

Phase 06 — Test

Vertrauen in klassische Software und KI-Verhalten herstellen

Die Test-Phase sichert die Qualität der Ziellösung auf zwei Ebenen: klassisches, risikobasiertes Testen nach ISTQB/ISO 29119 für die deterministischen Anteile — und systematische Evaluation (Evals) für das Verhalten der KI-Agenten. Testfälle werden agentisch aus den Anforderungen generiert und bleiben mit ihnen verknüpft.

Ziel der Phase

Ergebnis ist ein belastbarer Qualitätsnachweis: erfüllte Testabdeckung entlang der Testpyramide, bestandene Abnahmetests mit dem Fachbereich und dokumentierte Eval-Ergebnisse für alle agentischen Komponenten.

Vorgehensmodell

- 01

Teststrategie & Risikoanalyse
Ableitung der Teststrategie aus den Qualitätsszenarien der Design-Phase; risikobasierte Priorisierung der Testtiefe je Komponente und Prozess nach Schadenspotenzial und Fehlerwahrscheinlichkeit.
- 02

Agentische Testfallgenerierung
Automatische Zerlegung der Spezifikation in atomare Gherkin-Anforderungen und Ableitung vollständiger Testfälle in Signoff; Abdeckungsanalyse gegen die Spezifikationsklauseln.
- 03

Testautomatisierung entlang der Pyramide
Unit-, Integrations- und End-to-End-Tests in der CI-Strecke; Contract-Tests für API- und Tool-Schnittstellen der Agenten; Mutationstests für kritische Logik.
- 04

Agent-Evaluation (Evals)
Szenariobasierte Bewertung des Agentenverhaltens: Aufgabenerfolg, Robustheit gegen Prompt-Variation, Guardrail-Einhaltung, Halluzinationsrate; Regression-Evals bei jeder Modell- oder Prompt-Änderung.
- 05

Abnahme & Qualitätsnachweis
Fachliche Abnahmetests mit dem Kunden (UAT), Konsolidierung aller Ergebnisse im Qualitätsnachweis; Freigabeempfehlung für die Deploy-Phase.

Methodik-Mix

METHODE	EINSATZZWECK	REFERENZ
ISTQB-Testprozess	Standardisierter Testprozess von Planung bis Abschlussbewertung; gemeinsames Vokabular im Engagement.	ISTQB
ISO/IEC/IEEE 29119	Internationale Normenreihe für Testprozesse, -dokumentation und -techniken; Rahmen des Qualitätsnachweises.	ISO/IEC/IEEE 29119-1:2022
Testpyramide	Verteilung der Testautomatisierung über Unit-, Integrations- und UI-Ebene; Kosteneffizienz und schnelles Feedback.	Vocke / Fowler (2018)

METHODE	EINSATZZWECK	REFERENZ
Risikobasiertes Testen	Steuerung von Testtiefe und -reihenfolge über Produktrisiken; fokussiert das Budget auf die kritischen Pfade.	ISTQB Advanced
Mutationstests	Empirische Bewertung der Testsuite-Güte durch injizierte Fehler; Pflicht für migrierte Geschäftslogik.	Jia & Harman (2011)
LLM-/Agenten-Evaluation	Systematische, szenariobasierte Bewertung von KI-Verhalten (Holistic Evaluation); Grundlage der ReqPOOL-Eval-Suiten.	Liang et al. (2022), HELM

Plattform-Unterstützung

ReqPOOL Suite: **Signoff** (KI-gestützte Abnahme)

Signoff — Spezifikationsanalyse	Die Spezifikation (PDF, DOCX, Text) wird in atomare, prüfbare Anforderungen im Gherkin-Format zerlegt — jede mit eindeutiger ID und wörtlicher Referenz auf die Quellklausel.
Signoff — Testfallgenerierung	Vollständige Testfälle inklusive Testdaten, Rand- und Negativfällen — automatisch verknüpft in einer Traceability-Matrix; der Abdeckungsgrad wird transparent ausgewiesen.
Signoff — agentische Testdurchführung	Ein KI-Agent testet das Zielsystem direkt und sammelt Evidenz (Screenshots, Logs); alternativ arbeitet der Anwender ein geführtes Playbook ab, dessen Bildschirmaufnahme multimodal ausgewertet wird.
Signoff — unabhängige Bewertung	Ein unabhängiges KI-Prüfmodell bewertet jeden Testfall (erfüllt / nicht erfüllt / nicht prüfbar) und schlägt Bugfixes vor.
Signoff — reversionssicherer Qualitätsnachweis	Abnahmebericht (PDF) mit vollständiger Traceability von der Spezifikationsklausel bis zum Testergebnis; Excel-Export und Audit-Log.

Best-Practice-Checkliste

	Testfälle und Testläufe vor der Ausführung durch den Projektmanager freigeben — keine ungenehmigten Läufe gegen Kundensysteme.
	Generierte Testfälle wie generierten Code behandeln: fachliches Review der kritischen Fälle, keine ungeprüfte Übernahme.
	Testszenarien aus realen Prozessdaten und Edge Cases der Analyse-Phase ableiten — synthetische Happy-Path-Tests erzeugen Scheinsicherheit.
	Abdeckungsgrad gegen die Spezifikation ausweisen; ungeprüfte Klauseln blockieren das Test-Exit-Gate.
	Jede Bewertung nur mit Evidenz (Screenshot, Log) akzeptieren; Bewertungen des Prüfmodells nicht übersteuern, sondern Abweichungen fachlich klären.
	Nach Bugfixes die betroffenen Testfälle erneut ausführen und den Regressionsnachweis im Bericht dokumentieren.

Artefakte & Ergebnisse

- Teststrategie mit Risikobewertung
- Automatisierte Testsuiten entlang der Pyramide
- Eval-Berichte je agentischer Komponente
- Äquivalenznachweis für migrierte Systeme
- Auditfähiger Qualitätsnachweis (ISO 29119)

Quality-Gate-Kriterien

☐	Risikobasierte Abdeckungsziele erreicht und nachgewiesen
☐	Alle Abnahmetests vom Fachbereich bestätigt
☐	Eval-Schwellwerte aller Agenten eingehalten
☐	Keine offenen Defects der Prioritäten 1 und 2

Wissenschaftliche Vertiefung

ISTQB: Syllabi und Glossar — Internationaler Standard der Testerausbildung und -terminologie.

<https://www.istqb.org/>

ISO/IEC/IEEE 29119-1:2022 — Software Testing — Normenreihe für Testprozesse und -dokumentation.

<https://www.iso.org/standard/81291.html>

Vocke: The Practical Test Pyramid (martinfowler.com) — Praxisreferenz zur Testautomatisierungs-Verteilung.

<https://martinfowler.com/articles/practical-test-pyramid.html>

Jia & Harman: An Analysis and Survey of Mutation Testing (IEEE TSE 2011) — Umfassender Forschungsüberblick zu Mutationstests.

<https://doi.org/10.1109/TSE.2010.62>

Liang et al.: Holistic Evaluation of Language Models (HELM) — Wissenschaftlicher Rahmen der mehrdimensionalen LLM-Evaluation.

<https://crfm.stanford.edu/helm/>

Anthropic: Statistical Approach to Model Evals — Statistische Fundierung von Eval-Vergleichen.

<https://www.anthropic.com/research/statistical-approach-to-model-evals>