

Microsoft
**Most Valuable
Professional**

When CI Fails

Governed investigation
with GitHub Copilot and MCP

Nikila Fernando

Microsoft MVP for Azure and Developer Technology

Senior Platform Engineer - Sysco Labs

The developer's investigation work



Investigation brings this evidence into one reviewable explanation.

What this session demonstrates

01

CI failure

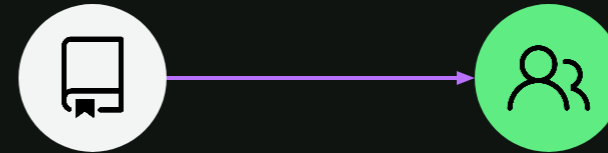
Explain independent build and policy failures.



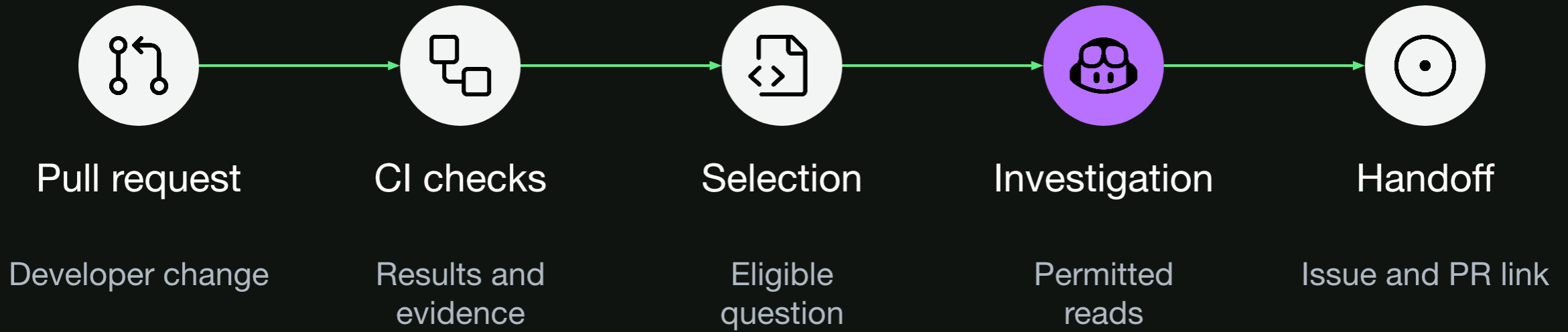
02

Shared change

Check the applications that use a shared component.



CI and investigation



Required checks keep control of merging.



A build failure and a policy violation

Check	Result	What CI observed
Application tests	PASS	The tested application behavior works
Build validation	FAIL	A caller references a removed output
Organization policy	FAIL	The change enables public ingress

Two independent failures need two corrections.



The investigation result



A report the developer can act on

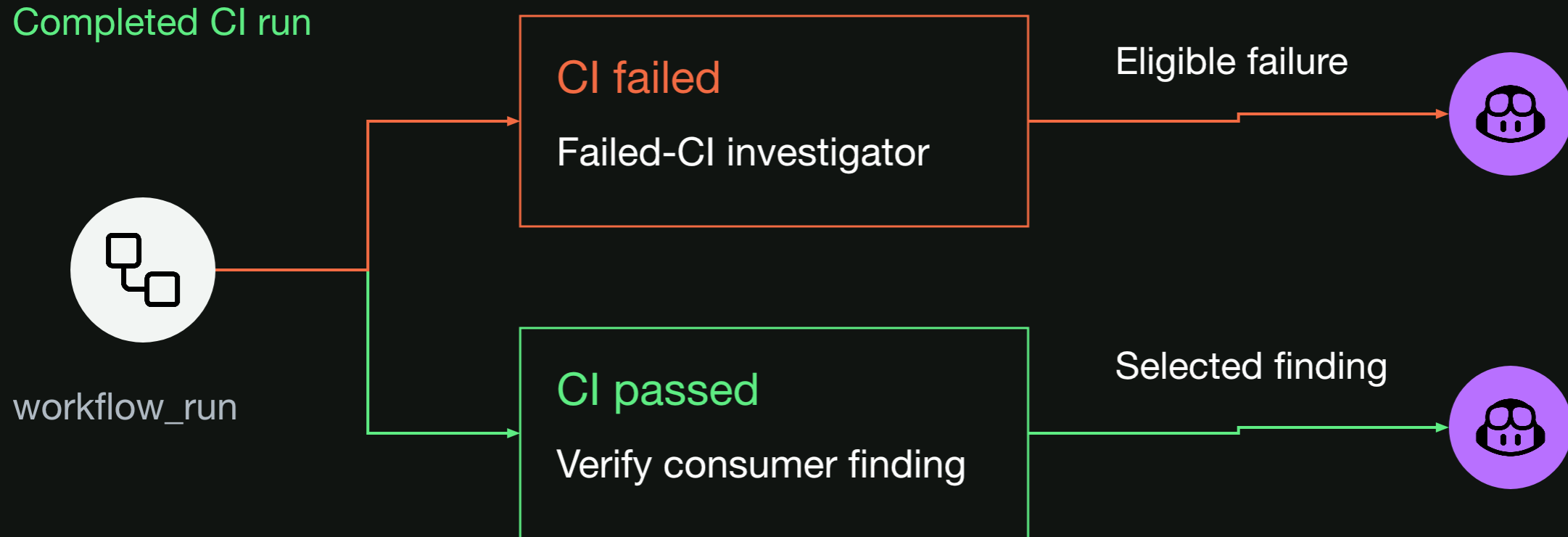
Finding	Suggested correction	Verify
Build configuration	Restore the output or update callers	Build validation
Policy violation	Restore the approved access setting	Policy check

Each finding cites its source and states missing evidence.

Demo 1: PR #20, Issue #21 and the PR feedback link



The event that starts investigation



PR events start CI. Eligibility and the kill switch control follow-up.



The workflow definition

```
on:  
  workflow_run:  
    workflows: ["CI"]  
    types: [completed]  
  
tools:  
  github:  
    mode: gh-proxy  
  
max-ai-credits: 120
```

Markdown source

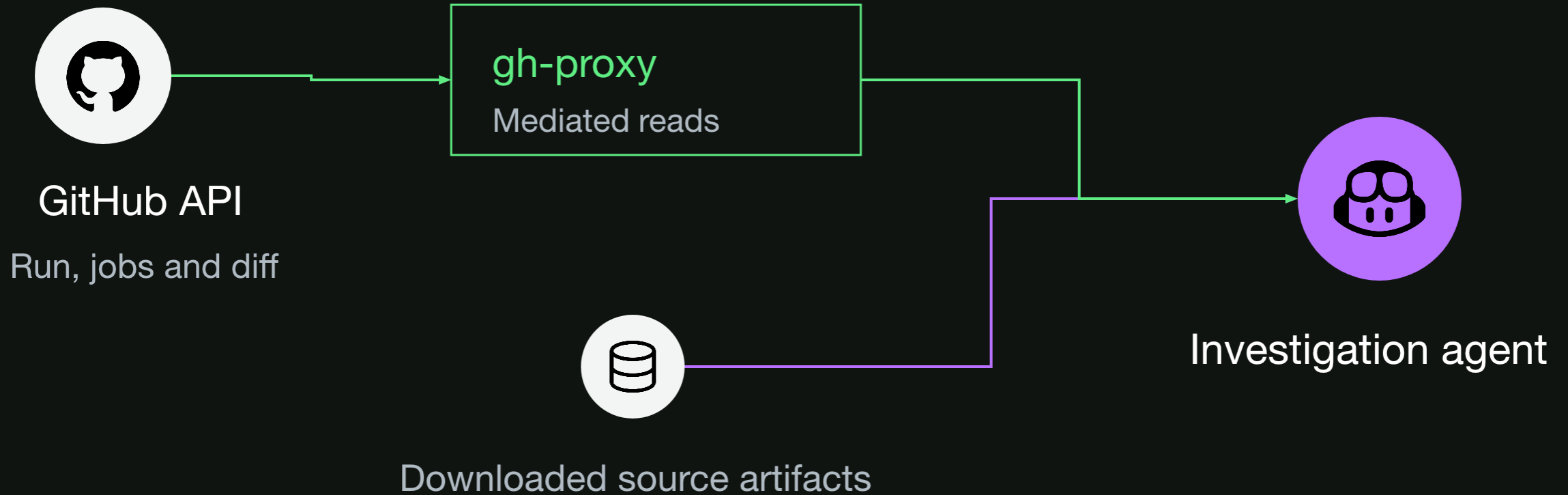
Event and conditions
Task instructions
Permitted tools
Controlled outputs

gh aw compile

The generated .lock.yml is the executable Actions workflow.



Evidence from GitHub

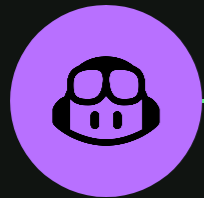


Bind evidence to the source run, attempt and tested revisions.



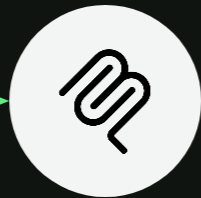
Organization context through MCP

Demo CI worker



Agent

<http://localhost:8765/mcp>



Policy MCP



Sample catalog

Enterprise extension



Authorized hosted service
Approved policy sources
Identity and access checks

The workflow starts the demo service for the investigation.



Tool contracts

Request

get_investigation_context

service
environment
ruleIds
policyVersion

Response

status: SUCCESS or ERROR
policyVersion
data or error
evidence references
evaluatedAt
correlationId

A contract defines the format. Evidence supports the claim.



Controlled report delivery

ANALYSIS

OUTPUT CONTROLS

DELIVERY



The publisher checks source identity and the current PR head.



A shared module has several callers



A consumer is an application that uses the shared component.



The same change affects callers differently

Shared default: true becomes false

Caller	Input	Before	After	Result
Application A	Explicit true	true	true	PASS
Application B	Use default	true	false	FAIL

Sample requirement: the effective setting must remain true.



Consumer checks before model use



No finding

Skip the model

Verified finding

Investigate and report

Invalid evidence

Request manual review

The shared-review selector runs before model use.



The shared-module investigation result

Demo 2: PR #30

Signal

Observed result

Baseline CI

PASS

Consumer safety gate

FAIL

Investigation report

DELIVERED

One of two registered sample callers inherits the unsafe default.



What controls merging

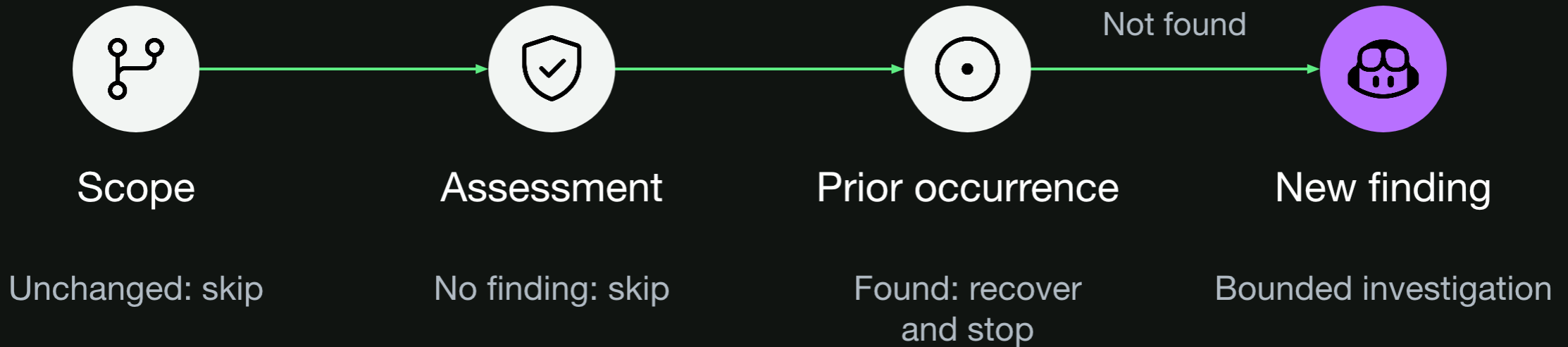
Check	Result	Meaning
Baseline CI	PASS	The baseline checks completed
Shared Module Safety Gate	FAIL	A consumer requirement is violated
Shared Module Review	SUCCESS	The report reached the PR



Merge blocked until the required finding is fixed



Avoiding unnecessary model runs



Shared review selects before inference and recovers recorded reports.

The failed-CI path currently investigates eligible failed runs.



Models, budgets and audit

Agent configuration

```
max-ai-credits: 120  
timeout-minutes: 10
```

Detection has a separate budget.

Current model setting: auto

Evaluate smaller models using the same correctness cases.

Observed usage in AIC

Prepared run	Agent	Detection
Demo 1	15.8	0.597
Demo 2	14.6	0.538



Measuring organization value

Developer effort avoided

less report review effort
less operating cost

Measure against ordinary CI
with useful deterministic context.

What to measure

Accepted findings

Correction and review effort

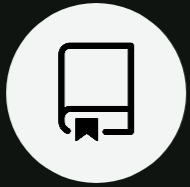
Time to a useful handoff

Inference and runner cost



Connecting enterprise services

Demo today



One repository
Two sample callers
Sample policies

Enterprise connection points



Authorized
consumer
inventory



Policy and
ownership
tools



Versions
Inputs and
relevant state

Access scope, ownership and freshness remain explicit.



What to take back to your team



CI checks keep enforcement.

Selected investigation explains the evidence.

Engineers receive a reviewable next step.

Questions

