From: Henri Van Ryn
Date: April 12, 2026 at 9:44:03 PM EDT
To: Daniel Capra
Subject: Supplemental Technical Submission: Rule 707 and the Statistical Horizon for Admissibility

Dear Professor Capra,

I am submitting the following technical brief for the Committee's consideration. I am currently finalizing the formal v3.6 technical report for publication and would be pleased to provide the full dataset and probability curves during our meeting in May.

Technical Brief: The Statistical Horizon & Rule 707 Admissibility
Author: Henri Frederic Van Ryn
Reference: Technical Report v3.6 (Feb 2026)

I. The Problem: The Reliability Deficit

Current forensic tools for AI text attribution suffer from a Signal-to-Noise (SNR) decay as text length decreases. Proposed Rule 707 currently fails to distinguish between "Machine Measurement" (deterministic) and "Machine Inference" (probabilistic). Without a quantitative floor, the court risks admitting "forensic noise" that violates Rule 702 reliability requirements.

II. The Mechanism: The n=180 Forensic Safety Floor

The Statistical Horizon is the point where machine-generated signals become mathematically distinguishable from the entropy of natural human language.

The Entropy Equation (H): We define the horizon where the entropy of the sample (Hs) exceeds the threshold for unique attribution: $H(X) = -\Sigma P(xi) \log2 P(xi)$

The Admissibility Probability (Pa): Admissibility is achieved only when the False Positive Rate (FPR) falls below the forensic significance level: $P(\text{False Positive}) = (1 - \text{Signal Strength})^n < 0.01$

Empirical Finding: In testing across multiple LLM architectures (GPT-4o, Claude 3.5, Gemini 1.5), the 180-token threshold represents the "Physical Limit" of reliable attribution. Below 180 tokens, the probability of "collisions" with natural human text increases exponentially, rendering the evidence stochastically opaque.

III. Proposed Regulatory Framework

To resolve the overbreadth of Rule 707, I propose a Bright-Line Standard:
The Deterministic Zone (< 180 tokens): Machine outputs are treated as simple tools; probabilistic inference of authorship is deemed per se inadmissible for lack of reliability.

The Inferential Zone (>= 180 tokens): Machine outputs trigger Rule 707 scrutiny, allowing for the cross-examination of the underlying "analytical reasoning" of the model.

The full v3.6 report includes the complete datasets and the Admissibility Probability Curves that define this 180-token floor. I will be in New York from May 5th through the 14th and am available for a brief technical briefing or to answer any questions the Committee may have ahead of the May 7th vote.

Respectfully,

Henri Frederic Van Ryn
Independent Researcher

From: Henri Van Ryn
Date: May 6, 2026 at 10:11:05 PM EDT
To: Daniel Capra
Subject: Statistical Horizon and Proposed Rule 707 — Updated Framework

Professor Capra,

I previously submitted a technical brief concerning AI-generated text attribution and Proposed Rule 707. Since that submission, I have refined the original fixed-threshold proposal into a more general scaling framework: the minimum text sample needed for reliable attribution depends on watermark strength, detector response, and the evidentiary confidence required.

I am sending this short update directly in case it is useful to you or the reporters as the Committee continues its work on machine-generated evidence. The central point is simple: before courts, regulators, platforms, or schools rely on AI-text detection, they need to know where detection becomes statistically meaningful.

I would be glad to provide a short practitioner/research perspective on reliability, attribution thresholds, or admissibility standards if that would be useful.

Respectfully,

Henri Frederic Van Ryn
Independent Researcher