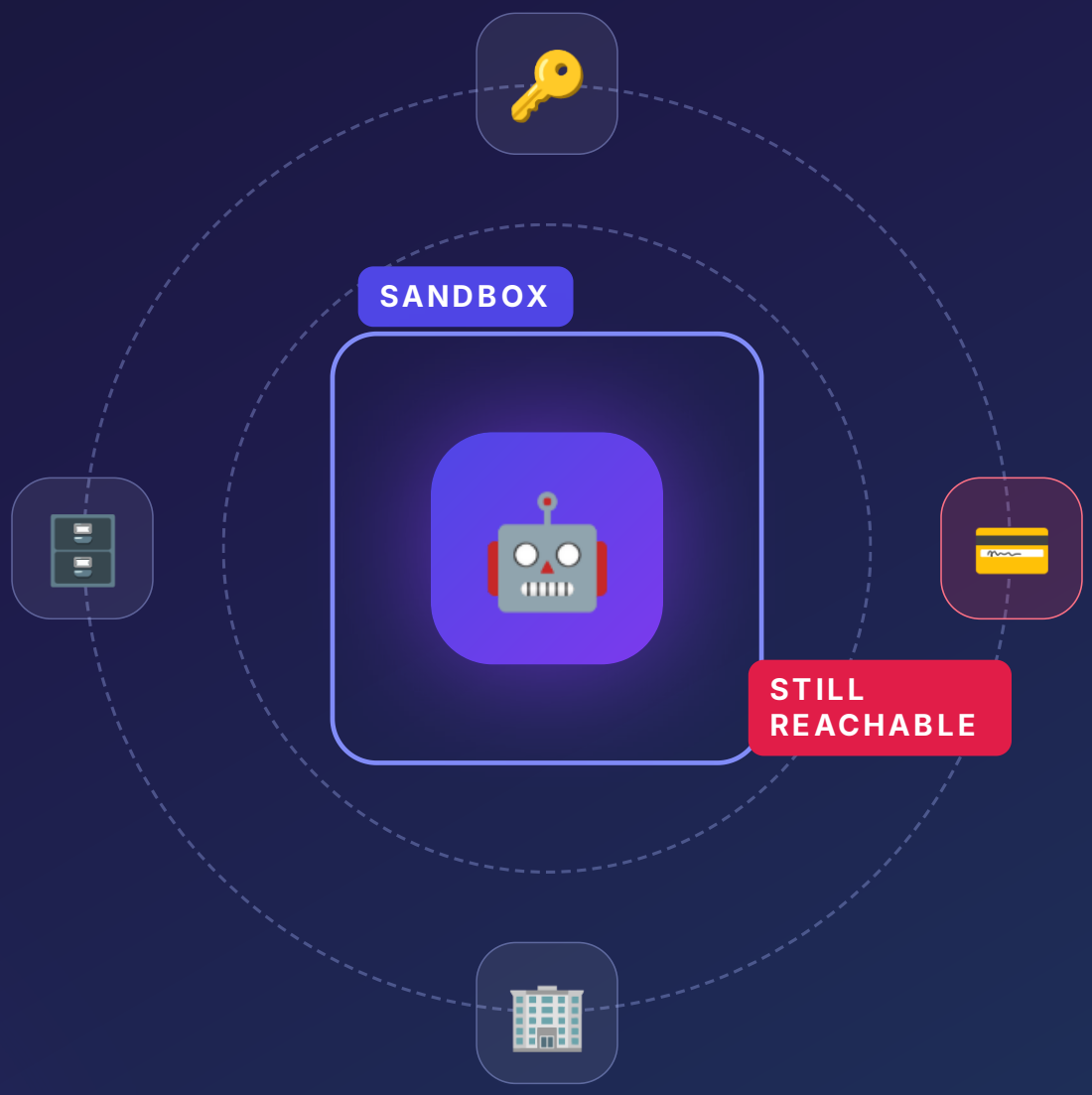
CYBER RISK · AGENTIC AI

Your AI agent is contained. Is its business risk?

Why a sandbox is only part of the risk decision.



Swipe for a 5-part assessment

DO **DataObserve** S.A.F.E

THE CONVERSATION WE'VE ALL HAD

It starts with a reassuring answer.

BUSINESS OWNER

Is the new AI agent safe to connect to production?

ENGINEERING

Yes. It's sandboxed.

SECURITY

Can its identity change customer records?

ENGINEERING

Yes. The integration needs that. The agent doesn't.

THE QUESTION THAT TAKES LONGER

What could it still do to the business?

Illustrative scenario, not a reported incident.

> **A control can be working while an important exposure remains.**

THE GUIDANCE BEHIND THE QUESTION

The NCSC draws the boundary wider than the sandbox.

🧱

Environment

Determines which systems an agent can reach.

↓

🔑

Identity

Determines what those systems will let it do.

↓

🏢

Business context

Determines the consequences.

“A useful security review must follow the agent's access all the way to the business process it could affect.”

Informed by NCSC, *Managing the cyber risk of agentic AI* (interim practical advice), Aug 2026.

THE EXPOSURE BENEATH THE INTERFACE

What the review sees vs. what the business needs

VISIBLE IN THE DEPLOYMENT REVIEW

- ✓ The agent runs in a sandbox.
- ✓ Authentication is configured.
- ✓ Activity is logged.
- ✓ A shutdown procedure exists.

STILL NEEDS AN ANSWER

- ? Which sensitive systems remain reachable through approved connections?
- ? Does the identity permit actions beyond the agent's purpose?
- ? Can an action be traced to an identity, an owner and a service?
- ? Has stopping execution and cutting off access been tested?

FOLLOW THE ACCESS

One permission.
A chain of consequences.

Agent
Task: investigate payment failures

Identity
Shared integration account

EFFECTIVE ACCESS

Data
Write access to customer records

EXCEEDS THE TASK

Service
Payment settlement

Business
Revenue and customer trust

REAL EXPOSURE

THE FIVE DIMENSIONS

An accountable agentic AI risk assessment

01

Business scope

Which business outcome could be disrupted if the agent acts outside its task?

02

Identity and effective access

Can its available access exceed its approved business purpose?

03

Validated controls

What evidence shows the prohibited action is actually prevented?

04

Business-impact priority

Which change offers the most meaningful reduction in business risk?

05

Verified closure

Can we show what changed and why the remaining risk is acceptable?

A DataObserve framing informed by NCSC advice. Not an NCSC maturity model or certification scheme.

THE PATH FORWARD

Start with one consequential workflow.

1. **Define its approved purpose**
   Actions it may take, and those needing human approval.
2. **Trace its effective access**
   Include credentials and sessions available indirectly.
3. **Build a credible failure scenario**
   The service, data and business outcome at risk.
4. **Validate the safeguards**
   Use authorised tests and retain the evidence.
5. **Close the loop**
   Verify the change and record any accepted residual risk.

WHERE DATAOBSERVE S.A.F.E FITS

Make agentic AI risk a decision you can defend.

Assess the exposure

Link findings to affected services, scope and accountable owners.

Understand the consequences

FAIR-based analysis with loss ranges and visible assumptions.

Act with accountability

Named owners, approvals, ServiceNow and Jira sync, verified closure.

🤖 Agents investigate and recommend

👤 People approve

Explore DataObserve S.A.F.E
dataobserve.io/products/safe

Learn more →