DROPSOLID
ARTIFICIAL INTELLIGENCE

WHITEPAPER 03

THE SOVEREIGN AI SERIES

WHITEPAPER 1 · PROCURE
AI sovereignty, procurement

WHITEPAPER 2 · DEPLOY
Two deployments in production

WHITEPAPER 3 · OPERATE
Transparency and governance

SOVEREIGN AI · TRANSPARENCY AND GOVERNANCE

Glass box vs black box

The enterprise AI transparency playbook. What observability, traceability and steerability require at the architectural level.

Published by
Dropsolid AI · 2026

EXECUTIVE SUMMARY

You cannot govern what you cannot inspect

Most enterprise AI today runs as a black box. Prompts go in. Responses come out. The people accountable for those responses, legal, compliance, information security and the CIO, cannot inspect what happened between input and output. They cannot evaluate the answer. They cannot audit the decision. They cannot influence the model's behaviour without rebuilding it.

The alternative is a glass box. Transparent by design. Observable by default. Traceable per interaction. Steerable without redeploying.

This whitepaper explains what glass box AI actually requires at the architectural level, why most AI stacks are structurally opaque, and how enterprise organisations can procure and operate glass-box systems today. Two production examples anchor the argument: **EPSO and Pidpa, both running sovereign AI as glass boxes** with audit traces satisfying EU institutional governance requirements.

IN THIS REPORT

01 Why governance models built for deterministic software break

02 The six architectural properties of a glass box

03 Evidence from production, and how to move an existing estate

BLACK BOX

- Input and output visible, nothing between
- Weekly reports, after the fact
- Vendor dashboard is the only record
- Behaviour changes require a rebuild

GLASS BOX

- Every step in the chain observable
- Live signal, drift caught as it starts
- Exportable trace per interaction
- Behaviour steered through configuration

> Transparency has to be built into how the system runs, not generated afterwards as a report.

PART ONE

The governance gap

Your governance model was designed for deterministic systems, testable in isolation, with reproducible outputs. AI applications break all three assumptions. Nobody sets out to build a black box, but three procurement routes produce them reliably.

AI CHANGES EVERYTHING

Why AI breaks the assumptions your controls were built on.

HOW YOU END UP WITH A BLACK BOX

The three routes that lead to black boxes, and how to recognise yours.

PART ONE · AI CHANGES EVERYTHING

Correct is a distribution, not a value

Legacy software is a system. You provide input, it converts to output. Deterministic. When there is a bug, you know because for some input you get unexpected output.

LEGACY SOFTWARE	AI-ENABLED APPLICATION
One input, one output	Probabilistic reasoning
Deterministic behaviour	Autonomous agents and external tools
Testable in isolation	Context managed across many turns
Reproducible results	Multiple models in one workflow

Modern AI-enabled applications work differently. The output for a given input can vary. **"Correct" is a distribution, not a value.**

Your governance model was designed for the legacy world. It expects deterministic systems, testable in isolation, with reproducible outputs. AI applications break all three assumptions.

> You either update your governance model, or you accept that you don't have governance over AI.

PART ONE · HOW YOU END UP WITH BLACK BOXES

Nobody sets out to build a black box

They emerge from three procurement routes that seemed reasonable at the time. Most organisations will recognise at least one of them in their own estate.

ROUTE ONE

Commercial API adoption

Your team starts calling commercial model APIs directly. Fast to prototype.

WHY IT IS OPAQUE

Impossible to introspect. You do not know why the model gave one answer over another, you cannot log the internal reasoning, and you cannot audit under the EU AI Act.

ROUTE TWO

Opaque SaaS AI

You buy a product that has AI built in. The vendor treats the AI as a feature.

WHY IT IS OPAQUE

No visibility into how it works, no control over which model runs, and no audit trail beyond the vendor's own dashboard.

ROUTE THREE

RAG without governance

Retrieval-augmented generation on a general-purpose model. Better than direct API calls, because the answers are grounded.

WHY IT IS OPAQUE

Still a black box, because prompt handling, retrieval logic and response generation remain opaque to the compliance team.

> All three routes produce systems that satisfy the immediate business need and fail the audit requirement.

PART TWO

What a glass box requires

Six architectural properties define a glass box AI system

6 PROPERTIES

The six architectural properties a glass box requires.

THE LIFECYCLE

Where each property shows up in the life of a single AI request.

3 PILLARS

Trust is built on three pillars: transparency, observability, traceability.

PART TWO · SIX PROPERTIES OF A GLASS BOX

1-3 : properties that decide what you can see

Six architectural properties define a glass box AI system. The first three govern visibility: what the system records, what it surfaces while running, and what it exposes to a human reviewer.

1 Traceability of every interaction

Every input, every retrieval, every model call and every output is logged with sufficient detail to reconstruct what happened. Not just that the user asked X and the system said Y.

The full chain: the prompt, the retrieved passages, the model version, the response and the citations.

2 Observability of AI operations

Live dashboards showing model usage, cost tracking, latency and response quality. Not weekly reports; live signal.

When something starts drifting, you see it before an incident.

3 Transparency of the decision process

The reasoning chain is exposed to a human reviewer. If the model retrieved this passage rather than that one, why. If confidence was low, on what dimensions.

A reviewer who cannot see the retrieval cannot evaluate the answer.

THE COMMON THREAD

All three are recording decisions, not describing them. Anything not captured at the moment of the interaction cannot be recovered afterwards.

PART TWO · SIX PROPERTIES OF A GLASS BOX

4 - 6 : Three properties that decide what you can do

Visibility alone is passive. The remaining three properties are what turn an observable system into a governable one.

4 Feedback that corrects the source

User feedback and expert review are captured against each interaction. Answers that drifted are corrected manually at the source content, as part of the improvement cycle.

The system improves through operation, one correction cycle at a time.

5 Auditability against a compliance framework

The audit trail can be filtered, exported and produced for an inspector on demand.

Not an ad-hoc export exercise; a standard operational capability.

6 Steerability without rebuild

You can influence the model's behaviour through configuration, prompts, context templates and retrieval thresholds, without a code change and a redeploy.

This is what makes the system operationally manageable.

Guide 03 turns these six properties into an assessment you can run against a system you already operate.

> The EPSO and Pidpa deployments (covered in Whitepaper 2) run with all six properties in place, which is why they could satisfy institutional governance without bespoke work.

PART TWO · THE LIFECYCLE OF AN AI REQUEST

In a black box, most of the chain is invisible

Every AI request follows the same chain. What separates a glass box from a black box is how much of it is observable to the people accountable for the answer.

THE CHAIN	BLACK BOX	GLASS BOX
Request a visitor or an editor performs a request	●	●
Application sends input and context to the AI provider	○	●
AI provider queries the language model	○	●
Model requests additional context from the vector database	○	●
Model returns an answer to the AI provider	○	●
AI provider produces an observable trace	○	●
Response returned to the requester	●	●

● Observable ○ Opaque

WHO THIS IS FOR

An internal user, in compliance, legal or IT operations, inspects, monitors, understands and influences AI operations on behalf of the organisation. In a glass box every step is observable, every trace auditable, every configuration steerable.

PART TWO · TRUST BUILT ON THREE PILLARS

Missing any one leaves you exposed

Enterprise-grade AI requires three capabilities together. They are not three products; they are three properties of one architecture.

PILLAR ONE

Transparency

Full visibility into the data flowing through the system

Certified security around the platform

An ethical, protected foundation for all AI operations

PILLAR TWO

Observability

Real-time visibility into AI performance

Every interaction accurate, efficient, fully optimised

Drift caught at the moment it starts

PILLAR THREE

Traceability

A complete audit trail of every AI decision

Total accountability and regulatory compliance

Verifiable data lineage

ONE ARCHITECTURE

Each pillar depends on the other two. A trace without live signal goes stale, live signal without a trace cannot be audited, and neither means anything if the data path is not visible in the first place.

PART THREE

Proof, and what to do next

One trace from a live institutional deployment, then the sequence for moving an estate you already run. Most organisations are not starting from scratch.

GLASS BOX CASE IN PRODUCTION

EPSO as a glass box, shown through the record a single interaction leaves.

FIVE STEPS TOWARDS A GLASS BOX

5 steps to move an existing system towards a glass box

PART THREE · GLASS BOX CASE IN PRODUCTION

Case: EPSO as a glass box

EPSO's deployment satisfies every one of the six properties. The clearest way to show that is not to re-tell the deployment, which WP2 documents end to end, but to look at the record one candidate question produces.

THE TRACE, PER INTERACTION

Question	As asked, in the candidate's own language
Retrieved passages	The sections of EPSO's procedural documents the answer was drawn from
Model used	Selected by the governed gateway, recorded per call
Answer and citations	The response returned, with the sources a candidate can click to verify
Export format	The shape the EU's information-governance framework expects

Inspection and export

Internal EPSO staff can inspect any interaction, filter by cycle, and export subsets for audit purposes.

Correction that carries forward

Feedback loops let selection support staff flag answers that drifted; corrections are applied manually to the source content in the improvement cycle.

Configuration, not deployment

Model configuration is externalised, so changing how the system handles a class of question is a configuration change.

Read alongside Whitepaper 2

The deployment itself, and the Pidpa case, are documented in *Sovereign AI in production*. This page is the governance view of the same system.

> That is what glass-box AI in production looks like. Not a report about the system; but the system's own record.

PART THREE · FIVE STEPS TOWARDS A GLASS BOX

Most organisations are not starting from scratch

You have AI in production already. The question is how to move from where you are to where you need to be, without stopping the work that is already running.

01 **Inventory**

List every AI-enabled system your organisation runs. Include prototypes, third-party features and embedded functionality. Under the EU AI Act, everything counts.

02 **Categorise**

Which are high-risk under the Act? Which touch personal data? Which produce output visible to citizens, customers or regulated audiences?

03 **Audit trace inventory**

For each system: what trace exists per interaction, where is it stored, can it be exported? If the answer to any of these is unclear, you have a black box.

04 **Gap analysis**

For each system with an incomplete trace, decide: rebuild, migrate to a glass-box platform, or retire.

05 **Procure glass-box for new**

The old ones you fix over time. The new ones you build right from day one.

THE RULE TO ADOPT

Every new AI use case ships as a glass-box or it doesn't ship at all.

DROPSOLID
ARTIFICIAL INTELLIGENCE

ABOUT DROPSOLID

Dropsolid is the Drupal Diamond Partner behind the sovereign AI use cases in production at EPSO and Pidpa.

Since 2013
building on Drupal

ISO 27001
certified: Dropsolid NV with Dropsolid AI NV

Ghent, BE
headquartered in Belgium

THE SOVEREIGN AI SERIES

Whitepaper 1 · PROCURE
AI sovereignty: a CIO's procurement guide for the EU AI Act

Whitepaper 2 · DEPLOY
Sovereign AI in production: EPSO and Pidpa, documented end to end

YOU ARE HERE
Whitepaper 3 · OPERATE
Glass box vs black box: the enterprise AI transparency playbook

Test a system you already run

Guide 03 turns the six glass-box properties into an assessment you can run against a system you already operate.

Download the glass-box health check →

Talk to Dropsolid AI

The same six properties both deployments in this series run in production.