Wopee.io

wopee.io/false-pass-checklist

The False Pass Checklist

Seven questions to ask any AI testing tool before you trust its green.

A false pass is a test that reports PASSED without proving it. These seven questions tell you whether your AI tester can produce one, and whether you would notice.

PASSED proved by the run | INCOMPLETE not proven yet | FAILED proved broken

Wopee.io's answers describe verdict integrity, which is in experimental preview.

1. **When the agent says PASSED, what did it record as evidence for each check?**
   Why it matters: A pass with no record is an opinion.
   > **Wopee.io's answer:** Every run keeps the agent's steps, checks and screenshots.

2. **Does anything check the pass against that record, or only the agent's own account of it?** MUST ANSWER
   Why it matters: An agent can describe a check it never ran.
   > **Wopee.io's answer:** Integrity gates compare the reported pass with what the run actually recorded.

3. **What happens to a pass the record doesn't back up?** MUST ANSWER
   Why it matters: If it stays green, the dashboard is lying.
   > **Wopee.io's answer:** In Enforce, an unproven pass becomes Incomplete.

4. **Is "unproven" a state of its own, separate from passed and failed?**
   Why it matters: Unproven is not broken: it needs a better check, not a bug ticket.
   > **Wopee.io's answer:** Yes: Incomplete (amber), with the reason in the report.

5. **Can you try the stricter verdict on your own runs before it changes any result?** MUST ANSWER
   Why it matters: Teams adopt what they can trial without risk.
   > **Wopee.io's answer:** Report only, on by default: it flags disagreements and never changes a stored result until you switch on Enforce.

6. **Can a product owner trace a green result to its evidence in one click?**
   Why it matters: Trust has to reach people who don't read logs.
   > **Wopee.io's answer:** The run report shows the Verdict Integrity section with each gate's finding.

7. **Does the vendor say where the tool is still weak?**
   Why it matters: Honest limits are a sign the rest is true.
   > **Wopee.io's answer:** Yes: one switch per project today, per-check switches in design.

If a tool can't answer 2, 3 and 5, its green is a claim, not a result.

See a real run answer 2, 3 and 5: “When AI tests pass but prove nothing”
youtube.com/watch?v=DF_2-VGVQKk

Start free at wopee.io • no credit card

wopee.io/no-proof-no-pass