

Corporate · Business Readiness

OAT Test Script Generator

CAI Score L2 · Draft — Human reviews & approves all output

▲ this agent

L1 Assist	L2 Draft	L3 Operate	L4 Decide	L5 Autonomous
TIER	PILLAR	AUTOMATION	HUMAN ROLE	
L2 · Draft	Corporate	≈55% of routine effort	Reviews and approves all output	

01 — Executive summary

OAT Test Script Generator is a CAI Score L2 (Draft) agent for corporate. Generates Operational Acceptance Testing scripts from process flows, improving coverage while cutting authoring effort. It operates with a human review gate: all output is approved before use. Indicatively, it can redirect around 55% of the routine effort this work consumes.

02 — Problem & context

- Test scripts are authored by hand, so coverage depends on who is free.
- Thin test coverage means defects slip into production.

03 — Representative use cases

- Generates OAT scripts from process flows, covering all documented steps and exception paths.
- Improves test coverage consistency: script quality no longer depends on who is available to author them.
- Reduces OAT script authoring time from days to hours, freeing testers for execution and analysis.
- Produces scripts in a consistent format that can be imported directly into test management tools.
- Produces the evidence to track script coverage, defect escape rate, and authoring effort saved.

04 — How to use it

- 1 Connect it in your tenant to the systems this work already touches (business readiness).
- 2 Confirm scope, the named reviewer, and the approval the tier requires.
- 3 Receives the task trigger or request
- 4 Produces the draft or analysis output
- 5 Human reviews and approves before use
- 6 Logs every output for audit
- 7 Review quality and acceptance rate after each cycle; tune scope as you learn.

05 — Where it fits in the process

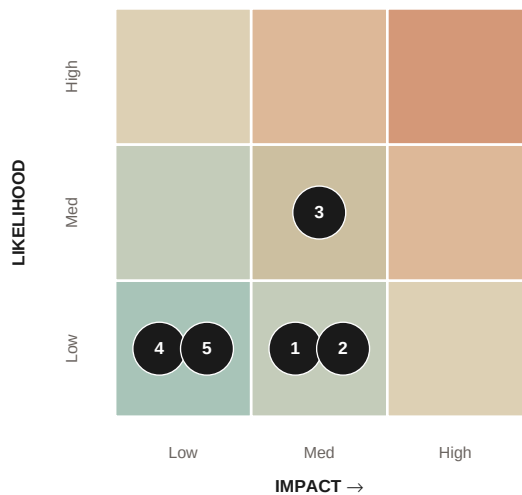
Sits in the readiness and testing step — generating OAT scripts from the process flow so testers refine rather than author from scratch.



Drafting loop for a L2 agent — human approves all output.

06 — Risk assessment

#	RISK	LIKELIHOOD	IMPACT	MITIGATION / CONTROL
1	Incorrect or incomplete draft	Low	Medium	Draft is reviewed and approved by a human before any use; errors are caught at the review gate.
2	Over-reliance / reduced diligence	Low	Medium	Tier design keeps a human in the approval seat; acceptance-rate KPIs are tracked each cycle.
3	Exposure of sensitive data	Medium	Medium	Runs inside your own tenant; Colleague AI processes none of your data. Access stays under your existing controls.
4	Scope creep or misuse	Low	Low	Scope is bounded and documented; the agent produces output only — it takes no autonomous action.
5	Quality drift over time	Low	Low	Periodic sampling and KPI review detects drift; prompts are version-controlled and change-managed.



Residual likelihood × impact after the tier's built-in controls. Numbers map to the table above.

07 — Regulatory & compliance impact

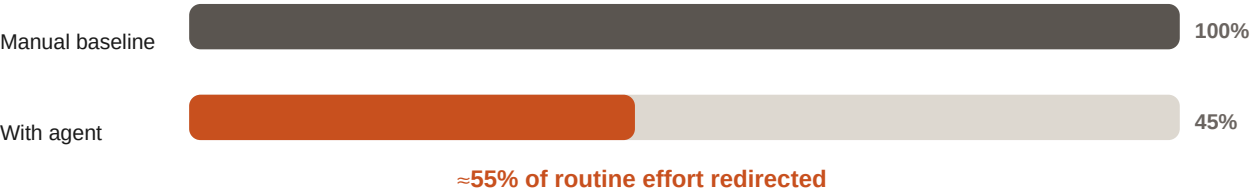
FRAMEWORK	RELEVANCE	HOW THIS AGENT ALIGNS
EU AI Act	Operational use under human oversight; logging and risk-tiering apply.	Documented risk tier, human oversight and event logging per agent — designed to map to the obligations.
DORA	Relevant where the function is in scope for financial-sector operational resilience.	Actions are logged and attributable, supporting resilience and traceability requirements — designed to map to.
GDPR	Applies wherever personal data is in scope of the workflow.	Data stays in your tenant; Colleague AI processes none of it. Purpose limitation and minimisation are preserved.
ISO/IEC 42001	AI management-system controls: roles, monitoring, review.	The agent operates within a documented management system — defined tier, named roles, logging and review. Designed to meet; not certified.

“Designed to map to” / “designed to meet” describe alignment with the CAI Score framework, not external certification. Confirm obligations with your compliance function.

08 — Value & illustrative ROI

Generates Operational Acceptance Testing scripts from process flows, improving coverage while cutting authoring effort.

Assumptions (illustrative — replace with your own)	Result
3 FTE engaged · 1840 productive h/FTE/yr · 55% addressable · €60/h blended	≈ 3,036 hours/yr redirected · ≈ €182,000/yr indicative value



Figures are an illustrative model to be validated against your own volumes and rates – not a guarantee.

KPIs to track

Improved test coverage and reduced authoring effort.

09 — Recommendations

- Connect the agent to the process flow repository and test management tool before deployment.
- A tester reviews and approves every generated script before it is used in a test cycle — this is an L2 agent.
- Prioritise high-risk, high-complexity processes for first script generation.
- Baseline test coverage and defect escape rate before go-live.

10 — Governance & method

This agent runs in your own Microsoft/Azure tenant; Colleague AI hosts only the control plane (score, policy, audit) and processes none of your data. A person reviews and approves all output before it is used. Every action is time-stamped and attributable. The CAI Score classification is documented and open to challenge.

This dossier is an evaluation document. ROI and automation figures are illustrative models, not commitments. Compliance statements describe the CAI Score framework and are not a certification of your compliance with any law or standard. © Colleague AI – proprietary. CAI Score™.