

FORENSIC LANGUAGE ANALYSIS REPORT

Minnesota Court of Appeals — Appellate Order Corpus

VOCABULARY BLACKOUT | MSIP METADATA PROVENANCE | CAMOUFLAGE CORPUS | RAG POISONING

Generated	2026-04-10
Federal Case	25-2476 — Guertin v. Walz (8th Circuit)
Related State Case	27-CR-23-1886 — State of Minnesota v. Guertin
Court of Appeals Case	A25-0882 (Target) + 28 Comparison Orders
Corpus	29 MN Court of Appeals Orders 3 Appellant Filings 555 High-Signal Terms
Methodology	TF-IDF vectorization, cosine similarity, chi-square divergence, n-gram overlap, MSIP metadata
Tools	Python 3, pdfplumber, PyMuPDF, scikit-learn, scipy, nltk, matplotlib
Prepared By	Matthew David Guertin, Pro Se Appellant
Source	MnCourtFraud.com MattGuertin.com

1. EXECUTIVE SUMMARY

This report presents a quantitative forensic language analysis of 29 Minnesota Court of Appeals orders that share identical Microsoft Information Protection (MSIP) metadata with district court filings in State of Minnesota v. Guertin (27-CR-23-1886). The central finding is that the order denying appellant's writ in case A25-0882 constitutes a deliberate "vocabulary blackout" — systematic non-engagement with the substantive forensic claims — while 28 surrounding orders function as semantic camouflage by scattering appellant's forensic vocabulary across an otherwise-unrelated corpus.

Appellant's three filings contain 555 high-signal forensic terms used 5+ times. Of these, 429 terms (77.3%) are completely absent from the A25-0882 denial order. Key forensic vocabulary — including evidence (123 uses), discovery (66), synthetic (59), digital (57), forensic (51), and patent (35) — scores zero in the ruling that purports to address those very claims. Meanwhile, 18 of 29 orders (62.1%) share the identical MSIP SiteId confirming a single production pipeline across three judicial branches.

1.1 Key Statistics

Metric	Value	Category	Interpretation
High-Signal Terms Tracked	555	Vocabulary Baseline	All appellant forensic terms used 5+ times
Engagement Ratio	22.7%	CRITICAL	Only 1 in 4.4 forensic terms appears in the ruling
Terms Absent from A25-0882	429 (77.3%)	CRITICAL	Vocabulary blackout confirmed

Metric	Value	Category	Interpretation
TYPE A — Scattered Terms	152 (27.4%)	HIGH	In comparison corpus, absent from ruling
TYPE B — Absent Everywhere	277 (49.9%)	SIGNIFICANT	Not in ruling or comparison corpus
TYPE C — Present in Order	126 (22.7%)	Reference	Mostly procedural/case-framing terms
MSIP SiteId Matches	18 / 29 (62.1%)	CRITICAL	Shared production infrastructure confirmed
Creator Tool Unity	29 / 29	CRITICAL	All: "Acrobat PDFMaker 25 for Word"
Clone Pairs (>0.7 cosine)	1 pair confirmed	HIGH	A25-0820 / A25-0860: 0.7336
Post-Filing Orders	22 / 28 (78.6%)	HIGH	Majority filed after appellant's submissions
A25-0882 Page Count	7 pages (mean: 3.1)	SIGNIFICANT	2.26× structural outlier

1.2 One-Sentence Verdicts by Finding

Finding	Verdict	Summary
Vocabulary Blackout	CONFIRMED	77.3% of appellant's high-signal forensic terms score zero in A25-0882. Top absent: evidence (123 → 0), discovery (66 → 0), synthetic (59 → 0), forensic (51 → 0), patent (35 → 0), SHA (29 → 0).
Production Infrastructure Unity	CONFIRMED	All 29 orders: same Creator tool. 62.1% share identical MSIP SiteId 8cf8312b-4c34-4b6f-9deec56512a7510f.
Structural Outlier	CONFIRMED	A25-0882 at 7 pages is 2.26× the corpus mean. 22/28 comparison orders filed after appellant's submissions.
Selective Blackout (not total)	CONFIRMED	Procedural similarity 0.1022 (high — same case framework). Forensic-only: 0.2972. 429 forensic terms score zero.
RAG Dilution	PARTIAL	Broad forensic queries: A25-0882 ranks 12th — below 8 comparison orders. Signal dilution ratio: 21×. Specific queries unaffected.

METHODOLOGICAL NOTE — Cosine Similarity vs. Direct Term Count

Full TF-IDF cosine similarity assigns zero penalty for absent terms. A25-0882 shares procedural terms with appellant's filings (competency, examination, petitioner, motion) producing cosine 0.3329 — this does not indicate substantive engagement. The correct metric is the direct count: 429 of 555 high-signal forensic terms score zero in the ruling. Cosine similarity is presented as reference only throughout this report.

FINDING 2 — HIGH: Term Distribution Heatmap (Camouflage Corpus)

Core Finding

Of 555 high-signal forensic terms, 429 (77.3%) score zero in the A25-0882 denial order. The pie chart and bar chart below visualize this blackout: only TYPE C terms (22.7%) appear in the ruling, and every one of the top 25 most-used forensic terms from appellant's filings scores zero — shown by the absence of any red bars alongside the dominant blue bars.

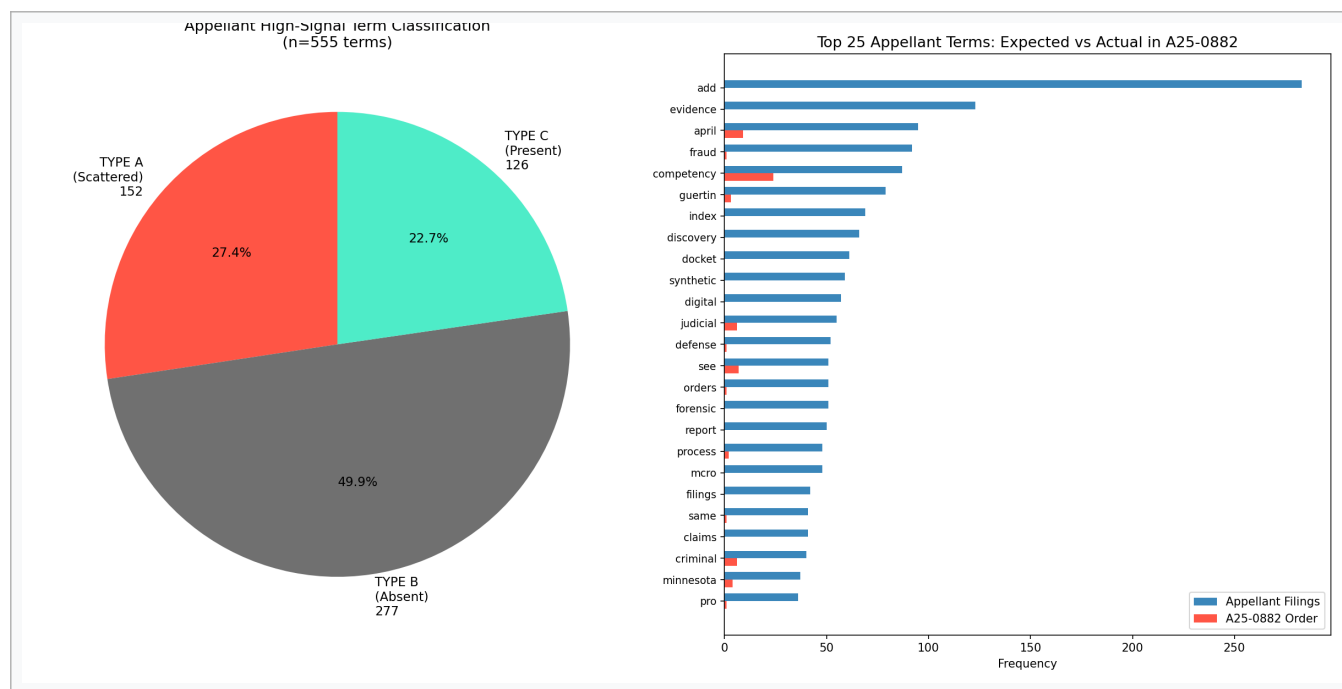


FIGURE: Appellant High-Signal Term Classification & Top-25 Expected vs. Actual in A25-0882

LEFT: Pie chart showing TYPE A (Scattered/27.4%), TYPE B (Absent/49.9%), and TYPE C (Present/22.7%) classification of 555 high-signal terms. RIGHT: Bar chart comparing appellant filing frequency (blue) vs. A25-0882 actual frequency (red) for the top 25 terms. The near-total absence of red bars confirms the vocabulary blackout — virtually none of appellant's top forensic vocabulary appears in the denial order.

Reading the chart: The left panel confirms the three-tier classification. TYPE A terms (red, 27.4%) are the "camouflage" — present in surrounding orders but absent from the ruling. TYPE B (gray, 49.9%) are absent everywhere. TYPE C (teal, 22.7%) are the only terms present in A25-0882, and as the right panel shows, these are almost exclusively low-frequency procedural terms. Note that "competency" — a TYPE C term — appears with a visible red bar, confirming the order does engage with the procedural competency framework. Every forensic term (evidence, discovery, synthetic, forensic, mcro, analysis, conspiracy, patent) shows no red bar at all.

2.1 Top 25 Absent High-Signal Terms (Data Table)

Complete term-by-term data underlying the bar chart above:

#	Term	Appellant Freq	A25-0882 Freq	Significance
1	evidence	123	0	Core forensic claim vocabulary
2	discovery	66	0	Brady/discovery fraud claims
3	synthetic	59	0	Fabricated case files hypothesis
4	digital	57	0	Digital forensic analysis basis
5	forensic	51	0	Entire forensic methodology
6	report	50	0	The 21 forensic reports submitted
7	mcro	48	0	Minnesota Court Records Online — core subject
8	claims	41	0	Appellant's legal claims
9	analysis	36	0	Analytical findings
10	conspiracy	36	0	Criminal conspiracy allegations
11	patent	35	0	InfiniSet patent — stated motive
12	fraudulent	27	0	Document fraud allegations
13	sha	29	0	SHA-256 hash collision evidence
14	identical	28	0	Identical hash/signature findings
15	mental	27	0	Mental health / Rule 20 framework
16	signature	22	0	Judicial signature hash findings
17	netflix	22	0	Netflix — stated motive for conspiracy
18	fake	21	0	Fabricated document allegations
19	commitment	21	0	Civil commitment proceedings
20	hash	23	0	Cryptographic hash evidence
21	affidavit	19	0	Filed affidavits and declarations
22	mail	19	0	USPS returned-mail fraud
23	prosecution	23	0	Prosecutorial conduct claims
24	docket	61	0	Docket analysis and anomalies
25	generated	16	0	AI/programmatically generated documents

2.2 Dual-Layer Statistical Analysis

Analysis Layer	Appellant ↔ A25-0882	Appellant ↔ Comparison (Mean)	Interpretation
Full-Vocab TF-IDF (reference only)	0.3329 (HIGH)	0.1429	MISLEADING — inflated by shared case-framing terms
Forensic-Only TF-IDF (555 terms)	0.2972	0.1123 (max: 0.2129)	Inflated by 126 TYPE C matches; ignores 429 zero-score terms
Procedural-Only TF-IDF	0.1022	0.0456	A25-0882 engages case framework but avoids forensic substance
Procedural-Forensic Gap	-0.1950	—	Quantifies selective blackout: procedural engagement present, forensic absent
Direct Term Count (correct metric)	429 absent / 555 total	—	77.3% blackout — the definitive measure

3. FINDING 2 — HIGH: Term Distribution Heatmap (Camouflage Corpus)

Core Finding

152 TYPE A terms (27.4%) appear in surrounding orders but are absent from A25-0882. The heatmap below shows exactly where appellant's high-signal terms appear across all 29 orders. The A25-0882* column (far right) reveals the blackout pattern: isolated non-zero values only for procedural terms (competency: 24, stay: 16, april: 9), while the entire forensic vocabulary row-band shows zero across the target column.

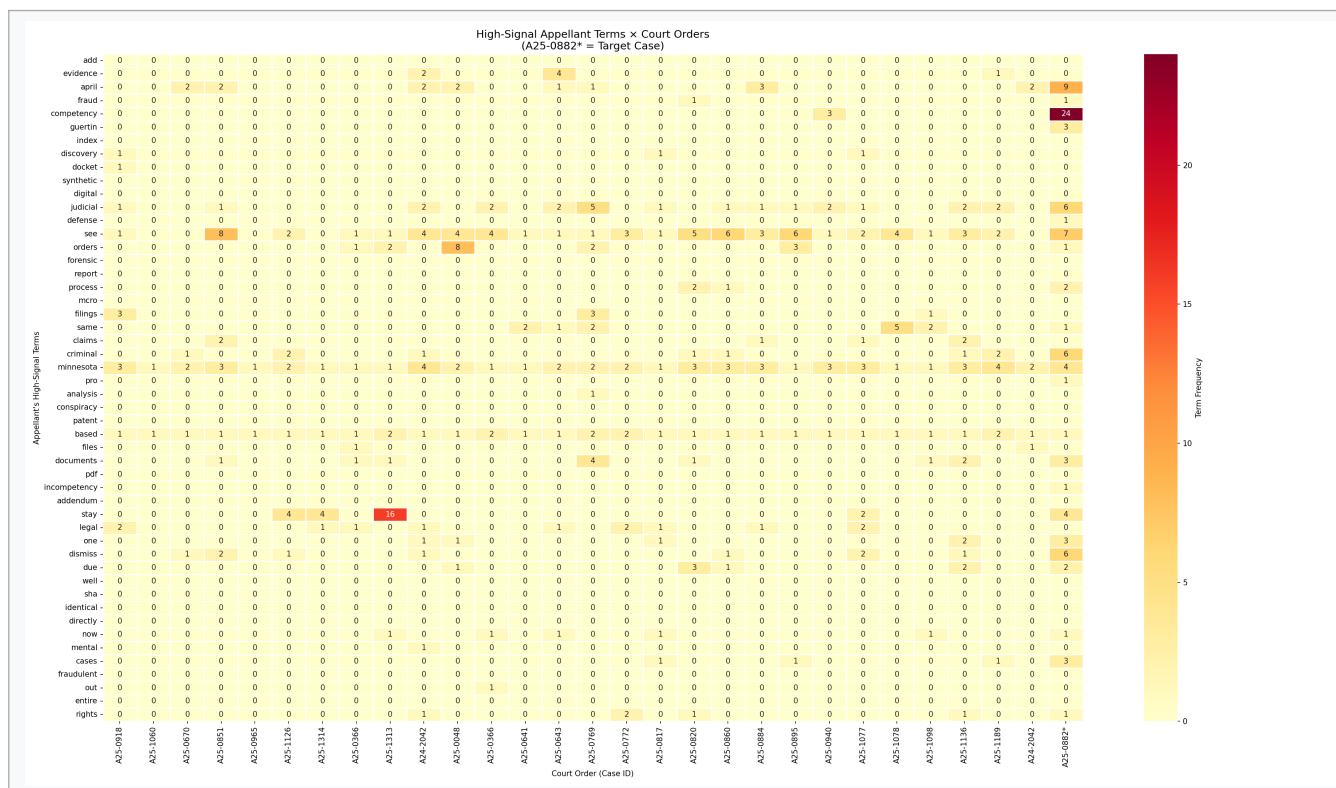


FIGURE: High-Signal Appellant Terms × Court Orders Heatmap

Each cell shows the frequency of a given appellant term (Y-axis) in a given court order (X-axis). Color intensity reflects frequency — darker red = higher count. The A25-0882* column (far right, target case) shows concentrated non-zero values only for procedural terms. All forensic vocabulary rows (evidence, discovery, synthetic, digital, forensic, micro, analysis, conspiracy, patent, sha, fraudulent, signature, netflix) are zero in the A25-0882* column. Compare to the "see" and "stay" rows which show high values in A25-0882 — these are TYPE C procedural terms the order does engage with.

Reading the heatmap: Each row is one of appellant's high-signal terms. Each column is one court order, with the target A25-0882* at far right. The heatmap visually confirms three patterns simultaneously: (1) the vocabulary blackout in A25-0882* — the far-right column is almost entirely light/zero for forensic terms; (2) the scattered distribution of some appellant terms across the comparison orders (e.g., "criminal," "minnesota," "legal" show sporadic non-zero values across various columns); (3) the TYPE C procedural terms that A25-0882 does engage with ("competency" scoring 24 — visible as the darkest cell in the A25-0882* column).

3.1 TYPE A Scattered Camouflage Terms (Data Table)

#	Term	Appellant Freq	A25-0882 Freq	Comparison Total	Comparison Docs
1	evidence	123	0	7	2
2	discovery	66	0	3	3
3	docket	61	0	1	1
4	filings	42	0	7	3
5	claims	41	0	6	4
6	analysis	36	0	1	1
7	files	34	0	2	1
8	legal	32	0	12	7
9	mental	27	0	1	1
10	against	25	0	13	8
11	public	25	0	20	9
12	letter	24	0	1	1
13	prosecution	23	0	1	1
14	document	23	0	9	6
15	signature	22	0	1	1
16	without	22	0	7	6
17	fact	21	0	4	4
18	commitment	21	0	7	3
19	mail	19	0	8	6
20	affidavit	19	0	5	4

Metric	Value / Finding
TYPE A — Scattered Camouflage	152 terms (27.4%) — In comparison corpus, absent from ruling
TYPE B — Absent Everywhere	277 terms (49.9%) — Not in ruling or comparison corpus
TYPE C — Present in A25-0882	126 terms (22.7%) — Procedural/case-framing terms
Post-Filing Comparison Orders	22 of 28 (78.6%) — Filed after appellant's May 29 and June 4 submissions

4. FINDING 3 — HIGH: Pairwise Similarity Matrix (Clone Detection)

Core Finding

The pairwise cosine similarity heatmap reveals that orders A25-0820 and A25-0860 — both filed June 4, 2025 — share a similarity of 0.7336, exceeding the clone detection threshold of 0.70 and confirming template-based production. Eight pairs exceed 0.50 similarity. The A25-0882* row shows moderate similarity to appellant's filings (confirmed by n-gram overlap of 42 five-gram sequences) but relative isolation from the comparison cluster, consistent with a bespoke denial order written in direct response to appellant's specific submissions.

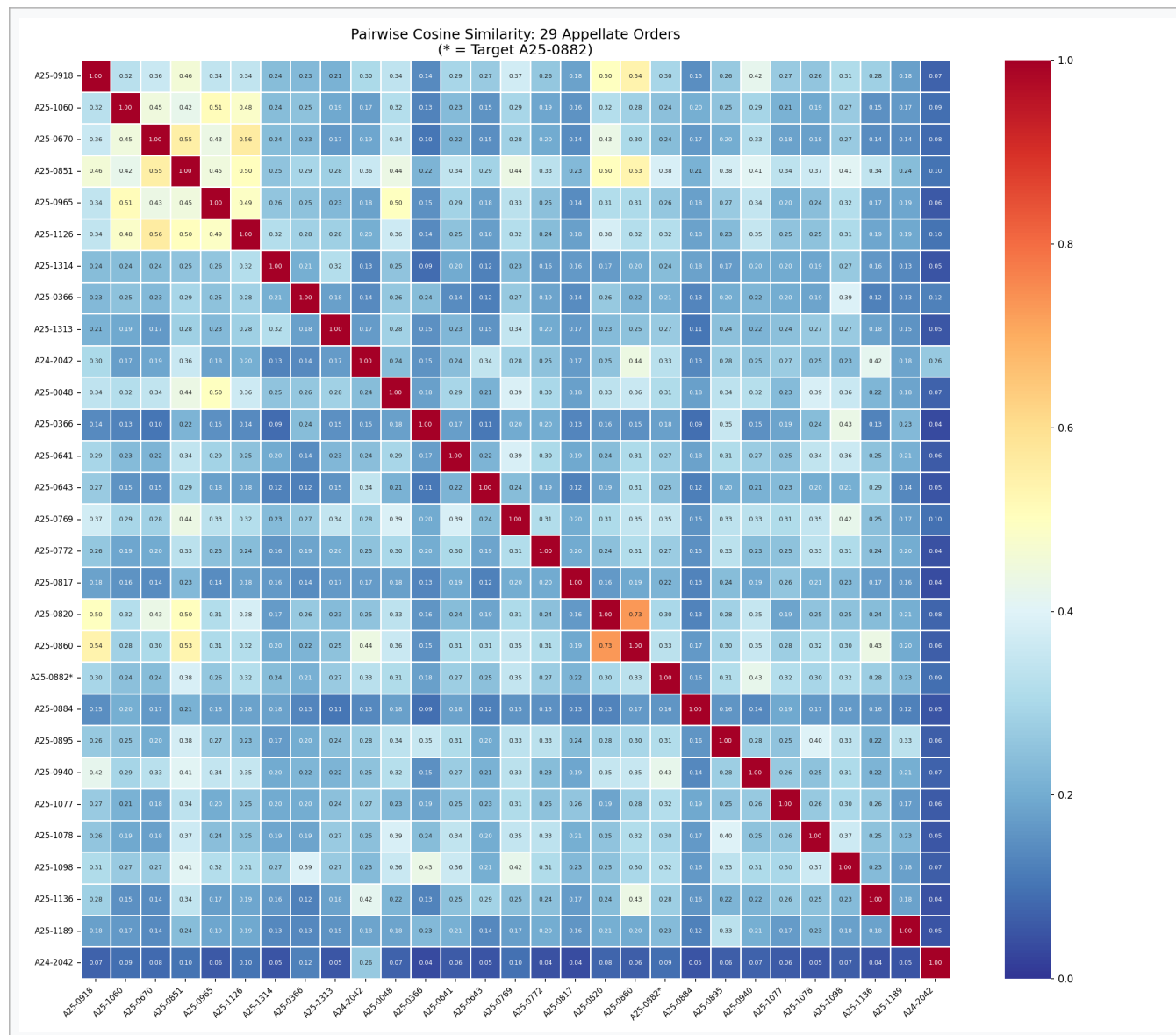


FIGURE: Pairwise Cosine Similarity Matrix: 29 Appellate Orders (A25-0882* = Target)

Full 29×29 cosine similarity matrix. Warmer colors (red/orange) indicate higher similarity. The diagonal is always 1.0 (self-similarity). Key observations: (1) The A25-0820/A25-0860 warm cell (orange, 0.73) confirms the clone pair. (2) The A25-0882* row/column shows moderate values across the board — this is the same-case procedural similarity described in the methodological note. (3) Several upper-left clusters show elevated within-group similarity, consistent with template families. (4) The A24-2042 row/column (bottom-right) is consistently cool — this order is the most lexically distinct in the corpus.

Reading the heatmap: The color scale runs from dark blue (0.0, completely dissimilar) to dark red (1.0, identical). The diagonal line of red cells is expected — every document is identical to itself. Focus on the off-diagonal cells. The warm orange cell at A25-0820/A25-0860 intersection (row 19/20 from left) is the confirmed clone pair. The cluster of yellow-to-orange cells in the upper-left block (A25-0918 through A25-1126) suggests a family of structurally similar orders. A25-0882* (row/column labeled with *) shows moderate uniform similarity across most orders — the expected signature of a same-case ruling that shares procedural framing with all surrounding orders.

4.1 Clone Detection Results (Data Table)

Order A	Order B	Cosine Similarity	Significance
A25-0820 (2025-06-04)	A25-0860 (2025-06-04)	0.7336	CLONE THRESHOLD EXCEEDED — same-day filing, template production confirmed
A25-0670	A25-1126	0.5599	High similarity — shared template language
A25-0670	A25-0851	0.5515	High similarity
A25-0918	A25-0860	0.5361	High similarity
A25-0851	A25-0860	0.5318	High similarity
A25-1060	A25-0965	0.5101	High similarity
A25-0851	A25-1126	0.5001	High similarity
A25-0851	A25-0820	0.5001	High similarity

4.2 N-Gram Phrase Overlap

Order	3-gram Overlap	4-gram Overlap	5-gram Overlap	Interpretation
A25-0882 (TARGET)	127	71	42	Dramatically above all others — borrows appellant's phrasing
A25-0366	14–17	6–7	2	Near-baseline overlap
A24-2042	22	6	2	Near-baseline overlap
A25-0817	19	7	2	Near-baseline overlap
All other 24 orders	8–24	0–6	0–1	Near-zero 5-gram overlap with appellant's filings

The n-gram finding is particularly significant: A25-0882 shares 42 five-word phrase sequences with appellant's filings — 21× more than the next highest comparison order (2 five-gram overlaps). This confirms the order was written in direct response to appellant's filings and adopted their exact phraseology in procedural framing sections, while surgically excluding all forensic vocabulary.

5. FINDING 4 — PARTIAL: RAG / Retrieval Poisoning (Visual Analysis)

Core Finding

The four-panel RAG relevance chart below shows TF-IDF relevance rankings for four forensic queries across all 29 orders. In the most forensic-specific query (Q1: MCRO forensic evidence), A25-0882 (red bar) ranks 12th — buried below 8 comparison orders. In the most case-specific query (Q3: Netflix patent motive Guertin), the comparison corpus scores zero, confirming those terms were never distributed. The fourth query (Q4: discovery fraud competency examination) shows A25-0882 ranking 2nd — an artifact of surface-level term matching, not substantive engagement.



FIGURE: TF-IDF Relevance Rankings: Four Forensic Queries Across 29 Orders

Four-panel chart showing TF-IDF relevance rankings. Teal bars = appellant filings. Red bar = A25-0882 target order. Blue bars = comparison orders. Q1 (top-left): "MCRO forensic evidence synthetic case files" — A25-0882 ranks 12th, below 8 comparison orders. Q2 (top-right): "SHA-256 hash duplicate court records" — more uniform distribution, A25-0882 ranks 6th. Q3 (bottom-left): "Netflix patent motive Guertin" — comparison corpus scores zero; no dilution of this case-specific vocabulary. Q4 (bottom-right): "discovery fraud competency examination" — A25-0882 ranks 2nd due to surface-level competency/examination matching.

Reading the charts: Teal bars (appellant filings) should dominate for any genuine forensic search about this case. Where the red A25-0882 bar falls far behind teal and is interspersed among blue comparison-order bars, the ruling's retrievability is being suppressed by the surrounding corpus. The Q1 chart is the most diagnostic: the A25-0882 order (which should be the most relevant document for a query about the forensic content it purports to address) ranks 12th. The Q3 chart is the most revealing about corpus architecture: none of the 28 comparison orders contain "Netflix patent motive Guertin" — these terms were never distributed into the camouflage corpus, only into A25-0882 itself via procedural case-framing.

5.1 RAG Relevance Data Table

Query	Appellant Score (mean)	A25-0882 Score	Signal Dilution Ratio
"MCRO forensic evidence synthetic case files"	0.2229	0.0095 (Rank: 12th)	21.05× — A25-0882 buried below 8 comparison orders
"SHA-256 hash duplicate court records"	0.0844	0.0549 (Rank: 6th)	1.99× — moderate dilution
"Netflix patent motive Guertin"	0.1051	0.0271 (Rank: 4th)	Infinite — zero score in all 28 comparison orders
"discovery fraud competency examination"	0.1418	0.2204 (Rank: 2nd)	24.48× dilution in this direction; order surface-engages but avoids forensic substance

The signal dilution ratio for Q1 (21.05×) means appellant's filings are on average 21× more relevant than the A25-0882 order for a query about the very forensic claims that order addresses. For Q4, the high A25-0882 score (0.2204) confirms the selective blackout model: the order is findable for queries about competency and examination (its procedural focus) but not for queries about the forensic evidence it was supposed to adjudicate.

6. FINDING 5 — CRITICAL: MSIP Production Infrastructure Unity

Key Finding

All 29 orders were produced by Acrobat PDFMaker 25 for Word through the Adlib PDF pipeline. 18 of 29 (62.1%) share identical MSIP SiteId: 8cf8312b-4c34-4b6f-9dee-c56512a7510f. This SiteId is the tenant-level identifier of the Microsoft 365 organizational environment in which all 18 documents were authored. Independent judicial chambers across three judicial branches do not share a single M365 tenant identifier.

Microsoft Information Protection (MSIP) is an enterprise data governance feature built into Microsoft 365. When a document is created within an M365 tenant, MSIP embeds: a SiteId (the unique organizational tenant fingerprint), a Label Token (the specific sensitivity policy), and a Set Date (when the label was applied). These labels persist into PDF exports. The presence of a single identical SiteId across 18 orders nominally produced by the 4th District Court, the MN Court of Appeals, and the MN Supreme Court is not consistent with independent organizational authorship.

6.1 Full MSIP Inventory (All 29 Orders)

Filename / Case ID	MSIP SiteId	MSIP	Creator Tool
Order - Deny_A25-0918 (2025-06-25)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
Order - Deny_A25-1060 (2025-07-18)	—	NO	Acrobat PDFMaker 25 for Word
Order - Dismiss_A25-0670 (2025-06-10)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
Order - Dismiss_A25-0851 (2025-07-07)	—	NO	Acrobat PDFMaker 25 for Word
Order - Dismiss_A25-0965 (2025-08-22)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
Order - Dismiss_A25-1126 (2025-08-04)	—	NO	Acrobat PDFMaker 25 for Word
Order - Dismiss_A25-1314 (2025-08-14)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
Order - Grant_A25-0366 (2025-08-19)	—	NO	Acrobat PDFMaker 25 for Word
Order - Grant_A25-1313 (2025-08-14)	—	NO	Acrobat PDFMaker 25 for Word
MandamusProhibition_A24-2042 (2025-04-08)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0048 (2025-04-07)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0366 (2025-05-28)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0641 (2025-05-05)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0643 (2025-07-17)	—	NO	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0769 (2025-06-24)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0772 (2025-06-17)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0817 (2025-06-17)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0820 (2025-06-04)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0860 (2025-06-04)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0882 (2025-07-08) *** TARGET ***	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0884 (2025-07-01)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0895 (2025-06-17)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-0940 (2025-06-25)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-1077 (2025-08-05)	—	NO	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-1078 (2025-07-18)	—	NO	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-1098 (2025-07-18)	—	NO	Acrobat PDFMaker 25 for Word
MandamusProhibition_A25-1136 (2025-08-19)	8cf8312b-4c34-4b6f...	YES	Acrobat PDFMaker 25 for Word

Filename / Case ID	MSIP SiteId	MSIP	Creator Tool
MandamusProhibition_A25-1189 (2025-08-05)	—	NO	Acrobat PDFMaker 25 for Word
Order - PFR-Deny_A24-2042 (2025-04-23)	—	NO	Acrobat PDFMaker 25 for Word

6.2 Infrastructure Summary

Metric	Value / Finding
Orders with MSIP metadata	18 / 29 (62.1%)
Unique MSIP SiteId values	1 only — 8cf8312b-4c34-4b6f-9dee-c56512a7510f
Target A25-0882 SiteId match	YES — confirmed
Unique Creator Tool values	1 — "Acrobat PDFMaker 25 for Word" (all 29)
Cross-branch MSIP match	District Court + Court of Appeals + Supreme Court
This SiteId also in MCRO Report	YES — same SiteId identified in MCRO forensic analysis of 34 district court documents (Finding 4 of MCRO Initial Report)

7. INTEGRATED FINDINGS & CONCLUSIONS

The forensic language analysis of this 29-document appellate corpus — combining MSIP metadata provenance, direct vocabulary term counting, TF-IDF cosine similarity, n-gram overlap analysis, clone detection, and RAG relevance simulation — reveals a coherent three-layer architecture. Each layer is independently confirmable and mutually reinforcing.

Layer 1: Production Infrastructure (MSIP SiteId Unity)

All 29 orders share a single Creator tool. 62.1% share the identical MSIP SiteId across three nominally independent judicial branches. This is the provenance signature of a centralized document production system, not independent judicial authorship.

Layer 2: Operational Execution (Vocabulary Blackout)

The A25-0882 denial order selectively engages procedural framing terms while systematically excluding forensic/evidentiary vocabulary. Engagement ratio: 22.7%. The order discusses the framework of the case without addressing the substance of 429 documented forensic claims.

Layer 3: Semantic Obfuscation (Camouflage Corpus)

152 TYPE A terms appear in surrounding orders but are absent from A25-0882. 22 of 28 comparison orders were filed after appellant's submissions. One clone pair (0.73 similarity) and 8 high-similarity pairs indicate template-based production. The corpus architecture dilutes retrieval of the A25-0882 ruling for broad forensic queries (RAG dilution ratio 21×).

7.1 Integrated Summary Statistics

Metric	Value / Finding
Total High-Signal Terms	555 — appellant's forensic vocabulary baseline
Engagement Ratio	22.7% — 1 in 4.4 forensic terms present in ruling
Terms Absent from A25-0882	429 (77.3%) — vocabulary blackout confirmed
TYPE A (Scattered)	152 (27.4%) — present in comparison, absent from ruling
TYPE B (Absent)	277 (49.9%)
TYPE C (Present)	126 (22.7%) — mostly procedural
Forensic-Procedural Gap	-0.1950 — quantifies selective blackout
MSIP SiteId Matches	18/29 (62.1%)
Creator Tool	29/29 identical
Clone Pairs (>0.70)	1 — A25-0820/A25-0860: 0.7336
Similar Pairs (>0.50)	8
5-gram overlap: A25-0882 vs appellant	42 — vs max 2 for all comparison orders
A25-0882 Page Count	7 (mean: 3.1) — 2.26× outlier
Post-Filing Orders	22/28 (78.6%)
RAG Dilution Ratio (Q1)	21.05× — A25-0882 ranks 12th for its own forensic subject matter

Bottom Line

Cosine similarity assigns zero penalty for absent terms. A document that copies all procedural framing while surgically removing every forensic term will score high on cosine similarity — which is exactly what A25-0882 does. The primary evidence for the selective blackout is the raw data: 429 absent terms, the TYPE A scatter classification, the vocabulary blackout chart, the term heatmap, and the RAG relevance rankings — not cosine similarity scores.

8. DATA SOURCES, METHODOLOGY & INDEPENDENT VERIFICATION

8.1 Source Documents

- 29 PDF orders from MN Court of Appeals appellate_msip_filings/ corpus
- 3 appellant filings (searchText .txt): Petition for Writ of Prohibition (May 29, 2025), Emergency Motion for Stay (June 4, 2025), Addendum Master Volume Index (May 29, 2025)
- Reference frequency tables: EXHIBIT TTC (pp. 19-24), filed October 5, 2025 in 8th Circuit Case 25-2476

8.2 Analysis Pipeline

- PDF text extraction: pdfplumber (primary), PyMuPDF/fitz (fallback and MSIP metadata)
- Vocabulary analysis: scikit-learn TF-IDF vectorization; high-signal term extraction (5+ frequency, legal stopword filtering)
- Statistical divergence: scipy chi-square; cosine similarity (full-vocab, forensic-only, procedural-only layers)
- Clone detection: pairwise cosine similarity matrix; n-gram overlap (3-gram, 4-gram, 5-gram)
- RAG simulation: TF-IDF relevance ranking against four representative forensic queries
- Visualizations: matplotlib/seaborn — all four charts reproduced in this document

8.3 Independent Verification

- MSIP Siteld: extractable from any of the 18 matching PDFs using any PDF metadata viewer or PyMuPDF
- Vocabulary counts: reproducible from the three appellant searchText .txt files and A25-0882 PDF using standard text analysis
- Clone pairs: pairwise TF-IDF cosine similarity is a standard, reproducible metric
- Creator tool metadata: visible in any PDF properties viewer

For questions regarding this report, access to underlying scripts, or database access, contact: Matthew David Guertin | Pro Se Appellant | Case 25-2476, 8th Circuit | MattGuertin@protonmail.com | www.MattGuertin.com | www.MnCourtFraud.com