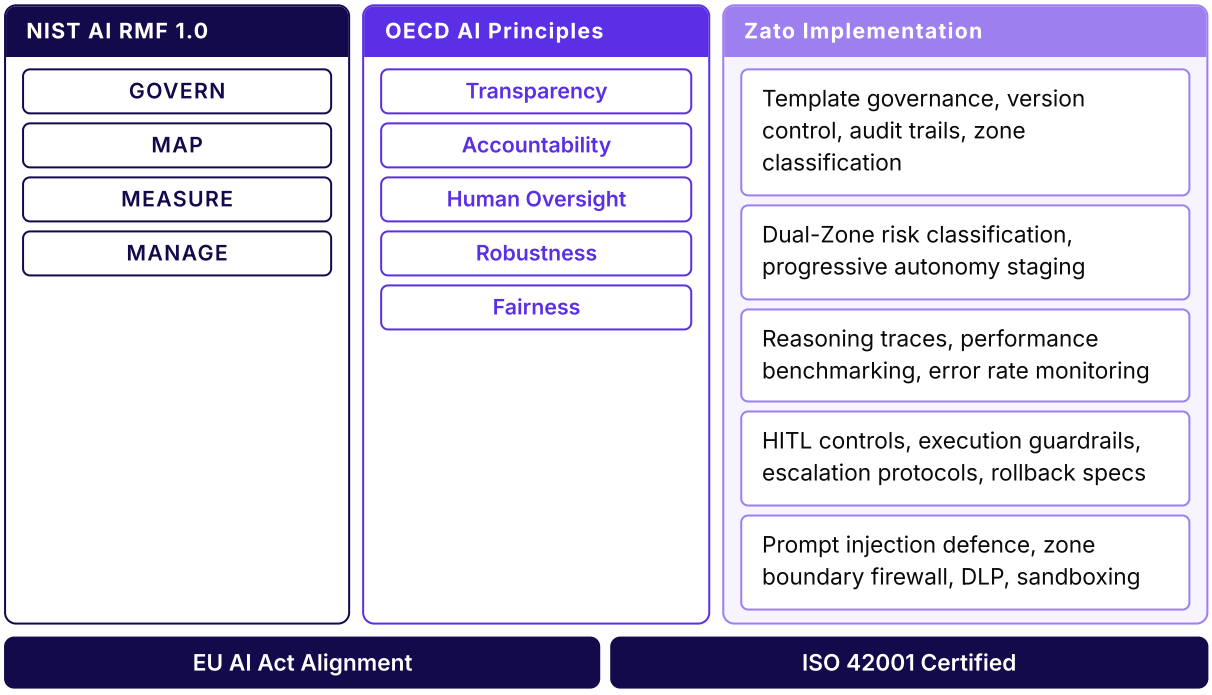


FIGURE 07 *Mapping from NIST functions and OECD principles to Zato implementation, plus EU AI Act and ISO/IEC 42001 alignment.*

- **NIST AI Risk Management Framework (AI RMF 1.0):** Zato's progressive autonomy model maps to NIST's GOVERN, MAP, MEASURE, and MANAGE functions. The Dual-Zone classification implements NIST's principle of proportionate risk management.
- **OECD AI Principles:** The framework's emphasis on transparency (reasoning traces), accountability (audit trails), human oversight (HITL controls), and robustness (execution guardrails) directly implements the OECD's five principles for trustworthy AI.
- **EU AI Act Risk Classification:** Zato's risk-tiered approach mirrors the EU AI Act's proportionate regulation model. ANZ firms that adopt Zato are positioned for compliance if comparable frameworks arrive in the region.
- **ISO/IEC 42001 (AI Management System):** Zato is certified to ISO 42001. Template governance, version control, and audit trail capabilities satisfy the standard's lifecycle requirements.

7.2

The Gartner Perspective on Agentic AI Governance.

Gartner's research on multiagent systems argues that governance frameworks must be in place before agentic deployments scale. The 2025 guidance highlights explicit agent boundary definition, human override capability, audit trail completeness, and progressive trust calibration. Every one of those is a core design principle in Zato's framework.

The Dual-Zone Architecture directly addresses Gartner's observation that agentic AI governance must distinguish between tasks that can be safely automated and tasks that require governed oversight: a distinction that many early agentic deployments fail to make explicitly.

SOURCE *Gartner*, "Multiagent Systems in Enterprise AI: Efficiency, Innovation and Vendor Advantage," December 2025.

7.3

Implications for ANZ regulators and professional bodies.

As AI agents become common in accounting, tax, and audit practice, ANZ regulators and professional bodies will assess AI-assisted professional output against documented governance, not vibes. Zato's architecture is built for that assessment. Clear delineation between autonomous preparatory work and governed professional judgment. Full reasoning traceability available for regulatory review. Template-based methodology documentation that lets a regulator inspect the governance framework itself, not just one output at a time.

REGULATOR-READY BY DESIGN

Zato's reasoning traces and version-controlled templates enable regulators to assess the governance framework itself: not just the individual output. That shifts AI assurance from sampling-based audit to architectural review.

08 SECTION · APPLICATIONS

Four practice areas where the Dual-Zone Architecture meets real engagement work: tax compliance, audit and assurance, AML/CFT, and strategic advisory.

Applications in ANZ Professional Services.

8.1

Tax compliance and advisory.

Tax engagements run the Dual-Zone Architecture end to end. Sub-agents gather data through the Xero connector, classify transactions through **GL Scrutiny**, prepare income and deduction schedules through **Workpaper Automation**, and produce first-pass computation of tax positions. Decision agents review the outputs, form conclusions on transfer pricing, R&D credits, international tax, and prepare client-facing advice. The autonomy stage is set by engagement maturity and risk profile. **Ziffy** sits over the top. Ask it anything about a position and it cites the workpaper, the regulation, and the reasoning trace.

8.2

Audit and assurance.

Audit benefits sharply from zone separation. Sub-agents handle the work that consumes the most junior time on every engagement: data analytics, workpaper preparation, reconciliation, anomaly detection through **GL Scrutiny**, evidence bundle assembly. Decision agents support the engagement team in forming conclusions on risk, materiality, and findings.

New engagements start in Shadow or Assisted mode. Opinion-forming sits at Critical-tier HITL regardless of autonomy stage. Dual sign-off, full reasoning trace, version-controlled template. The Review and Approve stages of the 8-stage workflow are where the partner's name goes on the work.

8.3

Regulatory compliance and AML/CFT.

AML/CFT compliance shows the handover at full stretch. Sub-agents gather customer due diligence data and detect suspicious patterns. Routing runs through **Command Centre**. Decision agents evaluate flagged items and prepare suspicious activity reports. Human sign-off is mandatory before any regulatory submission. The reasoning trace is the audit evidence.

8.4

Strategic advisory and the role of the practitioner.

Entity restructuring. Succession planning. Cross-border expansion. The highest-stakes engagements need the most professional judgment. Sub-agents provide research support, data analysis, and scenario modelling. Decision agents run in Shadow mode and offer structured analytical support. They do not substitute for the partner's judgment, and they are not built to.

This raises the question every partner asks: what happens to firm staff. The answer the architecture gives is direct. Sub-agents absorb the high-volume preparatory work that consumes the most junior time on every engagement. Decision agents support, not replace, the practitioner forming the conclusion. The capacity freed up at every level of the firm shifts upward. Juniors spend less time reconciling and more time learning to interpret. Seniors spend less time reviewing workpapers line-by-line and more time on advisory. Partners spend less time on routine sign-off and more time on the engagements that need a partner.

Zato's role in strategic advisory is to amplify the senior practitioner, not to replace one. In a profession where trust is what the client pays for, that is the only role worth building for.

09

Future directions.

The Dual-Zone Architecture in this paper is a foundation, not a finishing line. The field is moving, and the architecture extends naturally in several directions.

Dynamic zone reclassification. As agent capability and governance maturity grow, tasks now in the Governed Decision Zone become safe for autonomous execution. The boundary moves. The classification principle does not.

Inter-agent deliberation. Decision agents engage in structured debate before synthesis. Adversarial reasoning sharpens conclusions on complex positions. The reasoning trace records both sides of the deliberation, not just the answer.

Federated trust scoring. Standardised trust metrics across firms and jurisdictions let agent performance and governance maturity be compared like-for-like. Practitioners stop relying on vendor claims.

Regulatory sandbox engagement. Direct collaboration between vendors and ANZ regulators on sandbox environments where agentic systems are tested against real regulatory scenarios. The architecture is the conversation.

Continuous assurance. Periodic audit cycles give way to continuous assurance. Sub-agents monitor compliance in real time. Decision agents evaluate live issues as they arise. The audit shifts from sampling to streaming.

Each direction is consistent with the architectural commitments in this paper: separation of zones, progressive autonomy, traceable reasoning, programmatic guardrails. None requires a new framework. All extend the one Zato is building toward.

SOURCE *Gartner*, "Multiagent Systems in Enterprise AI," December 2025. *NIST*, "AI RMF Playbook," 2023.

10

Conclusion.

Agentic AI in accounting, tax, audit, and compliance advisory is not a question of if. It is a question of how. The firms that lead this transition are not the firms that deploy fastest. They are the firms that deploy responsibly, with architectures that draw a clear line between safe automation and governed judgment.

Zato's Dual-Zone Architecture draws that line. The Autonomous Execution Zone runs eleven categories of regulated-safe preparatory work at sub-agent speed. The Governed Decision Zone keeps professional judgment, opinion formation, and legally consequential action under progressive autonomy, execution guardrails, risk-tiered HITL controls, and reasoning traceability.

The four-stage progressive autonomy model earns decision agent independence through demonstrated performance. The Job Orchestrator sequences work across both zones with version-controlled templates. The Trust and Governance framework constrains every action, traces every decision, and produces regulator-ready evidence. ISO 27001, ISO 42001, ISO 9001, SOC 2.

For the firms that serve New Zealand and Australian businesses, Zato offers a path to agentic AI adoption that is technically capable and professionally defensible. In a profession built on trust, that is the only path that scales.

> Not the firms that deploy fastest. The firms that deploy most responsibly.

BACK MATTER

References.

01 Gartner, "Multiagent Systems in Enterprise AI: Efficiency, Innovation and Vendor Advantage," December 2025.

02 NIST, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," AI 100-1, January 2023.

03 OECD AI Policy Observatory, "Trustworthy AI Principles and Recommendations," 2024.

04 OECD, "Advanced AI Assistants: Governance Challenges and Policy Directions," 2024.

05 OWASP, "Top 10 for Large Language Model Applications," Version 1.1, 2024.

06 ISO/IEC 42001:2023, "Information technology — Artificial intelligence — Management system."

07 CPA Australia, "Well-funded IRD drives an aggressive compliance approach," 2025.

08 PwC New Zealand, "Tax Tips: Year in Review 2025."

09 NZ Inland Revenue, Taxation (Annual Rates for 2025–26, Compliance Simplification, and Remedial Measures) Bill Commentary, August 2025.

10 AASB, Climate Reporting and Assurance Standards, 2025.

11 ISACA, "No Looking Back: Transforming Audit with Artificial Intelligence," June 2025.

12 IBM Think, "AI Agents in 2025: Expectations vs. Reality," November 2025.

13 CPA.com/AICPA, "2025 AI in Accounting Report," June 2025.

14 Chartered Accountants Australia and New Zealand (CA ANZ), "Submission on the 2025 Occupation Shortage List Stakeholder Survey," 2025.

15 Bloomberg Tax, "Big Four Firms Roll Out AI That Can Handle Routine Tasks Solo," March 2025.

16 BDO, Audit Innovation Survey 2025.

17 Mordor Intelligence, "AI in Accounting Market Report," 2025.

18 Accountancy Age, "Big Four lag behind smaller firms in AI adoption, says ex-EY chair," August 2025.

19 EU Artificial Intelligence Act, Regulation (EU) 2024/1689, August 2024.

CITATION

Zato Origin Limited. *Agentic Orchestration for ANZ Professional Services: Trust, Guardrails, and Progressive Autonomy in Accounting, Tax, Audit, and Compliance Advisory.* Q2 2026, Version 2.1. Auckland, NZ.