

The login was valid. The data still walked out.

Lesson: valid logins still need MFA, export limits and task-level access. Argos tests what a login can reach and traces what leaves.

165+

organizations potentially exposed: Snowflake customers, 2024

- ✗ No MFA on the accounts
- ✗ Credentials unrotated for years
- ✗ No network allow lists
- ✗ Infostealers, some on contractor systems

Not a breach of Snowflake's platform. Source: Mandiant

Building now · the center of Argos

1 · Leak tracing and contamination detection

- Unique canary per recipient: explicit items plus subtle per-copy variations.
- A leaked sample is matched against a registry to name the source. Deterministic.
- Model probe: give a canary prefix, test for the unique suffix. Statistical, against a baseline.
- Limits: a negative is not proof of clean; rewriting can strip markers.

Verified · 2 sample apps · 4/4 demo fixes

2 · Access control proven by attack

- Static pass collects routes, DB access points and owner/auth guards. No spec, no crawling.
- Dynamic pass sends real requests as users A and B, and with no token, to a local app.
- Checks: ownership (BOLA), missing authentication, SQL injection. Only reproduced findings are reported.
- Fix: patch a copy, the same attack must now fail, then a fix PR opens.
- Runs in a disposable container: no network, read-only source, non-root.

Planned

3 · Setup the AI does for you

- Pillar 1 starts from a file: argos canary issue, argos canary scan. Any stack.
- AI drafts argos.config.yaml for a person to review.
- AI can raise a verdict, never lower it. Verdicts come from reproduction, markers and statistics.
- One evidence report: what was tested, results, limits.

Issue canaries



Distribute to vendors



Attack-test the serving system



Leak surfaces



Trace to a recipient



Probe the model



Evidence report

Companies that share data need **proof, not promises.**

Argos turns vendor trust into evidence, before an incident, not after.

165+

orgs exposed via valid
logins

Snowflake customers, 2024

7+

class actions after a vendor
breach

Mercor, 2026 (WSJ)

1

untraced leak can void an
eval

Scores nobody can trust

Why Argos wins

Reproduced, not guessed

We report only what we reproduce, then prove the fix.

Traceable by design

A unique canary per recipient names the source of a leak.

Starts from a file

No spec, no crawling, no security team (planned).

We found no commercial product offering per-recipient tracing plus
contamination detection.



VERIFIED

Attack-tested access, 4/4
demo fixes



BUILDING NOW

Canary tracing + model
probe



NEXT

Vendor-level access matrix



LATER

Customer-run workers

Looking for design partners with
private eval sets