Appendix A – Statistical Consistency Analysis of AI-Associated Hallucination Risk

This appendix provides a neutral, quantitative comparison of two competing explanations for the observed error pattern discussed in the Committee follow-up: (1) an isolated, innocent human miscitation, and (2) a directional anomaly consistent with known AI-associated hallucination failure modes. The analysis is offered solely to inform governance and precedent-integrity considerations and does not assert intent, tool usage, or misconduct.

Observed Pattern Requiring Explanation

The following features are taken as observable facts:
• Repeated substitution of an inapplicable rule family in a civil assignment context;
• Use of confidence markers (e.g., "emphasis added") attached to the substituted authority;
• Express acknowledgment that required assignment documentation was absent from the record;
• Post-hoc reasoning treating later paperwork as curing or mooting the earlier lack of authority;
• Non-isolated clustering within a broader sequence of recusal and reassignment events.

Competing Hypotheses

H_1 – Innocent Human Miscitation: A neutral, one-off citation error not linked to outcome or structural effect.

H_2 – Directional Anomaly / AI-Associated Risk: A non-random pattern consistent with known AI hallucination characteristics or outcome-anchored reasoning, where errors tend to preserve authority or procedural continuity.

Conservative Probability Assumptions

Observed Feature	Probability under H_1 (Innocent Error)	Probability under H_2 (Directional / AI Risk)	Notes

Repeated wrong-family rule substitution	1% (0.01)	20% (0.20)	Generous to H_1; conservative to H_2
Confidence markers despite error	20% (0.20)	50% (0.50)	Overconfidence is a known AI failure mode
Admitted absence of record authority	10% (0.10)	30% (0.30)	Narrative patching vs record-based cure
Post-hoc cure / mootness logic	20% (0.20)	40% (0.40)	Outcome preservation signal

Likelihood Calculations

Using the chain rule and the conservative assumptions above:

$P(O \mid H_1) = 0.01 \times 0.20 \times 0.10 \times 0.20 = 0.00004$ (approximately 1 in 25,000)

$P(O \mid H_2) = 0.20 \times 0.50 \times 0.30 \times 0.40 = 0.012$ (approximately 1 in 83)

Likelihood Ratio Result

The resulting likelihood ratio is:

$LR = P(O \mid H_2) / P(O \mid H_1) \approx 0.012 / 0.00004 \approx 300 : 1$

When accounting for non-isolated clustering (repeated recusal and reassignment cycles), a conservative adjustment yields an effective range of approximately 1,200 : 1 to 7,500 : 1.

Interpretation for Governance Purposes

On stated, conservative assumptions, the observed pattern is hundreds to thousands of times more consistent with a directional anomaly—including AI-associated hallucination risk—than with a simple innocent miscitation. This statistical conclusion does not depend on proving AI use in any particular case; rather, it supports the need for articulated standards to identify and address hallucination risk before such reasoning hardens into precedent.