PUBLIC COMMENT SUBMISSION

Targeted Redline Recommendations on the NAIC AI Risk Evaluation Supplement — Version 7.0 Draft

TO:	NAIC Big Data and Artificial Intelligence (H) Working Group
VIA EMAIL:	Scott Sobel (ssobel@naic.org) and Miguel Romero (maromero@naic.org)
FROM:	Mebrat Hill, Premium Audit Consultant & Insurance Assurance Practitioner Mebrat.hills@gmail.com | 510-626-7028
DATE:	November 16, 2026
RE:	**Targeted Redlines to AI Risk Evaluation Supplement (Version 7.0 Clean Draft)**

Dear Chair Houdek, Vice Chairs, and Members of the Working Group:

Thank you for the opportunity to comment on the Version 7.0 draft of the Artificial Intelligence (AI) Risk Evaluation Supplement (the "Supplement"). I appreciate the Working Group's continued refinement of the Supplement and its emphasis on risk-based regulatory inquiry, examiner discretion, and coordination with the existing Handbooks.

Several refinements in Version 7.0 are consistent with evidence-based governance principles I outlined in my September 28 and October 28, 2026 submissions, including the importance of evaluating human oversight as an operating control rather than relying solely on the stated presence of a human reviewer.

In particular, the Version 7.0 Human in the Loop guidance is a meaningful refinement. Its focus on reviewer authority, qualifications, and capacity and evidence—including whether overrides and escalations occur and are documented—provides a more practical basis for evaluating the design and operating effectiveness of human-oversight controls.

Two other provisions provide useful foundations that were already present in the Version 6.0 draft and are retained in Version 7.0: the guidance under "Reviewing Company Documentation," which recognizes that regulators may use documentation to corroborate management assertions, and Exhibit B, Narrative Item 2b, which asks how human-review adequacy is assessed in high-volume transactions and how exceptions or errors are found and resolved. My recommendations below build on those existing provisions; I do not characterize them as newly introduced in Version 7.0.

To make these principles more consistently examinable while preserving the Supplement's risk-based and technology-neutral approach, I respectfully submit four targeted, optional redline recommendations. They are intended to clarify the distinction between control design and control operation, support review of selected material decisions, and improve the clarity of relevant data and governing-rule periods. They are not intended to create new substantive legal duties or prescribe a particular technical architecture.

1. General Guidance — "Reviewing Company Documentation" (printed page 4)

Rationale: The existing paragraph recognizes that documentation may be used to corroborate management assertions. The proposed revision clarifies the distinction between evidence that a control was designed and evidence that it operated when required.

Proposed replacement for the first paragraph:

> *In several instances, the Supplement suggests a mix of both regulatory inquiries and requests for documentation as a means of allowing regulators to corroborate management's assertions. Where appropriate to the risk and materiality of the use case, regulators may distinguish between evidence of control design (such as policies, procedures, and assigned responsibilities) and evidence of control operation (such as contemporaneous records and exception handling demonstrating that controls functioned as intended). As regulators decide whether inquiry alone is sufficient or corroboration is needed, regulators should consult with the MRH, FCEH, and FAH, and any other NAIC resources as needed.*

2. Exhibit B — AIS Program: Narrative Item 1d (printed page 10) and Checklist Item 3p (printed page 14)

Rationale: These complementary changes would make the narrative and checklist options more explicit about control exceptions, unauthorized or bypassed actions, and whether material remediation was effective—not merely whether routine assessments occur.

Proposed revised Narrative Item 1d:

> *1d. Discuss the process by which the effectiveness of the AIS Program, AI Systems, and individual models are assessed and modified, including how control exceptions, bypassed reviews, and required remediations are detected, escalated, addressed, and verified.*

Proposed revised Checklist Item 3p:

> *3p. Evaluates whether AI Systems are suitable for their intended use and should continue to be used as designed, including whether controls operate effectively and material exceptions are appropriately resolved.*

3. Exhibit C — High-Risk AI Models: Item 9 (printed page 16)

Rationale: This addition would connect the model-level inquiry to the operational context in which a material decision is made, helping an examiner understand how required human oversight is confirmed and how material decisions are documented.

Proposed revised Item 9:

> *9. Does the AI model automate, augment, or support decision-making? For material decisions, describe how the insurer verifies that required human review or approval occurs before a decision takes effect, and what records document the relevant model version, governing rules, human review, and final disposition.*

4. Exhibit D — Data Used in High-Risk AI Models: Column (2) (printed page 19)

Rationale: For material insurance decisions, the applicable policy period and effective date of governing rules may be as important as the data itself. Identifying relevant periods and versions can reduce ambiguity when an examiner later reviews the decision.

Proposed revised Column (2) heading:

> *(2) Describe how the AI Model uses this data, including how the output is used within operations or how input data is used for training and testing. Where relevant to the use case, identify applicable policy*

periods and effective dates of governing rules, relevant data and rule versions, and how the data relates to resulting decisions.

These focused additions are intended to support consistent, reproducible examination while preserving examiner discretion, existing Handbook authorities, proportionality, and technology neutrality. They rely on evidence appropriate to the risk and materiality of the use case and do not require a prescribed platform, reporting architecture, or universal record format.

The Working Group may use these policy and drafting recommendations; this submission does not disclose, dedicate to the public, or grant any license or rights in any proprietary implementation technology, software, or assurance methodology.

I appreciate the Working Group's consideration and would be pleased to assist with any drafting questions.

Respectfully submitted,

Mebrat Hill
Premium Audit Consultant & Insurance Assurance Practitioner
Mebrat.hills@gmail.com | 510-626-7028

VERDICT
Verdict Systems Inc. · Houston, Texas · founder@verdict.systems · +1 832 641 1892

November 12, 2026

Commissioner Nathan Houdek, Chair
Doug Ommen and Mary Block, Co-Vice Chairs
Big Data and Artificial Intelligence (H) Working Group
National Association of Insurance Commissioners
By email to Scott Sobel (ssobel@naic.org) and Miguel Romero (maromero@naic.org)

Re: Version 7.0 of the AI Risk Evaluation Supplement, and the record behind a human override

Dear Chair Houdek and Members of the Working Group:

Verdict Systems Inc. supports adoption of Version 7.0 of the AI Risk Evaluation Supplement at the Fall National Meeting. This letter asks for no change to the text. The comment period on Version 6.0 closed on October 30, and the draft minutes of the November 2 meeting record that no further substantive changes will be made before adoption.[1] We write about the work those minutes say comes next: bringing the Supplement into the examination handbooks, and further study of agentic AI.

Verdict builds technology that records automated and AI-influenced actions at the moment they occur, in a form a third party can check. We have a commercial interest in this subject and say so at the outset. Nothing below requires a particular vendor or product. We commented on Version 5.0 on September 29 and submitted statements for the record on October 8.

1 What Version 7.0 added

Version 7.0 adds guidance on human review under the heading "Human in the Loop." It tells regulators who are evaluating human oversight to consider whether the reviewer "can reverse, modify, or escalate the AI-assisted decision, and the record shows that overrides actually occur and are documented with reasons." It also asks whether effective oversight "can be demonstrated via override rates, escalation rates, etc."

We support this guidance. A written policy can say that a person reviews the AI system's work. The new language has the examiner look for proof that the person did, case by case. Commenters at the November 2 meeting drew the same line between a control that is documented and evidence that the control operated.[2]

Version 7.0 also says where the next question belongs. It states that the Supplement "does not provide guidance on standards of fieldwork or related practices," and the staff note beside that sentence explains that this is why commenters' input on standards of evidence was not incorporated.[3] We agree with that placement. How an examiner tests a record is fieldwork, and the handbooks are where fieldwork guidance belongs.

2 The question the new guidance raises

"The record shows" assumes a record the examiner can rely on. Examiners already judge this for decisions people make. An adjuster's note entered on the day a claim was denied and a memo written the week the examination was announced can describe the same decision in the same words. An examiner would give more weight to the note, because it was made at the time and before anyone knew it would be reviewed.

With AI systems the two are harder to tell apart. The usual record of what an AI system did, and of what a person did about it, is a log: entries in a file or database inside the company's own systems.

Someone with administrator access can change or delete an entry, and unless the system was built to prevent that, the changed entry looks the same as an original. An override reason typed in after a complaint arrives can carry the same date as one entered when the decision was made.

The same applies to the rates the guidance mentions. An override rate is a count of override entries divided by a count of decisions. If entries can be added or removed afterward, the rate can move without anyone's conduct having changed.

Order of events is the other difficulty, and it grows with agentic AI, which Version 7.0 defines as systems that "pursue objectives across multiple steps without requiring human input at each stage." When a system takes several steps on its own, the examiner needs to know which came first. Did the person approve the payment before it went out, or after? Version 7.0 keeps the definition and the Exhibit A column for agentic AI, and leaves the more detailed agentic questions that appeared in Version 6.0 for further study.[4]

3 Three questions an examiner can ask of any record

An examiner can test a record without a technical background. Three questions cover most of it, and each has an everyday counterpart.

(a) *Was it made at the time?* A record created when the action happened is worth more than one assembled later. The everyday counterpart is a time-stamped receipt set against a note written from memory.

(b) *If someone changed it afterward, would the change show?* Software can compute a short code from a record, called a hash and often described as a digital fingerprint. Change one character of the record and the fingerprint changes. If the fingerprint is kept where the record's owner cannot alter it, any later edit to the record can be detected. A bound ledger with numbered pages works the same way, because a missing or replaced page is obvious.

(c) *Can someone outside the company confirm the first two answers?* Recording a deed with the county is the everyday counterpart. The recorder's office keeps a dated public entry that neither buyer nor seller can go back and change, and anyone can look it up. Public transparency logs do this job for digital records. What goes to the log is the fingerprint, signed by the sender, and the log publishes it with the time it arrived. The record itself, and any consumer information in it, stays with the company. An examiner who is later handed the record can compute its fingerprint and compare it with the public entry. If the two match, the record has not changed since the time shown in the log.

A company can meet all three with openly published methods, and a regulator does not have to choose among them.[5]

4 What such a record does not show

These records have limits. A record that passes all three questions shows that it existed in a given form at a given time and, where records are linked, in a given order. It does not show that the event happened as described, that every action was recorded, or that the decision was correct. Examiners would still test completeness and accuracy by sampling, as they do now. They would be sampling records whose time, integrity and order can be checked without relying on the company's word. A company that answers that it cannot produce records like these has still told the examiner something useful about inherent risk.

5 Where this could fit in next year's work

We offer two suggestions for the work that follows adoption. Neither changes Version 7.0.

Handbook guidance. When discussion resumes on incorporating the Supplement into the Market Regulation Handbook or the Financial Condition Examiners Handbook, the three questions could be written as a short procedure for testing a sampled override. One possible wording:

> SAMPLE WORDING
>
> "For a sample of AI-assisted decisions with Direct Consumer Impact, including decisions the company reports as overridden, obtain the company's record of the AI System's output and of the human review. For each record, determine (a) whether it was created when the action occurred, (b) whether a later change to it could be detected, and (c) whether its time of creation can be confirmed by a party other than the company, without access to the company's systems."

Further study of agentic AI. For systems that act in several steps, a fourth question helps: can the records show the order of the steps? One open method is for each later record to include the fingerprint of the record before it, which fixes the order. Appendix A shows two records linked this way. Section 5.3 of our September 29 letter proposed a definition, "Contemporaneous Action Record," that could serve as shorthand in that study.

6 A worked example staff can check

Appendix A reproduces two linked records of a fictional two-step decision: an AI recommendation, then a human override with a reason. The first check needs only a web browser. The others use standard tools and a public log that Verdict does not operate. The record in Appendix A of our September 29 letter (public log index 2951501338) can still be checked the same way today.

The records were produced on Verdict's public sandbox, which signs with a short-lived key generated for each request. They demonstrate the checks described above. They do not demonstrate the identity or key custody of a production signer.

We ask only that this letter be kept with the Working Group's record for the work ahead. We would welcome questions from members or staff, and would be glad to walk through the example with anyone who wants to try it.

Respectfully submitted,

Shayne Beavan
Founder, Verdict Systems Inc.
founder@verdict.systems · +1 832 641 1892

Disclosure. Verdict Systems Inc. is a Delaware corporation based in Houston, Texas. It has filed U.S. patent applications relating to its products; none has been examined. It is not an insurer or producer. The suggestions above do not depend on any Verdict technology.

1 Draft minutes of the Working Group's November 2, 2026 meeting (draft dated November 10), Attachment A to the materials for the November 16 session. The draft minutes also say the Supplement "will continue to evolve, including through work with the Market Regulation and Consumer Affairs (D) Committee and the Financial Condition (E) Committee, as well as further study of agentic AI," and that discussion of incorporating it into the two handbooks "will continue next year."

2 Same draft minutes, item 3.

3 Comment SS6 in "AI Risk Evaluation Supplement v7.0 (Edited)," posted on the Working Group's page and retrieved November 12, 2026.

4 Version 6.0's Exhibit A question on actions taken "without human approval before the action" and its Exhibit C Ref 3a do not appear in "AI Risk Evaluation Supplement v7.0 (Clean)," retrieved November 12, 2026.

5 Cryptographic hashing, trusted timestamps under RFC 3161, and public append-only transparency logs of the kind web browsers rely on through Certificate Transparency (RFC 6962).

Appendix A. Worked example: an AI recommendation and a human override

DEMONSTRATION · FICTIONAL SCENARIO · PUBLIC SANDBOX TIER

Scenario (fictional). An AI credit model at a fictional bank re-scores a pre-approved loan offer after the customer's salary deposits drop, and recommends withdrawing the offer. Bank policy sends the recommendation to a lending officer before any message reaches the customer. The officer confirms with the employer that salary has resumed and approves a reduced amount with a 60-day review, and the reason is recorded. In the Supplement's words, this is an override "documented with reasons." Each step is its own record, and the second record contains the first record's fingerprint. The example is a loan because the records were built for an earlier demonstration. The mechanism is the same for a claim payment or an underwriting referral.

STEP 1 RECORD	`ser_a383778f88ae4f060279d8ba137926f8`
STEP 1 FINGERPRINT	`65723a7438b7c36686ed7658f8016b2dd5380eaa3500a40a10423fc6a5c9b99c`
STEP 1 PUBLIC LOG	Sigstore Rekor, log index 3247822463 · logged 2026-10-14 02:19:05 UTC · log value `7e0c76c0c9382c5237c3e4322e670623ff51e1b81f7feaab2eca9dc745a54104`
STEP 2 RECORD	`ser_5086c93c22af206a567fa06cf634682e`
STEP 2 FINGERPRINT	`192eb3dfe9fa65765ee508f0e29d00975cbc3370f81419ceef616179ef220ddf`
STEP 2 PUBLIC LOG	Sigstore Rekor, log index 3247822472 · logged 2026-10-14 02:19:06 UTC · log value `8e22075588e58e4111b7dc246a381f27762bfc3721e6fcaf81340ed72d7e6862`

How to check them

1. **In a web browser, with nothing to install.** Go to search.sigstore.dev, choose Log Index as the attribute, and enter 3247822463, then 3247822472. Each page shows the entry's hash, which matches the log value above, and the time the log received it, shown in your own time zone. Sigstore's Rekor is a public transparency log that Verdict does not operate.
2. **On any computer.** The log value is the SHA-256 fingerprint of the record's own fingerprint. Computing it yourself gives the same value, which shows the record's fingerprint existed no later than the log time.

   ```
   printf '%s' 65723a7438b7c36686ed7658f8016b2dd5380eaa3500a40a10423fc6a5c9b99c | shasum -a 256
   printf '%s' 192eb3dfe9fa65765ee508f0e29d00975cbc3370f81419ceef616179ef220ddf | shasum -a 256
   ```

3. **With the full records.** Both records are public. Verdict's open-source verifier recomputes each fingerprint from the record's content and checks it against the log, including the log's own proof that the entry is included, using a Rekor key taken from Sigstore's repository. A change to any byte of a record produces a different fingerprint that no longer matches.

   ```
   curl -sO https://verdict.systems/samples/ser_a383778f88ae4f060279d8ba137926f8.json
   curl -sO https://verdict.systems/samples/ser_5086c93c22af206a567fa06cf634682e.json
   git clone https://github.com/verdictsystems/ser-spec
   curl -sL -o rekor.pem \
     https://raw.githubusercontent.com/sigstore/root-signing/main/targets/rekor.pub
   python3 ser-spec/tools/verify.py ser_a383778f88ae4f060279d8ba137926f8.json \
     --anchor --rekor-key rekor.pem
   python3 ser-spec/tools/verify.py ser_5086c93c22af206a567fa06cf634682e.json \
     --anchor --rekor-key rekor.pem
   ```

   Each run ends with "RESULT: content self-consistent and anchor confirmed."

4. **Order.** The second record's content includes the first record's fingerprint (field `prior_merkle_root`). Because that value is part of the content that produced the second fingerprint, the override record was created after the recommendation it responds to, and neither record can be altered without breaking the match. The check prints True:

```
python3 -c "import json
a = json.load(open('ser_a383778f88ae4f060279d8ba137926f8.json'))
b = json.load(open('ser_5086c93c22af206a567fa06cf634682e.json'))
print(b['record']['metadata']['prior_merkle_root'] == a['merkle_root'])"
```

What these records show, and what they do not

STATEMENT	STATUS
The content of each record is the content that was logged	**Shown**
Each record existed no later than its log time	**Shown**
The override record was created after the recommendation record	**Shown**
The checks run on the examiner's own computer, against a log Verdict does not operate	**Shown**
The loan, the review and the override occurred	**Not shown (fictional)**
Every action in the workflow was recorded	**Not shown**
Who signed the records	**Not shown (sandbox key)**

The last three rows are why an examination would still sample and test.