

# Integrating AI & Testing



# Atomic Labs

Presented by Charlotte Haycock

Email: [charlotte.haycock@atomiclabs.co.nz](mailto:charlotte.haycock@atomiclabs.co.nz)

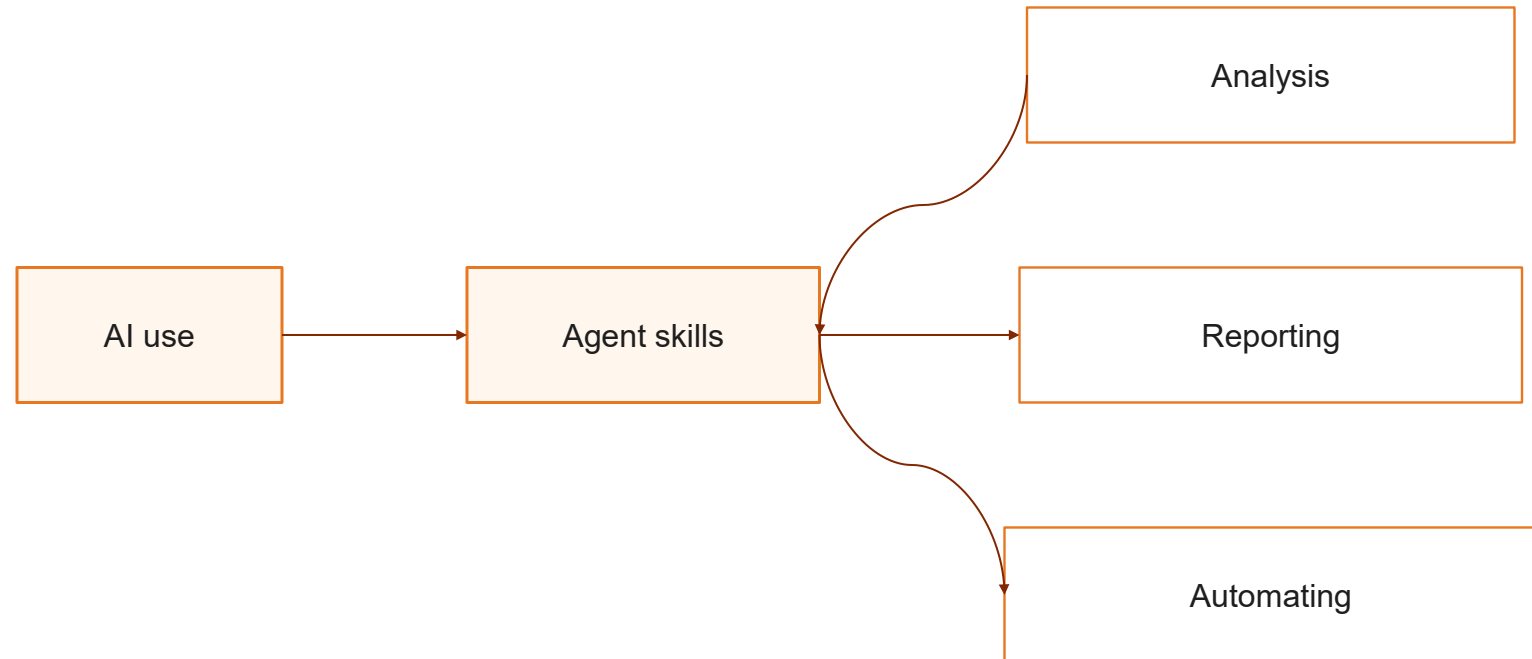
9 June 2026

# What we mean by AI and agent skills

AI becomes most useful when it is wrapped in a skill with a clear job and guardrails.

A skill is a packaged workflow that makes agent behaviour more consistent, safer, and easier to reuse.

**A good skill is clear, constrained, testable, and easy to maintain.**





# How we maintain the skills

- When users steer or correct the output, that is useful feedback.
- It usually means the skill is suboptimal and needs a better prompt, example, or guardrail.
- The maintainer skill reviews the interaction & suggests updates.
- Updates accepted, pushed into central repo, and available for the next use.

# How we test the skill itself

## Expected behaviour

- The skill stays within its guardrails
- It stops or escalates at the human-in-the-loop point
- It follows the intended skill flow
- It uses the expected supporting tools
- It does not take over work that should come back to a person

Verify the right skill path and tools are invoked

Review the workflow executed & output against expectations

Confirm the human-in-the-loop step happens at the right point

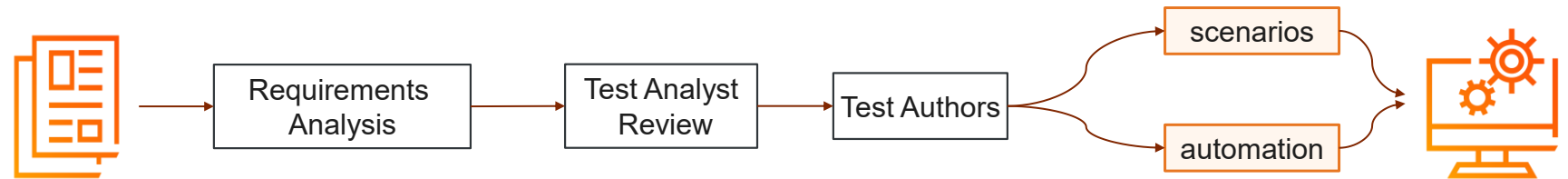
Compare sample runs against known-good outcomes

Capture evaluation results for reuse, including MLflow

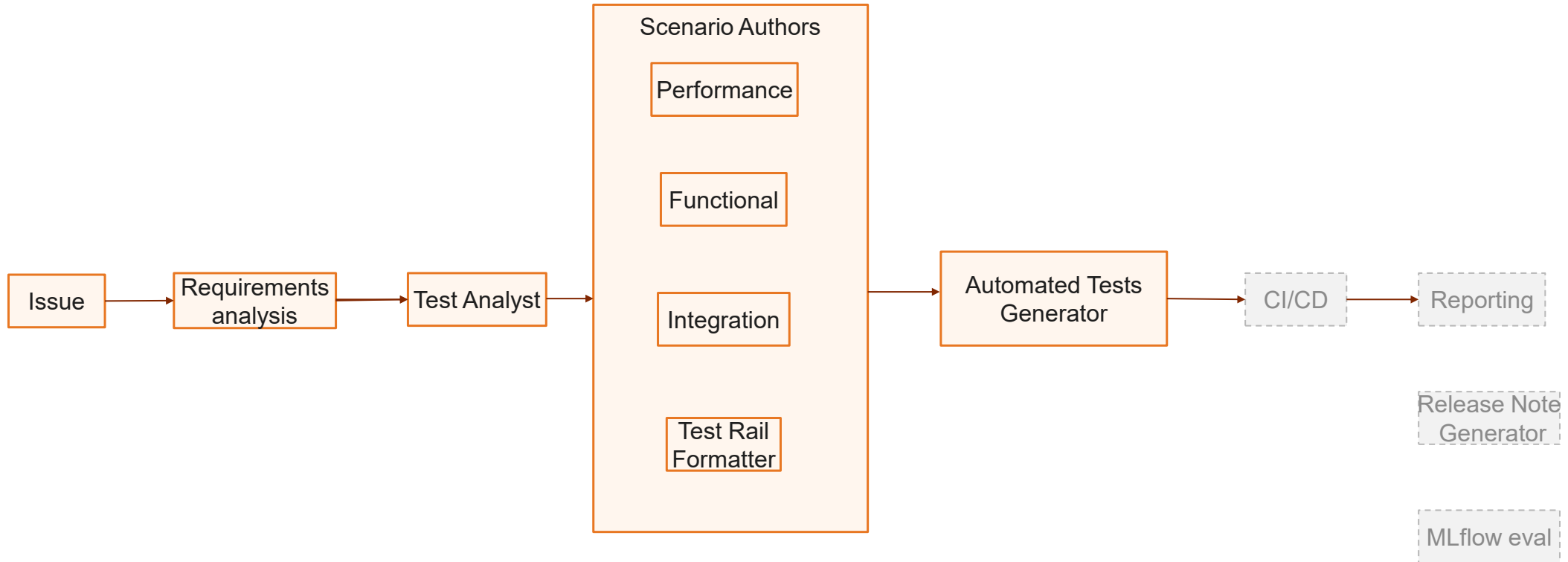
# Agents in practice: through the QA-analysis flow

We have a defect report & desired outcome but no acceptance criteria.

The QA-analysis flow turns that into a testable path.



# Current QA skills bundle and next expansion



# Thank you