DATA SYMPHONY · BRIEF

Independent validation for AI-assisted financial-services models

For CROs and heads of model risk. 4-minute read.

The problem

The existing validation playbook — methodology review, replication test, controls review, documentation review — was written for deterministic models. AI-assisted workflows in actuarial, credit and finance break four of its assumptions: outputs are not fully reproducible, behaviour drifts under retrieval freshness, prompts are themselves model inputs, and the human reviewer becomes part of the model. South African PA expectations and SR 11-7-style frameworks both already demand the extensions. Most internal playbooks have not been updated.

Five extensions to the validation playbook

- Drift testing on a fixed evaluation set. Re-run the AI-assisted step against a frozen pack of cases every month and track output stability, not just accuracy. Reject the version if drift exceeds an agreed band.
- Prompt sensitivity. Treat the prompt as a model input. Vary it across a published grid (rephrasings, ordering, length) and document the output dispersion. Lock the production prompt; version it like code.
- Retrieval freshness. If the workflow uses retrieval over assumptions, policy or product documents, validate the freshness, completeness and access controls of the index — not only the generator on top.
- Reviewer effort and override rate. Instrument how long the human reviewer spends and how often they override. Both signals should sit in the validation pack alongside accuracy. Falling effort plus rising acceptance is the failure pattern to catch.
- Scoping template. Use a one-page template that names the use case, the data, the prompt, the human in the loop, the failure modes, and the override authority. Reject any AI-assisted submission without it.

What good looks like

- The board pack reports drift, prompt sensitivity, retrieval freshness and reviewer override alongside the usual model performance metrics.
- No AI-assisted model reaches production without a scoping template signed by the model owner, the reviewer and model risk.
- When the PA asks how an AI-assisted output was produced, the answer takes one meeting, not three.