Secure AI Agents

SAIA | Action Contracts, independent verification and traceable decisions

THE NEED - AI use in EU enterprises rose from 13.5% in 2024 to 20.0% in 2025; Sweden reached 35.0%. Gartner forecasts agentic AI in 33% of enterprise applications by 2028 yet predicts that more than 40% of agentic projects will be cancelled by the end of 2027 because of cost, unclear value or inadequate risk controls. IBM found that 13% of 600 breached organisations reported AI-related breaches; 97% of those lacked proper AI access controls and 63% had no AI governance policy or were still developing one.[1, 2, 3]
The industrial problem is that AI agents can read data, call tools, modify files and trigger workflows. Existing security systems can control identity and access to some degree, but organizations still need a vendor-neutral way to define a bounded task, verify the outcome independently and determine which earlier tests or human decisions may need to be revisited after the agent changes.

CONTROL	VERIFICATION	CHANGE IMPACT
Existing identity and runtime policy controls restrict sensitive actions before execution.	Independent verifiers assess observable outcomes using protected verification rules and evidence. The agent must not be able to approve its own actions or alter the verification basis.	The project will investigate which previous tests and human decisions require reassessment after changes to models, instructions, tools or permissions, and evaluate both missed and unnecessary reassessments against controlled reference cases.

THE SOLUTION - The scalable Action Contract primitive

SAIA industrial R&D consortia will develop and validate a small, model-independent Action Contract. Each contract defines the intended result, permitted capabilities, explicit constraints, independent verifiers, retained evidence and the allowed decision: accept, retry, deny or escalate, with rollback where the affected system supports a defined recovery procedure. Models and tools may change; the contract, verifier and accountable decision remain stable.

1 CONTRACT	2 ACT	3 VERIFY	4 DECIDE	5 REASSESS
Intent, capabilities, constraints	Agent proposes or performs action	Independent checks test result	Accept, retry, deny or escalate; rollback where a defined recovery procedure is supported	Review affected contracts after change

Project scope

DEVELOP	VALIDATE
• Action Contract schema and assurance record • Local, model-independent verification • Reference adapter identity, runtime policy and traces • Change-impact rules that produce traceable reassessment	• Two bounded industrial agent use cases • One agent and two versions per use case • Three shared cybersecurity scenarios • Comparison with current review & incident-analysis work

What the project will deliver (TBD with full consortia)

Open method	Action Contract schema, assurance record, change-impact rules and trust boundaries.
Reference implementation	Local verification prototype builds on existing testing and evidence-management components, with limited adapters to external tools and their reports. The project can extend these capabilities where needed for agent actions and change-impact analysis.
Industrial demonstrators	Two pilots showing how actions are constrained, independently verified and reassessed after change.
Security evaluation	Repeatable attack, change and incident-reconstruction tests with documented scope and error rates.

[1]IBM, Cost of a Data Breach Report 2025: original source
[2]Gartner (2025), agentic AI forecast and cancellation risk: original source
[3]Eurostat (2025), EU enterprise AI use: original source

How value will be measured

Action coverage	Verifier quality	Change impact	Operational value
Sensitive test actions governed by a versioned contract	Reproducibility and blocked or escalated test actions	Missed and unnecessary reassessment candidates	Time to review change and reconstruct incident

SCOPE BOUNDARY

SAIA will not build a new agent platform, identity system or policy engine. It will develop and validate the Action Contract and assurance layer across existing components.

WHY NOW

The EU regulatory landscape is tightening. AI Act transparency obligations apply from 2 August 2026; Cyber Resilience Act (CRA) reporting obligations for manufacturers of products with digital elements have applied since 11 September 2026; and the revised Product Liability Directive explicitly covers software and AI for products placed on the market from 9 December 2026. For robotics and AI-enabled machinery, Regulation (EU) 2023/1230 (the Machinery Regulation) applies from 20 January 2027. Applicability depends on the organisation's role, product, and use case.[4]

Join the consortium

We are inviting select Swedish industrial partners that are developing or preparing to deploy or are deploying local AI agents with access to business data, engineering tools, software environments or operational systems. Each partner contributes one bounded workflow, a small set of sensitive actions, security requirements, a representative test environment and staff time for design, attack testing and validation.

1. EXPRESS INTEREST	2. SCOPE THE PILOT	3. CONFIRM PARTICIPATION
Call or send email of your interest and to explore a use case. Include estimated in-kind budget part of planned work.	In a call. Define workflow, environment, baseline and expected contribution with RISE.	E-mail agreed role, work package, budget details and CV of main participant before submission.

Project frame and consortia partner contribution

Funding and duration	10 million SEK in-kind* financing from industry consortia and 10 million from VINNOVA to develop a pilot for intended application for up to SEK 20 million SEK total eligible costs, 24 months.[5]
Partner contribution	*In-kind means all costs related to the project such as staff time, integration, software, testing, hardware, data preparation and evaluation and direct financing. Costs must be worked, paid, recorded and reported to VINNOVA mid- and end-of-project or if requested. Partner in-kind contributions cover all RISE costs, unlocking our expertise and essential technical components and tools.
Call fit and deadlines	VINNOVA “Reliable and secure industrial AI”; one Swedish company / at least two legal project parties, Interests to join SAIA sent preliminarily before 13th November 2026. Call text. None-binding consortia proposal submission 19 January 2027 at 14:00.[6] **Last day to verify join project 25th of March 2027**

Interested in joining?
+46 70 602 1745 / aslak.felin@ri.se to explore agent use case and the security problem the project should solve.

[4]Official EU texts:AI Act;Cyber Resilience Act;Product Liability Directive
[5]Vinnova call facts:original call page
[6]Vinnova call requirements and deadline:original call page