Corentis Shield

AI checkpoint for regulated workflows

R&D and innovation reviewers

Corentis ControlBench Innovation Brief

A benchmark and evaluation concept for runtime AI control.

AI needs a checkpoint before it acts. Corentis provides it.

An innovation-led brief explaining how Corentis can benchmark the difference between uncheckpointed and checkpointed AI-assisted workflows.

Generated April 2026

For discussion and pilot exploration only

Corentis Technologies Ltd | Company No. 17182737 | hello@corentis.co.uk

R&D AND INNOVATION REVIEWERS

Overview

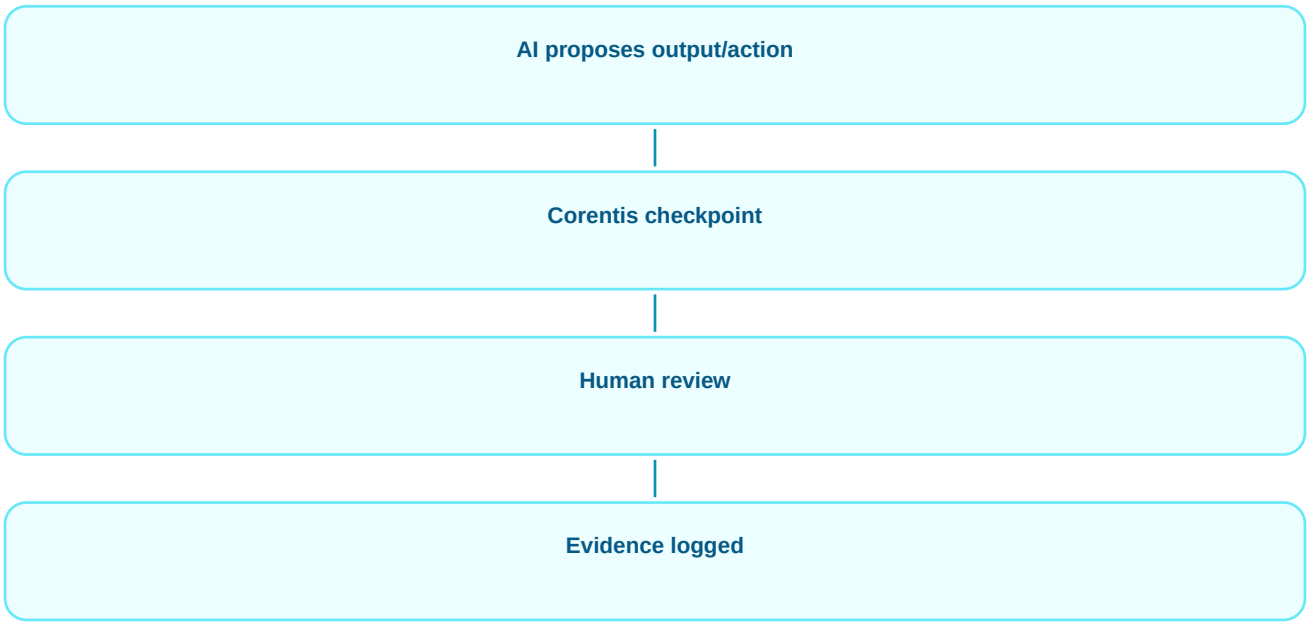
Core position

Corentis Shield is an AI checkpoint for regulated workflows.

AI needs a checkpoint before it acts. Corentis provides it. Corentis Shield is designed to help teams check AI outputs before they reach customers, teams or live systems.

VISUAL SUMMARY

Checkpoint flow



EVALUATION SHAPE

Baseline vs checkpoint

Baseline

AI proposes output or action without a runtime checkpoint. Review points and evidence gaps are assessed afterwards.

Checkpointed

AI proposes output or action. Corentis checks controls, pauses risky items, routes human review and records evidence before action.

Why benchmark runtime control?

AI-agent adoption needs more than impressive demonstrations. Regulated teams need to see whether checkpointing changes the quality, reviewability and evidence completeness of workflow decisions.

The innovation challenge

The challenge is to translate policy expectations into structured controls, run realistic scenarios, compare uncheckpointed and checkpointed workflows, and produce evidence leaders can inspect.

What ControlBench is

ControlBench is the proposed Corentis evaluation environment for testing runtime checkpointing. It uses scenario libraries, control schemas, baseline comparisons and evidence scoring to make the value of checkpoints measurable.

Baseline vs checkpointed workflow

The baseline run shows what happens when AI proposes outputs without a runtime checkpoint. The checkpointed run tests the same scenarios with Corentis controls, review routing and evidence capture in place.

What we would measure

The measurement path focuses on reviewability, control and evidence rather than inflated impact claims.

- Unsafe direct-action attempts caught.
- Vulnerable-customer escalation accuracy.
- Evidence completeness score.
- Human-review routing accuracy.
- False positive and false negative balance.
- Reviewer confidence.
- Scenario coverage.
- Policy-to-control mapping completeness.
- Blocked-action explainability.
- Audit artefact completeness.

Workstream 1: scenarios

Build a focused scenario library for complaints, hardship, vulnerable-customer signals, unsupported closure and sensitive customer communication.

Workstream 2: controls

Map policy intent into structured controls that can be checked at the action boundary.

Workstream 3: baseline comparison

Run uncheckpointed examples to identify evidence gaps, unsafe direct-action attempts and weak review routes.

Workstream 4: evidence scoring

Run checkpointed examples and score evidence completeness, escalation quality and reviewer clarity.

Workstream 5: feasibility outputs

Produce a feasibility report, scenario library, control schema, evidence scoring method and clear recommendation for the next validation stage.

Why this could become valuable infrastructure

A strong ControlBench result would create reusable assets: scenarios, controls, evidence standards and evaluation methods for regulated AI-agent workflows.

Next conversation

If your organisation is exploring AI agents in regulated workflows, Corentis is ready for a focused conversation about validation, pilot design and strategic support.

SELECTED SIGNALS

Evidence context

MCKINSEY GLOBAL AI SURVEY

88% of respondents in McKinsey's 2025 global survey reported regular AI use in at least one business function.

McKinsey & Company, 5 November 2025

MCKINSEY GLOBAL AI SURVEY

23% of respondents said their organisations are scaling an agentic AI system somewhere in the enterprise.

McKinsey & Company, 5 November 2025

MCKINSEY GLOBAL AI SURVEY

51% of respondents from organisations using AI said their organisations had seen at least one negative consequence.

McKinsey & Company, 5 November 2025

IBM / PONEMON

63% of breached organisations lacked AI governance policies to manage AI or prevent shadow AI.

IBM / Ponemon Institute, 2025

SALESFORCE AI CUSTOMER RESEARCH

72% of customers say it is important to know if they are communicating with an AI agent.

Salesforce, 2026 page accessed / report current at access

Selected sources

McKinsey & Company: The State of AI: Global Survey 2025
Date/status: 5 November 2025. Source domain: mckinsey.com.
Global cross-industry AI adoption context.

McKinsey & Company: The State of AI: Global Survey 2025
Date/status: 5 November 2025. Source domain: mckinsey.com.
Global agentic AI momentum context.

McKinsey & Company: The State of AI: Global Survey 2025
Date/status: 5 November 2025. Source domain: mckinsey.com.
Global AI risk and consequence context.

IBM / Ponemon Institute: Cost of a Data Breach Report 2025
Date/status: 2025. Source domain: ibm.com.
Security and AI governance-gap context.

Salesforce: State of the AI Connected Customer
Date/status: 2026 page accessed / report current at access. Source domain: salesforce.com.
Vendor AI trust and customer expectations context.

Company details and next step

Corentis Shield is provided by Corentis Technologies Ltd. Company No. 17182737. Company type: Private limited company. Registered office: Suite A, 82 James Carter Road, Mildenhall, IP28 7DE, United Kingdom. Contact: hello@corentis.co.uk.

Start a Conversation