

AIE-004 | Model Qualification Matrix — Task × Locale × Risk v1

Model Qualification: Task/Locale/Risk Benchmark

Bounded outcome: Build a reproducible qualification benchmark over representative task × locale × risk slices, declare metrics and hard failure gates before testing, retain row-level evidence and uncertainty, bind results to exact model/configuration/prompt/dataset versions, and issue bounded QUALIFIED / NOT_QUALIFIED / CONDITIONAL decisions for downstream consumption.

Evidence path

Boundary	Inspectable claim	Explicit non-claim
candidate id	Treat candidate id as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	No public leaderboard or single aggregate score is treated as sufficient product qualification evidence.
model version	Treat model version as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	Qualification is not a permanent global property of a model name; it is bounded to exact version and slice evidence.
configuration version	Treat configuration version as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	No runtime route selection, local-first policy or cost optimization ownership; AIE-005 and later lessons consume qualification evidence.
prompt version	Treat prompt version as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	Model-based judges are evaluation instruments, not automatic ground truth or final high-risk authority.
task slice	Treat task slice as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	No public leaderboard or single aggregate score is treated as sufficient product qualification evidence.
EVIDENCE BOUNDARY	Separate source facts, local observations, bounded inferences, and human decisions.	Unavailable production behavior remains NOT_OBSERVED or NEEDS VERIFICATION.

Acceptance gate: Accept only when every critical dimension has inspectable evidence, the negative probe behaves as predicted, and changed-context transfer preserves a residual limitation.

AIE-004 | Blank Evidence Record

Complete every field from retained observations; mark unavailable evidence OPEN.

Canonical field	Observed evidence / value	Evidence type	State
Candidate		OBSERVED / INFERRED	OPEN
Version binding		OBSERVED / INFERRED	OPEN
Task		OBSERVED / INFERRED	OPEN
Locale		OBSERVED / INFERRED	OPEN
Risk		OBSERVED / INFERRED	OPEN
Aggregate		OBSERVED / INFERRED	OPEN
Critical hard gate		OBSERVED / INFERRED	OPEN
Evidence gap		OBSERVED / INFERRED	OPEN
Decision		OBSERVED / INFERRED	OPEN

Source projection: derived_text/tables/model_qualification_matrix.md

AIE-004 | Evidence Index

Link every decision to a source, command, output, or explicitly bounded inference.

ID	Evidence type	Source / command	Observation	Supports / falsifies	State
E-01	BASELINE	derived_text/tables/ model_qualification_matrix.md			OPEN
E-02	RUNTIME	python3 lab/AIE004-LAB1/run_checks.py			OPEN
E-03	NEGATIVE PROBE	A public leaderboard or one aggregate average is not sufficient evidence that a model is.			OPEN
E-04	RECOVERY	Candidate Id evidence is inspectable and its non-claim remains explicit.			OPEN
E-05	TRANSFER	EX4			OPEN

Truth rule: Runtime and security claims require retained output from the supplied bounded fixture. Source facts and static readiness are not promoted into production observations.

AIE-004 | Decision Walkthrough

Run the evidence gates in order and preserve unresolved boundaries.

Case	Change	Prediction	Evidence to retain	Decision
SIMPLE	EX1	Treat candidate id as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	Retain bounded positive or negative evidence.	OPEN
REALISTIC	EX2	Treat model version as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	Retain bounded positive or negative evidence.	OPEN
EDGE OR FAILURE	EX3	Treat configuration version as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	Retain bounded positive or negative evidence.	OPEN
TRANSFER NEAR	EX4	Treat prompt version as a separate evidence boundary and close it only with inspectable observations from the bounded fixture.	Retain bounded positive or negative evidence.	OPEN

Allowed states: ACCEPT, REPAIR, DEFER

Guardrail: A local deterministic fixture does not prove production identity, authorization, confidentiality, integrity, availability, recovery, enforcement, or system-wide security.

AIE-004 | Failure and Repair

Falsify shortcuts before accepting the bounded decision.

Failure	Signal	Repair	Recovery check
A public leaderboard or one aggregate average is not sufficient evidence that a model is.	No retained evidence closes the candidate id boundary.	Retain the fixture output that distinguishes observed candidate id facts from inference, unknowns, and.	Candidate Id evidence is inspectable and its non-claim remains explicit.
A model can be qualified for one task/locale/risk slice and not qualified for another.	No retained evidence closes the model version boundary.	Retain the fixture output that distinguishes observed model version facts from inference, unknowns, and.	Model Version evidence is inspectable and its non-claim remains explicit.
This lesson does not choose runtime routing, local-first policy or cost optimization; AIE-005.	No retained evidence closes the configuration version boundary.	Retain the fixture output that distinguishes observed configuration version facts from inference.	Configuration Version evidence is inspectable and its non-claim remains explicit.
Model-based judges may be useful evaluation instruments but are not automatically ground.	No retained evidence closes the prompt version boundary.	Retain the fixture output that distinguishes observed prompt version facts from inference, unknowns.	Prompt Version evidence is inspectable and its non-claim remains explicit.

Transfer case: EX4

Transfer observation: _____

Recomputed decision: _____

AIE-004 | Review and Sign-off

Acceptance requires inspectable evidence plus an explicit residual limitation.

Review control	Required response	Reviewer evidence	Disposition
Canonical fields addressed	YES		OPEN
Negative probe retained	YES		OPEN
Recovery evidence retained	YES		OPEN
Changed-context transfer complete	YES		OPEN
Source facts separated from inference	YES		OPEN
Residual limitation explicit	YES		OPEN
Final bounded decision	ACCEPT / REPAIR / DEFER		OPEN

First explicit non-claim: A public leaderboard or one aggregate average is not sufficient evidence that a model is qualified for this product workflow.

Reviewer: _____ Date: _____ Record owner: _____