

Comments on the NAIC AI Risk Evaluation Supplement, version 5.0

Falsify OÜ · Reg. 17574308 · Narva mnt 5, 10117 Tallinn, Estonia · hello@falsify.dev · 9 September 2026 · Submitted to the Big Data and Artificial Intelligence (H) Working Group, 30-day exposure ending 29 September 2026

Dear Mr. Sobel and Mr. Romero,

Falsify OÜ (Tallinn, Estonia) submits two comments on version 5.0 of the AI Risk Evaluation Supplement exposed on 31 August 2026. Both concern a single point: the Supplement asks how a high-risk model was validated and when it was last tested, but does not appear to ask the company to identify the pre-deployment acceptance criteria against which validation was judged, or when those criteria were fixed.

Comment 1 — Exhibit C, new field after field 11

Exhibit C states that AI model risk criteria are set by the insurance company, and field 11 asks the company to discuss how the model was validated prior to deployment. A validation narrative is only as informative as the bar the model was measured against. Two models can report identical validation results while one was held to criteria fixed before testing and the other to criteria selected after the results were known. Field 12 records the last date of testing, but no field appears to record when the acceptance criteria were fixed, so an examiner cannot distinguish the two cases from the response.

Proposed field, designated optional, inserted after field 11:

"Pre-deployment acceptance criteria. State the metrics and thresholds the model was required to meet before deployment, when those criteria were fixed, and what record establishes that they were fixed before validation results were available. If the criteria were revised after validation results became available, state when and why. Where available, identify evidence whose chronology does not depend solely on the validation report itself, such as a version-controlled or independently time-evidenced record of the criteria."

This adds one row to a questionnaire the company already completes, and it parallels the Supplement's treatment of materiality, where the threshold used is disclosed to the regulator rather than left implicit.

Comment 2 — Exhibit B (Narrative), item 3a

Item 3a asks whether validation and testing are performed by someone independent from development. Independence of the tester addresses one source of bias; control over when the acceptance criteria are established addresses another.

Proposed addition at the end of 3a:

"...and whether acceptance criteria are established before validation begins, by whom, and how subsequent changes to those criteria are controlled and documented."

Interest statement. Falsify maintains an open specification for tamper-evident records of evaluation criteria and their timing. Neither proposal depends on PRML or on any particular tool. Evidence could come from an internal records system, an independently timestamped record, or another mechanism that preserves both the criteria and when they were fixed.

We consent to this submission being made public.

Kind regards,

Cüneyt Öztürk
Founder, Falsify OÜ
Tallinn, Estonia