

The claim

A board approves an AI system on a date. The data it was fitted to has a date, the process it was fitted to has a date, the population it will score has a date. Only the first is written into the minute. The authorisation carries no expiry, which silently asserts that the other three never move. They move.

Our position: validation is a cycle, not an event. Every AI authorisation should state a re-validation interval, name the performer and the budget, and state a decommission trigger — the condition on which the system is switched off rather than retrained. An approval with no expiry is not governance. It is a minute of a meeting.

The evidence

Decay is ordinary, not exceptional.

EVIDENCE

An analysis of temporal model degradation across 128 (model, dataset) pairs spanning healthcare operations, transportation, finance and weather, using four standard machine-learning model families, observed temporal degradation in 91% of cases, and distinguishes it from data concept drift and continuous learning as related but separate phenomena.

Source: Vela, D., Sharp, A., Zhang, R., Nguyen, T., Hoang, A. and Pianykh, O.S., "Temporal quality degradation in AI models", Scientific Reports 12, article 11654, 2022 · Geography: not geography-bound; datasets predominantly North American · Method: peer-reviewed empirical study across four industries · SEG-4 · Caveat: 91% is the proportion of studied pairs showing degradation, not a rate of decay and not a probability for any specific deployment; generative systems are out of scope.

It shows first in calibration, not accuracy, in clinical prediction, the most heavily validated applied modelling discipline there is.

EVIDENCE

In seven parallel regression and machine-learning models for hospital-acquired acute kidney injury developed on nationwide US Department of Veterans Affairs admissions, discrimination performance declines over time were statistically significant but small, while calibration deteriorated and observed drift was connected to shifts in the patient population.

Source: Davis, S.E., Lasko, T.A., Chen, G., Siew, E.D. and Matheny, M.E., "Calibration drift in regression and machine learning models for acute kidney injury", Journal of the American Medical Informatics Association 24(6), 1052-1061, 2017 · Geography: United States, Veterans Affairs hospital system · Method: peer-reviewed study of model performance over successive periods · SEG-4 · Caveat: one clinical outcome in one health system; we cite the direction and ordering, not a rate transferable to commercial models.

Credit limits, provisioning, pricing and fraud thresholds are set on calibration and monitored on discrimination: a board shown an accuracy chart is shown the metric that fails last.

The instrument for governing an expiring accuracy is older than machine learning: the argument was never whether to recalibrate, only how often.

EVIDENCE

ISO/IEC 17025:2017 requires a laboratory to establish a calibration programme, reviewed and adjusted as necessary to maintain confidence in the status of calibration (clause 6.4.7); ILAC-G24 / OIML D 10, Edition 2022 (E), sets out methods for determining calibration intervals, based on estimating the period over which equipment is likely to remain within prescribed limits after calibration and adjusting the interval to optimise the balance of risks and costs.

Source: ISO/IEC 17025:2017, clause 6.4.7; ILAC-G24 / OIML D 10, Edition 2022 (E), oiml.org · Geography: international · Method: standard and guidance text · SEG-5 · Caveat: written for measuring instruments, not statistical models; the transposition to AI re-validation is our reading and is endorsed by neither body.

Drift also arrives faster in the corridor.

EVIDENCE

Electronic money accounts in the WAEMU rose 18.99% in a single year, from 209 million in 2023 to 248 million in 2024, of which 76.8 million were active, an activity rate of 30.9%; over the same year the service point network fell 5.23%, from 1,678,067 points to 1,590,243, and digital financial services accounted for 57.2% of the Union's overall financial inclusion rate of 73.6%.

Source: BCEAO, "Rapport annuel sur les services financiers numériques dans l'UEMOA - 2024", bceao.int · Geography: eight WAEMU member states · Method: central bank annual statistical reporting from supervised issuers · SEG-5 · Caveat: sector aggregates describe the environment a deployed model scores, not the performance of any model; no drift measurement is claimed from them.

Nineteen per cent account growth against a five per cent contraction in service points is not the population any model was fitted to two years earlier. A vendor's home-market cadence is optimistic here by construction.

Evidence cards

CLM-TEMPORAL-MODEL-DEGRADATION · SEG-4. Claim: temporal degradation was observed in 91% of 128 (model, dataset) pairs across healthcare operations, transport, finance and weather, using four standard model families, and is distinguished from concept drift and continuous learning. Context: peer-reviewed, Scientific Reports 12:11654, 2022. Method: empirical cross-industry analysis. Contradictory evidence: 91% is a proportion of studied pairs, not a decay rate or deployment probability; datasets predominantly North American; generative systems out of scope. Causal confidence: none for any specific deployment. Transferability: establishes that decay is normal, not its magnitude. Review date: 2026-08-03.

CLM-CLINICAL-CALIBRATION-DRIFT · SEG-4. Claim: in seven parallel models for hospital-acquired acute kidney injury on nationwide US Veterans Affairs admissions, discrimination declines were statistically significant but small while calibration deteriorated, with drift connected to patient-population shifts. Context: peer-reviewed, JAMIA 24(6), 1052-1061, 2017. Method: analysis of performance over successive periods. Contradictory

evidence: one clinical outcome, one health system; the ordering may not hold for every commercial model. Causal confidence: association reported by the authors; no causal transfer claimed. Transferability: direction and ordering only. Review date: 2026-08-03.

CLM-CALIBRATION-INTERVAL-PRACTICE · SEG-5. Claim: ISO/IEC 17025:2017 clause 6.4.7 requires a calibration programme reviewed and adjusted to maintain confidence in calibration status, and ILAC-G24 / OIML D 10, Edition 2022 (E), gives methods for setting intervals by estimating how long equipment stays within prescribed limits and adjusting to optimise the balance of risks and costs. Context: international standard and guidance, current 2026. Method: standard and guidance text. Contradictory evidence: written for measuring instruments; neither body addresses AI, and the analogy carries no endorsement. Causal confidence: none — institutional fact. Transferability: as a pattern for expiry-bearing approvals. Review date: 2026-08-03.

CLM-BCEAO-DFS-GROWTH · SEG-5. Claim: WAEMU electronic money accounts grew 18.99% from 209 million (2023) to 248 million (2024), with 76.8 million active (30.9%), while service points fell 5.23% from 1,678,067 to 1,590,243, digital financial services contributing 57.2% of an overall financial inclusion rate of 73.6%. Context: BCEAO annual report on digital financial services in the WAEMU, 2024 edition. Method: central bank statistical reporting from supervised issuers. Contradictory evidence: sector aggregates say nothing about any model's inputs or performance; account counts include duplicates across issuers. Causal confidence: none; environmental change asserted, not measured drift. Transferability: WAEMU deployments scoring retail financial behaviour. Review date: 2026-08-03.

CLM-FIRM-REVALIDATION-INTERVAL · SEG-1 (firm position). Claim: every AI authorisation should carry a re-validation interval, a named performer and budget, and a decommission trigger; an approval without an expiry is not governance. Context: firm practice in the WAEMU-euro corridor. Method: interpretation built on the four cards above; no client outcomes cited. Contradictory evidence: fixed intervals can be worse than event-driven re-validation where inputs are stable; frequent re-validation costs money and can introduce instability through over-fitting to recent data; low-stakes systems may not merit the machinery. Causal confidence: none. Transferability: bounded to systems whose outputs set limits, prices or thresholds. Review date: 2026-08-03.

The limits

We are not arguing that models are unreliable or that initial validation was done badly, only that a validation result is dated evidence about a moving object.

The interval is easy to write and hard to fund. Re-validation needs three things in three places: labelled local outcome data, held by the operating company; access to the model or its case-level scored outputs, controlled by the vendor; and someone competent to compute calibration locally, who may be in neither. The budget line is usually in a fourth place, or nowhere. An interval nobody owns is an interval nobody keeps.

Sources and limitations. The caveats on the cards above bind: none of the four sources measures drift in a deployed commercial system, and none was written about AI re-validation. The re-validation rule is the firm's position, graded SEG-1, mandated by none of the cited evidence and unsupported by any dataset comparing authorisations with and without an expiry, because we found none meeting our standard. It is strongest where outputs set limits, prices or thresholds, weakest for low-stakes systems under stable inputs. This note is not valid as legal or regulatory advice, and nothing here suggests that a re-validated model is a correct one. Sources verified 2026-08-03.

What to do on Monday

- 1. The board secretary writes an expiry date into the authorisation itself and records what happens on that date if nobody acts.** If the system simply keeps running, the interval is decorative.
- 2. The model owner names three parties in writing: who re-validates, on whose data, and out of which entity's budget.** If one is unnamed, the interval will be missed for reasons later called unforeseeable.
- 3. The chief risk officer records the decommission trigger and the local officer who can pull it within the week.** Retraining is the default answer and not always the right one; a system nobody locally can switch off is not under our control.
- 4. The contract owner requires dated calibration on local outcomes in place of the vendor's accuracy pack, and obliges the vendor to disclose its own retraining cadence.** A model re-fitted offshore restarts the clock without telling anyone in Abidjan.