TEMPLATE P22 · VALIDATION AND PILOT RESULTS

Validation and pilot results

Gathers the results of the tests and of the pilot under real conditions, and compares them with the thresholds and the value hypothesis set before investing.

Fernando García Varela Version 0.1 (working draft) 16-09-2026 Draft for review

IN THIS DOCUMENT

01 Instructions for use

02 Identification

03 Test results

04 Specific evaluation of generative AI and agents (Enterprise)

05 Pilot design

06 Pilot results against the hypothesis

07 Observed value and projection

08 Stop criteria

09 G5 criteria according to the ambition level

10 Team conclusion and recommendation

11 Quality criteria

12 Approval and verification

13 Version control

Legal notice and disclaimer. SEVEN-G is a reference methodological framework provided "as is" and for information purposes only. It does not constitute legal, regulatory, financial or professional advice, nor does it guarantee compliance with any law or standard. References to general regulation (such as the EU AI Act, the GDPR, DORA or NIS2), to technical standards and to regulation specific to each sector or jurisdiction may be incomplete, may not apply to a particular case or may become out of date as a result of regulatory changes, interpretations or supervisory positions after their consultation date. **Each organisation that uses SEVEN-G is solely responsible for identifying the regulation that applies to it, verifying that it is current and certifying its own regulatory compliance**, with the appropriate qualified advice. This methodology is a generic, free aid shared with the community so that nobody has to start from scratch; each person or organisation can and should adapt it to its own use. It must not be inferred that its legally sensitive parts have been reviewed by legal counsel: those reviews, for each company or sector, are the ultimate responsibility of the company, consultant or organisation that uses it. Although every effort is made to keep it up to date, some regulation may have changed without being reflected here. To the fullest extent permitted by law, the author accepts no responsibility whatsoever for the effects of its application in any organisation or for its full applicability. The methodology does not grant certification of any kind. The author accepts no liability for any use made of this content or for any decisions taken on the basis of it. The data, figures, companies and cases in the examples are fictitious or illustrative.

01

Instructions for use

Aspect	Guidance
When it is used	Phase 5 (Delivery and validation), after performing the tests and the pilot and before requesting G5.
Who completes it	AI Product Owner (pilot and value) and AI Technical Owner (tests).

Aspect	Guidance
Who verifies it	AI Risk Owner under Lite; AI Auditor under Enterprise. Value is validated by management control when it is declared as *validated*.
At which *gate* it is reviewed	Go-live. (G5)
Lite versus Enterprise	Lite may omit the fields marked **(Enterprise)**.
Tool	T11 Value hypothesis canvas and calculator.
Relationship	P08 Value hypothesis canvas · P09 Baseline · P17 Human oversight · P18 Security design · P19 Rollback test · P20 Adoption · P21 Delivery report · P28 Value monitoring · document 40.

Rules:

- **The thresholds were set beforehand.** The metrics, success thresholds and stop criteria are those approved in phase 2 (P08). They cannot be relaxed to avoid a decision to stop without approval from the body that authorised the use case (01 §7.4, rule 6).
- **Production is the only truth.** Laboratory results do not replace the pilot under real conditions.
- Every amount has a formula, is incremental against the baseline or the control group and has a status (measurement rules 1 and 2).
- The annual extrapolation of a pilot is always **estimated** until it is measured in production.

02

Identification

Field	Content	Guidance
Use case code		IA-AAAA-NNN (T01).
AI system(s) and version tested		Identifier (T02) and version (P16).
Ambition level		Optimise · Augment · Transform.
Intensity		Lite · Enterprise.
Document version		
Author		Role and name.

Field	Content	Guidance
Date		DD-MM-AAAA.
Document status		Draft · Submitted for verification · Verified · Approved · Superseded.

03

Test results

Test type: Functional · Model performance · Bias and non-discrimination · Robustness · Security · Prompt injection · Load and latency · Usability · Human oversight · Rollback. **Result:** Passed · Passed with observations · Failed · Pending.

Test type	Metric	Acceptance threshold and date on which it was set	Result obtained	Status	Evidence
Model performance ***(illustrative example)***	Proportion of correct classifications on an independent test set	≥ 90 % · set in P08 on 03-03-2026	92.4 %	Passed	Evaluation report v2
Bias and non-discrimination ***(illustrative example)***	Difference in error rate between age groups	≤ 3 points	4.1 points in the over-65 group	Failed	Bias report v1
Prompt injection					P18 block 11
Rollback	Time to safe state	Target time from P19			P19 block 10

Every *Failed* test indicates its treatment and whether it blocks go-live:

Failed test	Does it affect a critical security, compliance or human oversight control? (Yes · No)	Treatment	Owner	Date

04

Specific evaluation of generative AI and agents (Enterprise)

Aspect	Evaluation method	Sample	Result	Threshold	Status
Responses grounded in the authorised sources					
Fabricated or incorrect content					
Inappropriate content or content contrary to the use policy					
Leakage of confidential or personal information					
Agent actions within limits and purpose					

05

Pilot design

Field	Content	Guidance
Period		Start and end dates.
Population and scope		Area, cases, users and affected persons included.
Attribution method		The one approved in P08: control group · before and after · other, justified.
Control or reference group		Composition and how it was assigned.
Sample size and justification (Enterprise)		Why the size makes it possible to distinguish the effect from noise.
Autonomy level during the pilot		A0 · A1 · A2 · A3 (P17).
Deviations from the pilot design		Changes during the pilot and their effect on the validity of the results.

06

Pilot results against the hypothesis

Metric	Baseline (P09)	Target	Success threshold	Pilot result	Control group result	Attributable difference	Does it meet the threshold? (Yes · No · Partially)
Primary metric:							
Secondary metric:							
Adoption (P20)					—	—	
Human override rate					—	—	

07

Observed value and projection

Type: Efficiencies · Return · Recurring cost. **Status:** Validated · Declared · Estimated.

Item	Type	Formula (units × unit value)	Amount in the pilot period (€)	Status	Annual projection (€)	Status of the projection
Materialised reduction in processing time ***(illustrative example)***	Efficiencies	2,300 cases × €3.10 of external cost avoided per case	7,130 (3-month pilot)	Declared	28,520 (× 4 quarters, same volume)	Estimated
Model consumption and licences ***(illustrative example)***	Recurring cost	3 months × €1,900	5,700	Validated	22,800	Estimated

Magnitude	Projected annual amount (€)	Status
Efficiencies		

Magnitude	Projected annual amount (€)	Status
Return		
Recurring cost		
Annual net value = efficiencies + return − recurring cost		Estimated
Released capacity (hours, not added)		
Risk avoided and compliance (not added unless there is a formula)		

08

Stop criteria

Stop criterion set in phase 2	Threshold	Result	Has it been triggered? (Yes · No)

If any criterion has been triggered, the team must recommend **Stop** or justify to the body that authorised the use case why it proposes a different result.

09

G5 criteria according to the ambition level

Level	What G5 requires (01 §7.6)	Evidence provided	Status (Met · Not met · Not applicable · Pending)
Optimise	Efficiency validated against the baseline; plan to materialise the released capacity.		
Augment	Actual adoption and performance improvement measured.		
Transform	Market or customer evidence: use, conversion, initial revenue or verified operational change.		

10

Team conclusion and recommendation

Field	Content	Guidance
Conclusion		Does the hypothesis hold under real conditions?
Recommendation to the *gate*		Proceed · Proceed with conditions · Iterate · Stop. Pivot is not possible at G5 (01 §7.3). It is a recommendation; the competent body decides in P29.
Proposed conditions		Only on non-critical aspects, with a deadline and an owner.

11

Quality criteria

Formal G5 criteria in document 21; checklist LV-G5 in document 22.

#	Check	Status (Met · Not met · Not applicable · Pending)
1	The thresholds and stop criteria are those set in phase 2, dated before the pilot.	
2	The pilot has been carried out under real conditions with the approved attribution method.	
3	The bias, robustness, security and prompt injection tests have been performed.	
4	Every amount has a formula, type and status; the annual projection is shown as estimated.	
5	Released capacity is reported separately and is not added.	
6	Failed tests on critical controls have been resolved.	
7	The recommendation is consistent with the results and with the stop criteria.	

12

Approval and verification

Function	Role	Name	Date	Signature or reference
Author (pilot and value)	AI Product Owner			
Author (tests)	AI Technical Owner			
Validation of amounts	Management control			
Verifier	AI Risk Owner (Lite) · AI Auditor (Enterprise)			
***Gate* decision-maker**	According to 01 §7.5 (recorded in P29)			

Segregation of duties: the verifier and whoever validates the amounts are not part of the team that builds.

13

Version control

Version	Date	Changes
0.1	16-09-2026	First version of the template.