

AI AND SOFTWARE ENGINEERING

99% Confidence, 100% Wrong

| When AI gets bold,
QA gets busy

A funny, practical look at QA in the age of AI.

PRESENTED BY

01

Sunita McCoy

Global QA Head &
professional bug hunter



02

Dev Panajkar

AI and product optimist



SoftEd

Live Virtual Event

AI and Software Engineering

September 24, 2026 • 1:00 PM - 5:30 PM EDT

THE NEW REALITY

AI is now part of the delivery chain

AI writes

Requirements, stories, release notes

AI summarizes

Tickets, calls, incident history

AI codes

Features, fixes, refactors

AI tests

Cases, data, automation scripts

AI explains

Defects, root cause, customer replies

AI automates

Triage, routing, approvals

And sometimes, AI makes things up with excellent grammar.

Good QA principles did not expire

01

Trust, but verify

Approval is a step, not a mood.

02

Shift left

Review the input before the output ships.

03

Test early, test often

Frequency beats heroics.

04

Automate what repeats

Humans for judgment, scripts for volume.

05

Review what matters

Depth follows risk, not habit.

06

Happy path and ugly path

Users are creative in the worst way.

07

Find it before the customer does

Early bugs are cheap bugs.

THE ANSWER

Start with the QA version of QA. Then add paranoia.

Requirements still need clarity. Test data still matters. Edge cases still exist.

MEET AI, THE CONFIDENT INTERN

AI is the most confident intern ever hired

PRO

Works 24/7

PRO

**Never
complains**

PRO

**Responds
instantly**

PRO

**Sounds very
senior**

CATCH

**Still needs
supervision**

It never says “I don’t know.” It says “Certainly,” then deletes the login button.

THE PROBLEM WITH CONFIDENCE

Confidence looks like quality

Clear	LOOKS FINISHED
Fast	LOOKS FINISHED
Structured	LOOKS FINISHED
Polished	LOOKS FINISHED
Persuasive	LOOKS FINISHED
Completely wrong	IS NOT FINISHED

WHY IT SLIPS THROUGH

A human mistake looks suspicious.

An AI mistake arrives with headings, bullets and a summary.

Confidence is not proof.

Confidence is just the bug wearing a blazer.

WHICH EXISTING QA FRAMEWORKS STILL APPLY

The old playbook still works, with upgrades

01	Test Pyramid	Balance the layers	06	Exploratory Testing	Matters more now
02	Shift Left Testing	Now covers prompts	07	User Acceptance Testing	Useful, not impressive
03	Risk-Based Testing	Sorts the AI use cases	08	Root Cause Analysis	Prompt, data or model
04	Regression Testing	Catches side effects	09	Defect Severity / Priority	Grades the damage
05	Boundary Value Analysis	More boundaries, unwritten	10	Equivalence Partitioning	Still trims the cases

AI did not cancel boundary values. It created more boundaries nobody wrote down.

Same frameworks. New test objects.

WHAT QA ALWAYS TESTED

Screens, Code, APIs, Data,
Workflows.....

WHAT AI-QA ALSO TESTS

Prompts

Model outputs

Assumptions

Human approvals

Source grounding

Guardrails

Escalations

Regression now asks: did yesterday's good answer become today's hallucination?

CLASSIC AI BUGS

New defects. Same old chaos.

BUG 01

Invents requirements

BUG 02

Skips dependencies

BUG 03

Misreads context

BUG 04

Generates weak tests

BUG 05

Misses edge cases

BUG 06

Fabricates facts

BUG 07

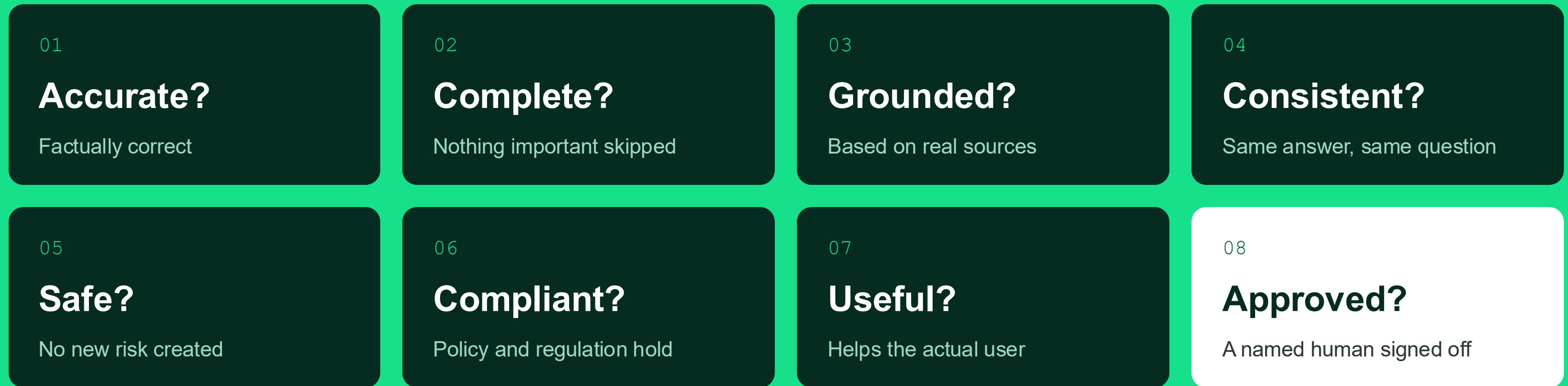
**Summarizes the
opposite of what
happened**

BUG 08

**Makes a risky output
sound harmless**

Before AI, bugs hid in code. Now bugs also hide in sentences.

Before using AI output, check the basics



“Looks good” usually means “I did not read it carefully.”

AI quality needs three checks

01 / INPUT QUALITY

Did we give AI the right context?

Prompt, source, data, constraints.

02 / OUTPUT QUALITY

Is the result correct and useful?

Accurate, complete, consistent, usable.

03 / CONTROL QUALITY

Are review and escalation clear?

Human review, audit trail, escalation.

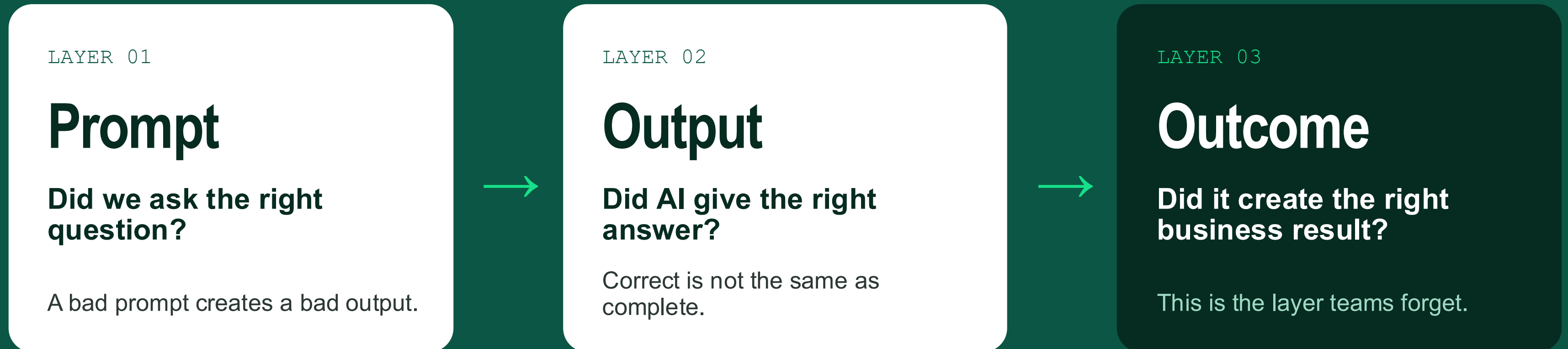
“The AI said so” is not a control.

Not every AI output needs the same review

RISK TIER	TYPICAL USE CASES	QA GATE
Low LIGHT TOUCH	Internal drafts, brainstorming, formatting	Light review before use
Medium VALIDATE	Customer support, summaries, recommendations	Check against sources and business rules
High GATED	Legal, financial, medical, security, compliance, production actions	Human approval, audit trail, sometimes blocked automation

The higher the risk, the stronger the QA gate.

Test the full chain



A good-looking output can still create a bad business outcome.

Decide where humans must step in

Customer-facing	High-risk
Low-confidence	Compliance-sensitive
Financially material	Irreversible
Weak or missing sources	ROUTE ALL OF THESE TO A NAMED REVIEWER

THREE MODES TO DESIGN

AI acts
Reversible, low stakes, logged

AI suggests
A human accepts or edits

AI waits
Nothing moves without approval

Design the gate, or the human is not in the loop. They are reading the incident report. 13 / 15

AI behavior can change on its own

WATCH 01

Answer consistency

WATCH 02

Source changes

WATCH 03

Model drift

WATCH 04

Prompt drift

WATCH 05

Data drift

WATCH 06

Policy changes

WATCH 07

**Performance
degradation**

WATCH 08

New failure patterns

Keep a golden test set: benchmark prompts and expected answer patterns you re-run forever.

The new regression question: did the same question quietly get a different answer?

QA becomes the trust layer of AI

AI

gives us speed

QA

gives us confidence

Humans

give us judgment

The best teams stop asking “can AI do this?” and start asking “can we trust this output enough to use it?”

When AI gets bold, QA gets busy.

Available at
amazon

