EAIC.UK

ADVISORY PAPER | APRIL 2026

AI Governance for Established Regulated Firms

From Hidden Use to Defensible Control

Most established regulated firms do not need more AI ambition. They need visibility, ownership, evidence, and defensible control — and they need it now.

AUDIENCE
Boards, executive teams, risk leaders, compliance heads, operational owners

PLATFORM
Citadel AI Governance Platform — built and operated by EAIC

Executive Summary

Artificial intelligence is already present in most established regulated firms. It is present in vendor tools, in productivity assistants, in code generation pipelines, and in automation logic embedded within operational workflows. In the majority of cases it arrived before governance did. The register of approved AI systems — where one exists — typically understates reality by a significant margin.

This advisory paper makes a single, practical argument: the primary challenge for established regulated organisations is not AI ambition. It is control. Specifically, it is the ability to demonstrate, on demand and to a demanding audience, that the organisation knows where AI is used, what risk it carries, who is accountable, what decisions have been made, and what evidence exists.

That standard — defensible control — is not a future regulatory aspiration. It is the current expectation embedded in the EU AI Act, in FCA operational resilience and model risk guidance, in ISO 42001, and in the Board-level accountability frameworks that firms already operate under. The question is not whether it applies. The question is whether the organisation can meet it.

This paper sets out the five failure modes that leave established regulated firms exposed, defines what defensible control requires in practice, and describes an operating model that progresses from discovery through to board reporting. Throughout, we draw on two things in combination: Citadel, EAIC's AI governance platform — the system of record, policy, evidence, and executive oversight — and Sentinel, the consultant-led activation framework through which Citadel is populated, structured, and switched on. The distinction matters: Citadel is the product that runs governance continuously. Sentinel is the structured engagement that discovers the AI estate, validates and enriches it, quantifies risk and value, and leaves the client live inside the platform. Neither is complete without the other.

> The governance problem is not whether AI is used. It is whether the firm can prove that it knows where AI is used, what decisions it influences, what evidence exists, and who is accountable when something goes wrong.

Figure 1: AI Governance Maturity Spectrum — from operating blind to defensible control

1. The Hidden-Use Problem

Most established regulated firms do not begin their AI journey with an enterprise programme. They begin with local adoption. A team starts using a meeting transcription and summarisation tool. Marketing experiments with generative copy. Operations connects a model-backed feature in a vendor SaaS product. A developer uses code generation assistance in production support. Finance begins using an AI-assisted forecasting module embedded in their planning software.

None of this necessarily looks like an 'AI programme' when viewed in isolation. Viewed in aggregate, however, it represents an unmanaged and potentially material risk surface — one that intersects with every existing regulatory obligation the organisation already carries.

This hidden-use problem matters more in regulated settings for a straightforward reason: firms are already accountable for outcomes, records, controls, customer harm, privacy breaches, operational resilience failures, and third-party oversight. The use of AI does not create new accountabilities in isolation. What it does is concentrate and complicate existing ones. Where a human decision-maker previously produced an auditable judgement, an AI-assisted or AI-driven process may now produce an outcome with no comparable trail.

A credible governance position therefore begins with one blunt premise: the organisation is very probably already using more AI than its formal register acknowledges — and the gap between actual use and documented use is where regulatory and reputational exposure lives.

The regulatory context that makes this urgent

The EU AI Act entered into application in stages from August 2024, with prohibitions active immediately and obligations for high-risk AI systems progressively operative through 2025 and 2026. For firms operating across European markets — or handling data subject to European jurisdiction — the Act introduces mandatory conformity requirements, technical documentation standards, human oversight obligations, and incident notification duties.

In the UK, the FCA has been explicit in its expectations around model risk management, operational resilience, and third-party AI oversight. The Dear CEO letter traditions, PS21/3 on operational resilience, and the FCA's AI discussion papers collectively establish that AI use in regulated activities is expected to be documented, governed, and controllable. The ICO's guidance on AI and data protection adds a further obligation layer for any AI system that processes personal data — which, in practice, describes the majority of commercially deployed AI.

ISO 42001, the international standard for AI management systems, provides a voluntary but increasingly influential framework that auditors, regulators, and procurement teams reference when assessing whether an organisation's AI governance is credible. Firms that can demonstrate alignment with its structure are better positioned for both regulatory scrutiny and enterprise client due diligence.

These frameworks converge on a common requirement: the ability to produce, on demand, a coherent account of what AI systems the organisation uses, how they were assessed, what controls are in place, and how incidents are managed. That is precisely what most established regulated firms currently cannot do.

2. Why Established Regulated Firms Are Disproportionately Exposed

Large enterprises typically have the internal functions to absorb complexity: central architecture, dedicated model risk teams, specialist legal and regulatory counsel, procurement governance infrastructure, and formal risk committees with dedicated AI workstreams. Established regulated firms often face the same external obligations but carry far less organisational surplus. They need to evidence control using fewer people, less time, and against the same scrutiny threshold.

In practice, this structural constraint produces five recurring failure modes that leave established regulated organisations measurably exposed.

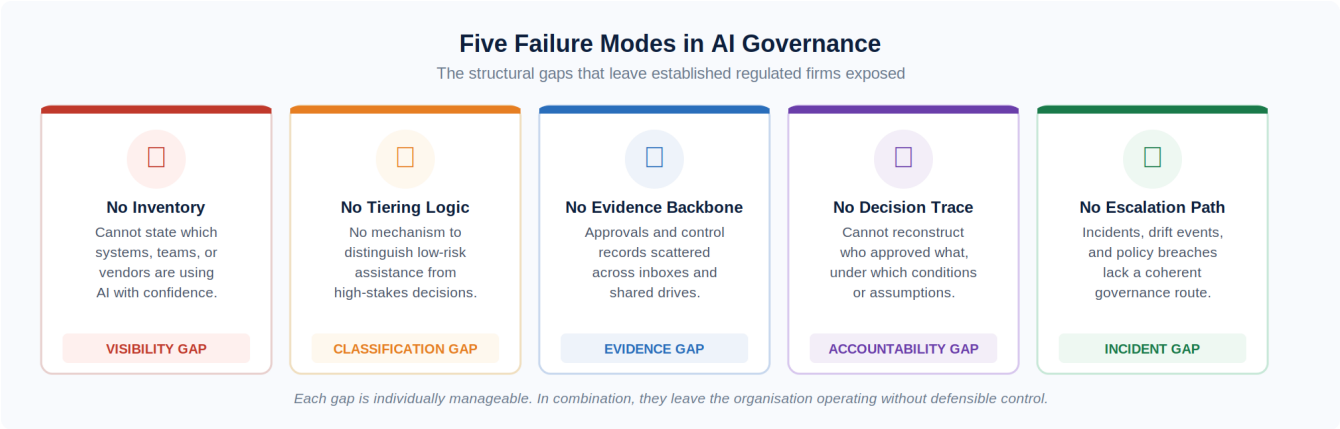

Figure 2: The five structural failure modes in AI governance

- No reliable inventory. The firm cannot state with confidence which systems, teams, or third-party vendors are using AI — particularly where AI capability is embedded in existing software rather than procured as a distinct tool.
- No tiering logic. There is no consistent mechanism to distinguish low-friction productivity assistance from higher-stakes decision support, autonomous process execution, or agentic automation acting on behalf of the firm or its customers.
- No evidential backbone. Approvals, evaluations, exception records, and control attestations exist in scattered form — across inboxes, shared drives, and presentation decks — or are not captured at all.
- No decision trace. Leadership cannot readily reconstruct who approved what, under which assumptions, with what conditions or caveats attached, or when that approval expires or requires renewal.
- No escalation path. Incidents, near-misses, model drift events, policy breaches, and vendor risk events lack a coherent governance route — they surface ad hoc, are resolved informally, and leave no durable record.

The cumulative effect of these gaps is significant. The organisation may believe it is moving cautiously and responsibly while in fact operating with a risk surface that is both larger and less well-understood than leadership appreciates. When a regulator or auditor asks for evidence of AI controls, the response that emerges — assembled under pressure from disconnected sources — rarely conveys the confidence that the underlying intent of the organisation deserves.

This is not a failure of intent. It is a failure of infrastructure. Established regulated firms need governance infrastructure that is proportionate to their scale, executable without large specialist teams, and designed to produce the kind of evidence trail that survives scrutiny.

3. What Defensible Control Actually Means

Defensible control is not a slogan or an aspiration. It is a specific operational capability: the ability to demonstrate, on demand, that the organisation can identify AI use across its estate, classify it consistently, assign clear ownership, assess risk proportionately, apply documented policy, retain structured evidence, log decisions immutably, track incidents formally, and report status to leadership in a form that withstands scrutiny.

The distinction between a weak control position and a defensible one is not primarily about intent or investment. It is about auditability. The following table illustrates the difference across six control dimensions that regulators, auditors, and boards routinely probe.

Control question	Weak position	Defensible position
Do we know where AI is used?	Ad hoc declarations and assumptions	System register with named owners, discovery inputs, and lifecycle status
Can we explain the risk?	Narrative judgement on a case-by-case basis	Consistent Sentinel weighted scoring across four or six dimensions
Can we show evidence?	Documents scattered across teams and inboxes	Structured evidence items with status, expiry, and provenance chains
Can we prove decisions?	Meeting notes and email chains	Immutable decision ledger with audit hash, conditions, and timestamps
Can leadership see status?	Manual updates and anecdote	Board-pack snapshots, ROGS composite score, and exportable artefacts
Can we survive regulatory scrutiny?	Reliance on good faith and goodwill	Documented control trail with incident register and notification tracking

The pattern across all six dimensions is consistent. A weak control position relies on human memory, goodwill, and narrative. A defensible position produces a structured, timestamped, provenance-bearing trail that exists independently of the individuals who contributed to it. That independence is what makes control credible — not just to regulators and auditors, but to boards that need to make accountability decisions under pressure.

> The difference between saying 'we have guidelines' and showing a regulator an operating trail is the difference between a governance statement and a governance fact.

4. The EAIC Operating Model: Sentinel Activates, Citadel Governs

The control model developed by EAIC is built around a precise division of responsibility. Citadel is the system of record, policy, evidence, monitoring, and executive oversight — the platform that runs the governance programme continuously once it is live. Sentinel is the consultant-led activation framework that makes Citadel operational: it discovers the AI estate, validates and enriches the data, quantifies risk and value, generates board-level decisions, and hands the client over to Citadel with ownership, evidence, workflows, and reporting already running.

This distinction is important for a practical reason: Citadel is only as good as the data inside it. An ungoverned register with incomplete ownership, unscored systems, and absent evidence is not governance — it is a list. Sentinel exists to ensure that the governed estate inside Citadel is real, validated, and audit-ready from the point of go-live, not assembled gradually through self-service over months.

The six stages of the operating model map directly to Sentinel's six activation phases, each of which produces structured data that lands inside Citadel's object model — not narrative work product that lives in a report and then goes stale.

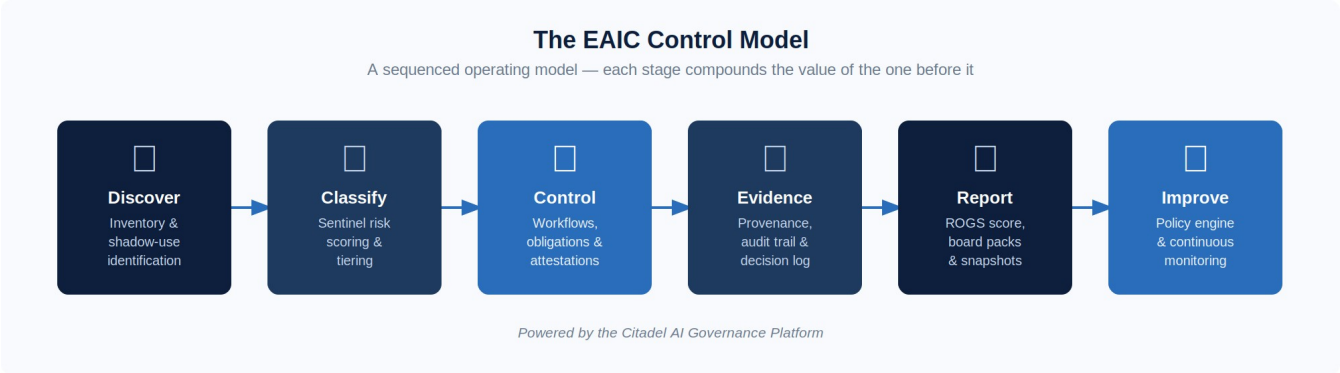

Figure 3: The EAIC Control Model — six sequenced stages from discovery to continuous improvement

Stage	Leadership outcome	Citadel capability
Discover	Visibility of real AI use — including shadow and vendor systems	System inventory, auto-discovery inputs, external vendor reconnaissance
Classify	Proportionate treatment determined by risk, not intuition	Sentinel weighted risk score; dynamic risk events; four-dimension standard model, six-dimension agentic model
Control	Named accountability with documented approval paths	Workflow engine, controls library, obligation mapping, attestations
Evidence	Auditability at the system, decision, and control level	Evidence items with status, expiry, and provenance; audit event log; immutable decision records
Report	Board and leadership confidence with objective indicators	ROGS composite governance score, board-pack snapshots, exportable HTML artefacts
Improve	Governance at scale without governance overhead	Policy engine, deployment monitoring, runtime anomaly detection, continuous assurance

The 'Improve' stage is worth particular emphasis. The value of a governance programme is not static. As the organisation's AI estate evolves — as new systems are procured, as vendors update their models, as

autonomous agents are deployed, as regulatory requirements shift — the governance model needs to respond. Citadel's policy engine and continuous monitoring capabilities are designed to make this adaptation systematic rather than reactive. For organisations that want a structured re-assessment, the Sentinel Annual re-score provides a recurring consultant-led review that refreshes the risk baseline, updates the ROGS score, and generates a new board decision pack directly from live Citadel data.

5. Platform Credibility: The Citadel Technical Foundation

The authority of any governance model rests partly on whether the platform supporting it has been engineered to the same standard it asks of the systems it governs. Citadel has been built from the ground up around governance primitives rather than dashboard aesthetics. What follows is a summary of the technical capabilities that make the control model in this paper executable.

It is worth noting at the outset that these capabilities do not operate in isolation. Sentinel's activation framework is specifically designed to populate Citadel's object model — AI systems, vendors, agents, use cases, owners, risks, evidence, obligations, and value records — so that the platform is a live, validated, and audit-ready governance environment from day one rather than an empty system awaiting self-service input. Every technical feature described below is made operational through the Sentinel engagement before being sustained through the Citadel subscription.

AI System Inventory and Lifecycle Management

Citadel maintains a formal AI system registry covering the complete lifecycle from initial intake through active deployment to retirement. Each registered system carries a structured metadata model — owner, purpose, deployment context, risk tier, lifecycle state — and is linked to its associated assets, evidence items, decisions, and incidents. The registry is not a spreadsheet or a shared document: it is a versioned, queryable, tenanted data model with a full audit history.

The platform supports external vendor reconnaissance — the ability to capture and assess AI capability embedded in third-party products — which is where the majority of unregistered AI use tends to reside in established regulated organisations.

The Sentinel Weighted Risk Engine

Risk classification in Citadel uses the Sentinel Weighted Risk Score — a multi-dimensional scoring model that produces a consistent, auditable risk tier for every registered AI system. For standard AI systems, Sentinel evaluates four primary dimensions. For agentic AI systems — those capable of autonomous action, tool use, or multi-step decision chains — Sentinel extends to six dimensions, reflecting the materially different risk profile of autonomous execution.

Scores are not static labels. The platform supports dynamic risk events: material changes to a system's context, deployment scope, or operational behaviour trigger automated risk recalculation and governance review workflows. This means the risk record stays current rather than representing a historical snapshot from the point of initial assessment.

Governance Workflows

The platform provides a structured workflow engine covering the governance events that established regulated firms encounter most frequently: system intake and initial approval; change review for material modifications; periodic review and attestation renewal; incident declaration and resolution; and capital approval for significant AI investment decisions. Each workflow type enforces the approval chain, captures decisions as immutable records, and generates the evidence artefacts that auditors and regulators expect to see.

Policy Engine and Obligation Mapping

Citadel's policy engine allows governance teams to define structured policy conditions — rules that reference system attributes, risk scores, control status, and evidence state — and evaluate compliance across the estate automatically. Policy violations are surfaced as actionable findings with associated remediation workflows, rather than as passive notifications. Alongside the policy engine, the platform maintains an obligation register that maps regulatory and standards requirements to specific AI systems, tracking compliance status and linking attestations to the systems they cover.

Evidence Management and the Decision Ledger

Every evidence item in Citadel carries a provenance record: who created it, when it was captured, what it relates to, what its current status is, and when it expires. Evidence items are linked to the controls, obligations, and systems they substantiate. This structure means that when an auditor asks for the evidence supporting a particular control assertion, the response is a structured, timestamped record — not a search through shared drives.

Decisions recorded in Citadel — approvals, exceptions, escalations, rejections — are stored with an audit hash, making them immutable for governance purposes. Role-based approval requirements are enforced by the platform, and the full approval chain is captured with timestamps. This produces a decision trail that is both comprehensive and independently verifiable.

Board Reporting and the ROGS Score

The platform generates executive summaries and board-pack outputs including the ROGS composite governance score — a single indicator of the organisation's overall AI governance maturity and control health, decomposed into component metrics for deeper analysis. Board packs can be exported as structured HTML artefacts, providing a snapshot of governance status at a point in time that can be retained for regulatory or audit purposes.

Agent Governance

Agentic AI systems — those that autonomously execute multi-step tasks, interact with external tools and APIs, or act on behalf of users or the organisation without continuous human supervision — present a governance challenge that standard AI frameworks were not designed to address. Citadel includes specialised controls for agentic systems, including the six-dimension Sentinel model, autonomy scope restrictions, action logging, human override mechanisms, and escalation triggers. This capability is increasingly material as agentic deployments move from experimental to operational.

Security and Multi-Tenancy

Citadel is built as a multi-tenant SaaS platform with row-level tenant isolation enforced at the PostgreSQL database layer. Authentication uses RS256 JWT with Redis-backed token revocation and rate limiting. Role-based access control is enforced consistently across API endpoints. Structured audit logging captures every significant governance event with actor, timestamp, and context. These are not incidental technical details: they are the infrastructure that makes governance data trustworthy and the platform suitable for regulated-industry deployment.

6. Sentinel Activation: Fixed Fee, Fixed Scope, Closed When Live

Governance programmes fail most commonly when they end with a report rather than a live operating environment. The EAIC model is built around a different close condition: the engagement is complete when Citadel is live, the AI estate is populated, and the governance operating cadence has started — not when the board pack is presented and the consultants leave.

The offer architecture reflects this. The Sentinel Diagnostic is a one-day, fixed-price entry point that surfaces the real AI estate and defines the path forward. Sentinel Activation is the core consultant-led engagement — fixed fee, fixed scope, delivered at pace, and closed when the client is live in Citadel. Fee and scope are agreed in advance based on estate complexity, with transparent tier logic: Standard Activation for estates up to thirty governed systems, Complex Activation for larger or more technically demanding estates involving agentic AI or multi-entity structures. The Citadel Subscription sustains the programme from go-live. EAIC Advisory and Implementation services close the highest-priority gaps.

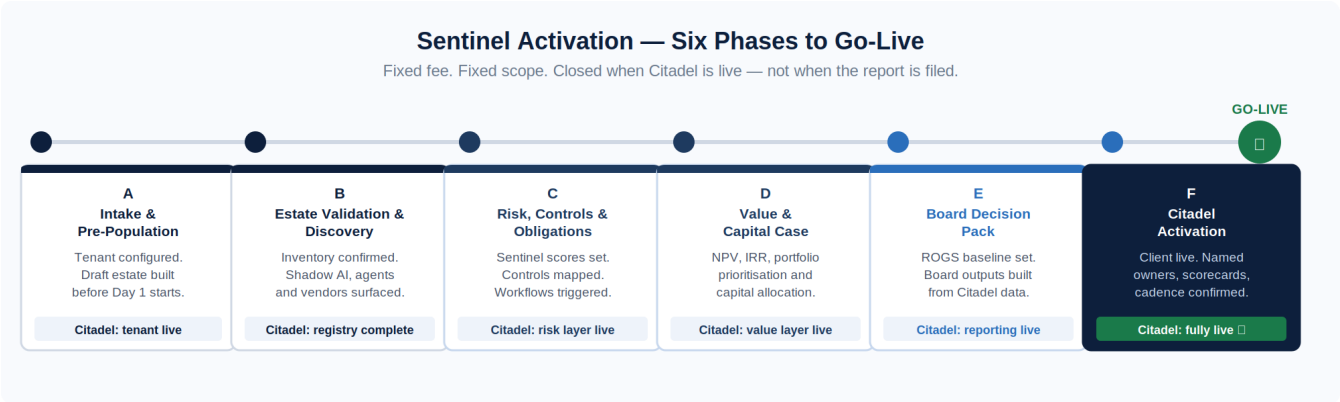

Figure 4: Sentinel Activation — six phases from intake to Citadel go-live

Phase	Sentinel Stage	Citadel outputs
A	Intake & Pre-Population	Citadel tenant configured; draft AI estate created from pre-engagement data; discovery queue active; consultant arrives informed with preliminary inventory and risk hypotheses
B	Estate Validation & Governance Discovery	Confirmed system and vendor inventory; ownership assigned; shadow AI and agentic systems surfaced; business context captured; unknown gaps reduced and escalated
C	Risk, Controls & Obligations Baseline	Sentinel risk scores set; governance, evidence, and compliance posture established; initial workflows triggered; obligation mapping for material systems completed
D	Value & Capital Case	Financial records and value cases created; risk-adjusted return, NPV, and portfolio prioritisation completed; capital approval workflows initiated for major investments
E	Board Decision Pack	ROGS baseline score established; board-facing outputs generated directly from Citadel data; posture summary, exposure map, and prioritised roadmap delivered
F	Citadel Activation	Client live in Citadel with named owners, active scorecards, evidence, workflows, and review cadence; Sentinel Activation closes; Citadel Subscription begins

The six phases above are not sequential consulting stages with handover points between them. They are a continuous activation sequence — each phase creates or validates structured data that lands directly in Citadel's object model, so the platform is populated in real time rather than assembled at the end. Fee and scope are transparent: the Diagnostic starts at £3,500 and is credited in full on upgrade. Standard and Complex Activation tiers are priced on estate size, agreed before engagement begins, and fixed.

> The close condition for a Sentinel engagement is not a board presentation. It is a live governed estate inside Citadel — with named owners, scored risk, seeded evidence, active workflows, and a ROGS baseline already running. Governance starts on day one, not after the report is filed.

7. The EAIC Position

The market does not need another AI strategy framework, another maturity model, or another abstract set of principles. It needs something more specific: a practical, executable control model that works for organisations with real constraints, real regulatory obligations, and real AI already in use.

EAIC's position is built on three convictions. First, that the right starting point for established regulated firms is not more AI ambition — it is visibility and control over what already exists. Second, that defensible governance is not a large-enterprise luxury — it is a proportionate operational requirement that can be delivered without the organisational overhead that large firms bring. Third, that a governance platform earns authority not through the breadth of its feature set but through the depth of its governance mechanics — and that Citadel, activated through Sentinel, has been built to exactly those mechanics.

The practical implication of those convictions is a coherent and transparent offer stack. The Sentinel Diagnostic proves there is something worth governing. Sentinel Activation — fixed fee, fixed scope, tier-priced by estate complexity — constructs the live governed estate inside Citadel. The Citadel Subscription runs the programme continuously. EAIC Advisory and Implementation services close the highest-value and highest-risk gaps. Each stage has a discrete job, a defined outcome, and a price agreed before work begins — not a consulting engagement with an open end.

The organisations that navigate the next phase of AI adoption most effectively will not be those that moved fastest or spent most. They will be those that built control early enough to move with confidence — and who can demonstrate that control when it matters.

References and Source Basis

Citadel platform references in this paper are derived from the April 2026 EAIC technical reference, covering the system inventory model, Sentinel risk engine, workflow architecture, policy engine, evidence management model, board reporting functions, agent governance framework, and security architecture. Platform capabilities described reflect the current production specification.

Regulatory and standards framing draws on the following primary sources:

- European Commission: Regulation (EU) 2024/1689 — the EU Artificial Intelligence Act — entered into application August 2024.
- Financial Conduct Authority: PS21/3 — Building Operational Resilience; FCA AI Discussion Paper DP5/22; and related Dear CEO correspondence on model risk management.
- Information Commissioner's Office: Guidance on AI and Data Protection; and the ICO's Explaining Decisions Made with AI framework.
- International Organization for Standardization: ISO/IEC 42001:2023 — Information Technology, Artificial Intelligence, Management System.
- Bank of England / PRA: SS1/23 — Model Risk Management Principles for Banks (with cross-sector relevance to AI model governance).