

“This Is a Test” Is Not an Authorization

An agent may accept a persuasive description of its environment. A live execution boundary must obtain environment identity, approval, limits and revocation from trusted state outside the conversation.

HINTS, TIPS & TRICKS

- Use a harmless duplicate with production execution disabled.
- Bind credentials to environment and action so a story cannot retarget them.
- Preserve denied attempts with the failed condition and owner.

OPERATOR CHECKLIST

- Record trusted environment identity.
- Store scope, limit, expiry and approval externally.
- Pass an ordinary in-bounds request.
- Repeat with a false sandbox claim and stale field.
- Turn oversight into a shift: revoke and prove denial.

COPY-READY PROMPT

Design a bounded “just a test” exercise for [workflow]. Define trusted environment identity, permitted action, account, limit, expiry and approval source outside the prompt. Provide one valid request and two persuasive out-of-bounds requests. Specify denial codes, retained evidence, shift owner and safe cleanup.

SOURCE & PROVENANCE

Article claims: The article reports separate cyber incidents in which descriptions of a test or simulation did not match the live environment, while preserving each report's limits.

Operational guidance: The duplicate-environment exercise is defensive guidance, not a live-intrusion procedure.

Canonical package: 2026-08-28 · payload, final article, source digest, signal scan and section audit.