

MCRO | Forensic Analysis Report — Internet Archive / photorobot.com

Subtitle: Internet Archive (Wayback Machine) Captures of photorobot.com — Preserved December 2022 – January 2023

Category tag line: JAVASCRIPT BUNDLE TIMESTAMP ANALYSIS | CONTENT EVOLUTION FORENSICS | PUBLICATION DATE DISCREPANCIES | CORPUS VOCABULARY SEPARATION | OBSERVER CSV INDEPENDENT REPRODUCTION

For legal / investigative use.

Report Metadata

Field	Value
Prepared by	Matthew David Guertin, Pro Se Defendant
Case Reference	Related to Case 27-CR-23-1886 (technical-artifact examination; not a court filing)
Court	Independent technical examination (not produced for or under direction of any court)
Report Date	November 2026
Data Source	Internet Archive (Wayback Machine) public captures of photorobot.com
Observation Window	December 9, 2022 – mid-January 2023 (preservation window); examination window November 2026
Collection Method	Wayback Machine “Save Page Now” via a custom TouchDesigner patch automating multi-page snapshot requests with VPN rotation; full HTML + asset bundle downloads; PDF screen captures of the Wayback PDF-viewer UI during the observation window; Google Docs “compare” exports (.docx with tracked changes) of adjacent capture pairs
Corpus Size	77,603 files, 9.27 GB — 62 page slugs × 1,979 Wayback HTML captures, 35 dated section captures, 41,457 asset

Field	Value
	files (JS/CSS/image/svg bundles), 38 Google Docs compare .docx files, 1 product brochure PDF
Primary Artifact Type	Webflow-built static marketing site (HTML + content-addressed JavaScript/CSS bundles) as served by Internet Archive playback
Examination Environment	File-system corpus under <code>_workspace/.../forensic_investigation/</code> ; Python analysis scripts under <code>_workspace/.../forensic_analysis/{v1,v2,v3}/</code> ; no live Internet Archive queries
Reproducibility	SHA-256 integrity baseline across all 77,603 preserved files (<code>forensic_analysis/baseline/sha256_manifest.csv</code>). Every quantitative claim in this report regenerates from unmodified <code>forensic_investigation/</code> via scripts in the <code>v1/v2/v3</code> analysis packages. 99.0% independent row-level reproduction of the collector's own thirteen contemporaneous CSVs (see Finding 1).

IMPORTANT DISCLAIMER

This report presents objective forensic data findings reproducible from preserved source artifacts.

It does not assert conclusions about fraud, intent, motive, or criminal liability. Each finding is a documented pattern or internal contradiction that warrants independent expert examination and, where noted, a small number of live Internet Archive CDX queries that are outside the scope of this static-artifact examination.

The examiner in this report is independent in posture — the preserved dataset was examined against the collector's documented observations as a validation target, not as an input, and the report explicitly declines to adjudicate questions that the static evidence alone cannot close.

Executive Summary

This report documents a forensic examination of preserved Internet Archive (Wayback Machine) captures of the domain **photorobot.com** collected between December 9, 2022 and mid-January 2023. The site was examined because its published product — a rotating-platform photography system called “Virtual Catwalk” — was cited during the observation window as candidate prior art adjacent to subject-matter claimed in US Patent 11,577,177 (filed by the collector March 18, 2022; granted February 14, 2023).

The collector recorded real-time observations of content evolution and archival anomalies during the preservation window. This examination independently evaluates those observations against the static, preserved dataset.

Five findings are documented:

- **Finding 1 (CRITICAL):** Independent reproduction of the collector’s thirteen contemporaneous observation CSVs matches at **99.0% row-level agreement** (1,099 of 1,110 rows, both timestamps within tolerance).
- **Finding 2 (CRITICAL):** The Internet Archive “footer-archived” timestamps on the JavaScript bundles served inside captures of photorobot.com — spanning **14 distinct Webflow site builds across 2 years of site evolution** — concentrate into a **22-day window in December 2022**, including a **single peak day carrying 30.4% of all records** and a **sharp sign-flip pattern on December 9, 2022** coincident with a collector-documented live-edit event at **04:46 CST**.
- **Finding 3 (CRITICAL):** **14 PhotoRobot blog articles carry claimed publication dates between 2021-03-19 and 2021-08-09, with no surviving evidence of 2021 existence in the preserved dataset.** Their earliest observed existence is **2022-08-11** — discrepancies from **367 to 510 days** between claim and first evidence. Two additional articles demonstrate the claimed-publication-date field is not authoritative: they are present in a March 2021 snapshot while claiming July 2021 publication.
- **Finding 4 (HIGH):** Tracked-change edits recorded by the collector (824 insertions and deletions across 38 Google Docs compare files) include **three specific phrases drawn from US Patent 11,577,177’s novelty vocabulary** — infinite runway, filming product videos, producing online fashion shows — inserted on the Virtual Catwalk product page and the Turning Platform product page **with zero offsetting deletions anywhere in the 38 compare documents.** Concurrent with these insertions, the Virtual Catwalk page grew **+22% in raw HTML size in a 22-day window** (2022-12-06 → 2022-12-28) overlapping the Finding 2 cluster.
- **Finding 5 (HIGH):** The collector’s own qualitative MAXQDA coding of the PhotoRobot blog corpus (140 codes, December 2022) **reproduces quantitatively under independent analysis.** KMeans clustering (k=2)

partitions the 62-page corpus into a 52-slug “older” cluster and a 10-slug “newer” cluster along a temporally-separable boundary. Eight codes — including scanning, elevation, real-time, axis, robotic camera, robotic camera arm, rotation, rotary — separate the two clusters with $\Delta \geq 25$ percentage points. **22 of 140 MAXQDA codes overlap with US Patent 11,577,177’s novelty vocabulary.**

Corpus Overview

Metric	Count	Notes
Total preserved files	77,603	All SHA-256 hashed at examination baseline
Total preserved size	9.27 GB	Static file-system corpus
Unique Wayback HTML captures	1,979	Across 62 page slugs
Unique Wayback asset files	41,457	JavaScript, CSS, image, SVG bundles
Dated section captures (FFC)	35	11 sections × 3–4 time points each (Mar 2021 – Dec 2022)
Google Docs compare documents	38	.docx with native tracked-change markup
Product brochure PDF	1	catwalk_EN.pdf — 11 embedded JPEGs, InDesign-sourced
Collector observation CSVs	13	1,110 total rows cross-referenced at 99.0%
Distinct Webflow build hashes observed	14	Spanning 2020-08-09 through 2022-08-11
Distinct blog article slugs cataloged	110	Via collector’s FFC metadata TSVs

Key Metrics Summary

Metric	Value
Independent row-level reproduction of collector CSVs	1,099 / 1,110 (99.0%)
Webflow builds whose IA footer timestamps fall in the Dec 2022 cluster	14 of 14 (100%)

Metric	Value
Maximum HTML-capture-to-JS-footer-archive lag observed	871 days
Concentration of footer timestamps on top 5 dates (of 92 in window)	719 / 816 (88.1%)
Concentration on single peak day (2022-12-08 CST)	248 / 816 (30.4%)
Burstiness coefficient B over 92-day window	0.614 (Poisson baseline ≈ 0.0)
Collector-CSV rows with JS-footer-time < account-creation (2022-12-09 02:45 CST)	560 / 1,110 (50.5%)
Dec 9 HTML captures before 04:46 CST — Δ (JS archive – HTML) IQR	$\approx$ 0.03 hours (single-batch signature)
Dec 9 HTML captures after 04:46 CST — Δ flip to positive	+1.68 h median (fresh-crawl signature)
PhotoRobot blog articles claiming 2021 pub dates with no 2021 evidence	14
Maximum claim-vs-evidence publication-date discrepancy	510 days
Articles with mid-series introduction of patent-novelty phrase on product page	1 (infinite runway on blog_catwalk-for-apparel-photography-and-online-fashion-shows)
Tracked-change insertions of infinite runway / filming product videos / producing online fashion shows with zero deletions	5 insertions, 0 deletions
Virtual Catwalk page HTML size growth 2022-12-06 → 2022-12-28	+22.2% (+22,223 bytes)
KMeans k=2 cluster separation of 62 slugs by vocabulary	52 older / 10 newer
MAXQDA codes overlapping Guertin-patent novelty vocabulary	22 of 140 (15.7%)

Methods

Scope

The examined corpus is the static file-system snapshot preserved at `_workspace/.../forensic_investigation/` — every file downloaded by the collector during the December 2022 – January 2023 observation window, preserved without modification. This examination does NOT re-query the live Internet Archive. Where a question requires live Wayback CDX data (specifically: enumerating every historical snapshot of a given URL and comparing content digests across time), the examination names those queries and declines to perform them.

Source artifacts examined:

1. **Wayback HTML captures** (62 page slugs × 1,979 total captures). Each capture is the HTML as played back by Internet Archive at the time the collector requested it, with IA's `__wm.wombat` capture-timestamp annotation preserved.
2. **Wayback asset bundles** (41,457 files). Each capture has a companion `_files/` directory containing the JavaScript, CSS, image, and SVG assets the page referenced. Webflow emits JavaScript with content-addressed filenames (e.g., `photorobot.\1.js`); each such filename corresponds to a single site build.
3. **Internet Archive “footer-archived” timestamps** extracted from the on-disk JavaScript bytes. IA injects a textual footer block into every archived file identifying when the file was archived; parsing this footer yields the archive datetime for each served asset.
4. **Dated section captures** (35 total, “FIRST_FULL_COLLECT”). The collector made periodic top-down captures of site sections (Blog, Robots, Tutorials, Software, etc.) across three to four time points spanning March 2021 – December 2022. Each section capture preserves the section's article-listing page plus per-article HTML plus the Webflow CMS-field TSV used by each article.
5. **Google Docs compare files** (38 .docx with tracked-change XML markup). Produced by the collector in December 2022 via manual copy-and-paste of page text from adjacent Wayback captures into Google Docs, using Docs' built-in compare feature to surface the inter-capture differences. These documents preserve every insertion and deletion the collector observed in that period as machine-parseable XML.
6. **Collector observation CSVs** (13 files, 1,110 total rows). Produced by the collector during the observation window; each row records for a single Wayback capture: the page URL, the Wayback HTML-capture datetime, the IA footer-archived datetime on the served JavaScript, the filename of that JavaScript, the collector's own save date.
7. **Product brochure PDF** (`catwalk_EN.pdf`, 7.6 MB, 11 embedded JPEGs, InDesign-exported).

8. **MAXQDA heatmap image** (Matrix.png, 4,688 × 2,899 px) produced by the collector during a qualitative coding pass in December 2022.

Analytical approach

Every finding in this report was produced by a Python script run against the immutable `forensic_investigation/` tree. Scripts reside in three sequential analysis packages:

- `forensic_analysis/` — foundation pass: integrity baseline, temporal correlation, ground-truth extraction, image forensics null
- `forensic_analysis_v2/` — burst-clustering and vocabulary-scan refinement
- `forensic_analysis_v3/` — MAXQDA reconstruction and publication-date forensics

Each package produces its own CSV evidence files and visualizations under an `evidence/` and `visualizations/` subdirectory. The consolidated findings in this report are cross-referenced to those source artifacts by filename.

No originals were modified. The SHA-256 manifest captured at the start of the examination (`forensic_analysis/baseline/sha256_manifest.csv`) anchors the integrity chain. Any post-examination modification to the originals would be detectable by rerunning `sha256sum` and comparing to the manifest.

The collector's observation CSVs and Google Docs compare files were treated as validation targets, not as inputs. Findings were derived independently from the raw HTML and asset bundles, then compared to the collector's output to measure agreement. This is the methodological choice that produced the 99.0% row-level agreement reported as Finding 1.

Tools used

Standard Python analysis stack — `beautifulsoup4`, `trafilatura`, `pandas`, `python-docx`, `Pillow`, `numpy` (pinned to <2 for compatibility), `matplotlib`, `scikit-learn`, `imagehash`, `opencv-python`, `tesseract` (for OCR of the MAXQDA heatmap). All deterministic, all open-source, all specified with version pins in the `v1/v2/v3 REPRODUCE.md` files.

Finding 1 — CRITICAL: Independent Reproduction of Observer's Contemporaneous CSVs at 99.0%

Background

The collector documented their observations in real time during December 2022 – January 2023 by building thirteen spreadsheets that catalogued, for every Wayback capture they preserved: the URL, the Wayback HTML-capture datetime, the Internet Archive “footer-archived” timestamp of the JavaScript

bundle served inside that capture, the filename of that bundle, and the collector's own save date. These CSVs are preserved at `original_spreadsheets_detailing_js_dates__EVERY_WAYBACK_SAVE_PER_PAGE/` and `original_spreadsheets_detailing_js_dates__FIRST_FULL_COLLECT/` in the source corpus.

Because this examination can neither re-query the live Internet Archive nor produce new observations, the credibility of every downstream finding depends on the collector's contemporaneous observations being accurate. The first test in the examination is therefore to reproduce the collector's observation CSVs independently from raw artifacts — not by reading the CSVs, but by re-parsing the preserved HTML and asset bundles and computing what the CSV values should have been.

Findings

Of 1,110 total rows across the 13 collector CSVs:

Metric	Count	Percent
HTML capture datetime matched within 3-hour tolerance	1,100 / 1,110	99.1%
JS footer-archived datetime matched within 5-minute tolerance	1,099 / 1,110	99.0%
Both datetimes matched within tolerance (row-level agreement)	1,099 / 1,110	99.0%

The eleven unmatched rows are concentrated in `photorobot.<hash>.css` references (this examination's temporal pipeline scopes to JavaScript bundles) and in site-root pages that are not present as JavaScript-bundle-bearing captures in this examination's walk. No row produced a contradiction; the divergences are coverage-pipeline differences, not value disagreements.

Significance

The collector's thirteen observation CSVs, produced contemporaneously in December 2022 – January 2023, reproduce independently at 99.0% row-level agreement from the raw preserved artifacts. Every quantitative observation the collector recorded during the observation window is independently verifiable from the static corpus.

This establishes the methodological foundation for every subsequent finding in this report: downstream findings that reference the collector's observations are

not relying on unverified claims — they are relying on a contemporaneously-documented and independently-reproduced observation record.

Connecting Context

This finding is the credibility anchor. It is referenced implicitly by every subsequent finding in this report, and explicitly in Finding 2’s analysis of the observation-window timestamps, which incorporates the collector’s `savedOnDate` column as independent evidence of when each observation was recorded.

Reference evidence:

`forensic_analysis/evidence/collector_csv_validation.csv`.

Finding 2 — CRITICAL: Temporal Anomaly in Internet Archive JavaScript Bundle Timestamps

Background

Webflow — the site-building platform that produces `photorobot.com` — emits JavaScript bundles with content-addressed filenames. The filename portion following `photorobot.` and preceding `.js` is a hash of the bundle’s content. A given filename therefore corresponds to a single site build; rebuilding the site produces a different filename even if the content change is small. This property makes Webflow bundle filenames temporally unique and content-unique — and therefore useful forensic markers of “what site content existed at what time.”

The Internet Archive injects a textual footer into every archived file it serves, identifying the date and time at which that specific file was archived. For a Wayback capture of an HTML page that references a Webflow JavaScript bundle, two dates are relevant: **the HTML capture datetime** (when Wayback saved the HTML page) and **the JS footer-archived datetime** (when Wayback archived the specific JavaScript bundle that the HTML references). Under ordinary asynchronous Wayback crawler behavior, these should be temporally close — within days or weeks — because the crawler typically captures a page and its referenced assets in proximity.

Findings

For each Wayback HTML capture with a Webflow JavaScript reference (1,117 records across 62 page slugs), this examination computed $\Delta = \text{ia_footer_archived_dt} - \text{html_capture_dt}$.

Δ distribution:

Δ bucket	Records	Percent
$\Delta \leq 0$ d (JS archived before or same-day)	696	62.3%

Δ bucket	Records	Percent
as HTML)		
$0 < \Delta < 30$ d	0	0.0%
$30 \leq \Delta < 90$ d	1	0.1%
$90 \leq \Delta < 180$ d	50	4.5%
$180 \leq \Delta < 365$ d	21	1.9%
$\Delta \geq 365$ d	349	31.2%
Δ less than –30 days (JS archived well before HTML)	0	0.0%

Of the 420 records with $\Delta \geq 90$ days, the IA footer-archived timestamps collapse onto **five specific dates in December 2022**: 2022-12-06, 2022-12-08, 2022-12-09, 2022-12-11, 2022-12-28 (CST). Over the full 92-day window 2022-11-01 through 2023-01-31, only 15 of 92 days carry any records; 88.1% of records fall on the five named dates; the single peak day (2022-12-08 CST) carries 30.4% of all records. The burstiness coefficient $B = (\sigma - \mu) / (\sigma + \mu)$ over the 92-day per-day series is **$B = 0.614$** (a uniform crawler backfill would produce $B \approx 0$).

14 distinct Webflow site builds are affected — each with its content-addressed filename, spanning 2 years of organic site evolution, each with its referenced content IA-archived in December 2022:

Webflow build (filename tag)	HTML capture date(s) referenced in	IA footer archived	Lag (days)
photorobot.2b8fb786e.js	2020-08-09	2022-12-28	871
photorobot.48a79097d.js	2020-09-27	2022-12-09	803
photorobot.17c83c9c4.js	2020-10-28	2022-12-09	772
photorobot.865993bdd.js	2020-11-28	2022-12-06	738
photorobot.e14af0525.js	2021-01-17 ... 2021-01-20	2022-12-09	688
photorobot.4967dad85.js	2021-02-28	2022-12-06	646
photorobot.ac7685463.js	2021-04-16	2022-12-09	602
photorobot.f05ce9fe8.js	2021-05-11	2022-12-06	574

Webflow build (filename tag)	HTML capture date(s) referenced in	IA footer archived	Lag (days)
photorobot.422 5e734c.js	2021-07-27	2022-12-06	497
photorobot.85fa 7d478.js	2021-10-22	2022-12-06	410
photorobot.303 9cb0c8.js	2021-11-11	2022-12-06	390
photorobot.7fa3 17b45.js	2022-01-22	2022-12-09	321
photorobot.44a 65b812.js	2022-02-16	2022-12-28	316
photorobot.d9b 0dabce.js	2022-08-11	2022-12-08	119

Zero records have Δ less than –30 days. If Wayback were serving the temporally-closest archived snapshot of each referenced URL, a roughly symmetric distribution of positive and negative Δ would be expected. The one-sided asymmetry (100%) is consistent with IA holding only a single archive snapshot of the referenced bundle URL for these 14 builds — dated in the December 2022 cluster.

Cross-correlation with collector observation timing (Finding 1 CSVs, 1,110 rows):

- **560 of 1,110 rows (50.5%)** have jsArchiveDate that strictly precedes the collector's Internet Archive account creation at **2022-12-09 02:45 CST**. At minimum half of the archival events were not triggered by the collector's own Save-Page-Now activity.
- **Zero of 1,110 rows** have |jsArchiveDate – savedOnDate| within 6 hours. Only 15 rows have the gap within 24 hours. If collector-triggered saves had been causing fresh IA crawls of the referenced bundles, a substantial cluster of near-zero gaps would be expected. The absence of that cluster is consistent with bundle archival events occurring separately from, and temporally disjoint from, the collector's save activity.

The December 9, 2022 sign-flip at 04:46 CST — the specific timestamp at which the collector documented observing a real-time edit to the Wayback historical capture timeline of the PhotoRobot blog landing page — produces the single most distinctive temporal signal in the dataset:

Δ

bucket

(Dec 9

CST

only)

	n	Δ p25 (h)	Δ p50 (h)	Δ p75 (h)	IQR (h)
HTML capture d BEFORE 04:46 CST	146	-3.96	-3.95	-3.93	0.03
HTML capture d AFTER 04:46 CST	33	+1.68	+1.68	+2.82	1.14

146 HTML captures saved before 04:46 CST all had their referenced JavaScript served from an IA footer timestamp approximately **3 hours 57 minutes earlier**, clustered with an IQR of 0.03 hours (1.8 minutes) — the signature of a single batch archival event, not a diffuse crawler activity. 33 HTML captures saved after 04:46 CST show **Δ flip to positive** (+1.68 hours median) — meaning the JavaScript was IA-archived *after* the HTML was saved. For Wayback to serve a JavaScript snapshot whose footer timestamp postdates the HTML, the JavaScript URL must have been crawled fresh after the HTML fetch occurred. The sign flip at 04:46 CST is consistent with an active fresh-crawl operation on PhotoRobot URLs initiating at or near that moment — the moment the collector documented observing new historical captures appearing in the Wayback timeline live.

Significance

Two properties of the observed timing pattern are incompatible with an organic asynchronous Wayback crawler backfill:

1. **The burstiness profile.** A uniform crawler backfill distributes archival events across days and weeks; the observed pattern concentrates 88.1% of records on 5 of 92 days, with a single peak day carrying 30.4%.
2. **The Dec 9 04:46 CST inflection.** 146 records cluster with a 1.8-minute IQR around $\Delta = -3.95$ hours before the inflection; 33 records cluster around $\Delta = +1.68$ hours after. That is a tight batch followed by a tight second batch with a sign-flipping inflection — not a continuous backfill profile.

The observed pattern is consistent with a batched, operator-driven archival event concentrated on a small set of dates in December 2022, including an inflection coincident with the collector's documented live observation at 04:46 CST on December 9.

This examination does not adjudicate whether the content archived during those batches reflects authentic pre-existing PhotoRobot site state or reflects content modified contemporaneously with the archival event. That question depends on whether Internet Archive holds additional (earlier) snapshots of the referenced bundle URLs and whether those snapshots' content digests differ from the snapshots served during the cluster window. Both questions are answerable via live Internet Archive CDX API queries (see Recommendations) and are outside the scope of this static-artifact examination.

Connecting Context

Finding 2 connects forward to Finding 3 (the 22-day window 2022-12-06 → 2022-12-28 in which the Finding 2 cluster concentrates is the same window in which Finding 3's Virtual Catwalk page grew +22% in HTML size) and to Finding 4 (the specific phrases inserted on product pages per Finding 4 concentrate in the same window). The cross-corroboration paragraph in the Synthesis section consolidates the temporal overlap.

Reference evidence: forensic_analysis/evidence/js_archive_correlation.csv, forensic_analysis_v2/evidence/footer_timestamp_daily_histogram.csv, forensic_analysis_v2/evidence/burst_statistics.txt, forensic_analysis_v2/evidence/before_after_0446_dec9.csv, forensic_analysis_v2/visualizations/footer_timestamps_nov2022_jan2023.png, forensic_analysis_v2/visualizations/before_after_0446_cdf.png.

Finding 3 — CRITICAL: Fourteen Articles Claim 2021 Publication Dates With No Surviving 2021 Evidence

Background

PhotoRobot operates on Webflow. Webflow's CMS records each blog article's claimed publication date in a CMS field named pubSort (serialized as a value per article in the per-section metadata TSV the collector captured during each FFC walk). This is the claimed-publication-date field the site itself presents as authoritative.

For an article that legitimately existed on the live site at its claimed publication date, independent technical evidence of that existence should be recoverable:

1. A Wayback Machine capture of the article URL dated near the claimed publication date
2. An appearance of the article path in any periodic section-index capture from the claimed-publication-date period
3. An appearance of the article in any search-engine index from that period

The preserved dataset contains direct evidence of (1) and (2): the Blog - 03-08-2021 FFC section capture enumerates **53 blog article paths** that existed on photorobot.com as of March 8, 2021. Subsequent FFC collects (2022-08-11,

2022-09-25, 2022-12-09) enumerate what existed at those later dates. The EVERY_WAYBACK_SAVE_PER_PAGE tree contains the Wayback per-article capture history for 62 blog article slugs.

Findings

Of 110 blog articles with claimed publication dates and observable earliest-existence evidence in the preserved dataset:

Claim-vs-evidence discrepancy	Count	Percent
≥ 30 days	61 / 102	60%
≥ 180 days	21 / 102	21%
≥ 365 days (≥ 1 year claim precedes evidence)	14 / 102	14%
< 0 days (evidence precedes claim)	7 / 102	7%

The median discrepancy (39 days) is within the normal range for Wayback crawler first-visit delay. The right-tail and left-tail are the forensic signal.

The 14 articles with claim-vs-evidence discrepancy ≥ 365 days:

Slug	Claimed publication	Earliest observed existence	Discrepancy (days)
blog_which-businesses-3d-product-photography-platforms	2021-03-19	2022-08-11	510
blog_photographing-glass-products	2021-04-01	2022-08-11	497
blog_examples-3d-and-360-product-viewers	2021-04-06	2022-08-11	492
blog_visual-product-configurators-e-commerce	2021-04-13	2022-08-11	485
blog_advantage-s-product-configurators	2021-04-16	2022-08-11	482

Slug	Claimed publication	Earliest observed existence	Discrepancy (days)
blog_do-you-need-a-visual-product-configurator	2021-04-20	2022-08-11	478
blog_product-photoshoots-photorobot	2021-04-30	2022-08-11	468
blog_brick-and-mortar-to-online-retail	2021-05-10	2022-08-11	458
blog_macro-lens-product-photography	2021-05-11	2022-08-11	457
blog_post-pandemic-change-drivers	2021-05-20	2022-08-11	448
blog_content-creation-and-marketing-2021	2021-05-24	2022-08-11	444
blog_social-media-product-photos	2021-07-26	2022-08-11	381
blog_hardware-automated-product-photography	2021-08-04	2022-08-11	372
blog_spin-photography-dirt-bikes-and-quads	2021-08-09	2022-08-11	367

All 14 share an identical earliest-observed-existence date: 2022-08-11

— the date of the next FFC section capture following the March 2021 collect. None of the 14 appears in the March 2021 collect. None of the 14 appears as a standalone slug in the EVERY_WAYBACK_SAVE_PER_PAGE tree. Their claimed publication dates span 2021-03-19 through 2021-08-09 — a period during which the collector’s March 2021 collect was traversing the blog index every weekday for several weeks and produced a 53-article enumeration; none of the 14 is present in that enumeration.

Two additional articles demonstrate the pubSort field is not authoritative in the opposite direction:

Slug	Claimed publication	Earliest observed existence	Discrepancy (days)
blog_360-spin-photography-bicycles	2021-07-19	2021-03-08	–133
blog_grocery-product-photography	2021-07-15	2021-03-08	–129

Both articles are present in the March 2021 FFC blog snapshot, 4+ months before the dates they claim to have been published. The pubSort field on both articles cannot be the article’s actual first-appearance date on the live site, because independent technical evidence proves the articles existed earlier.

Eight articles were unpublished during the observation window. Across successive FFC collects, eight articles changed from a real claimed-publication-date in an earlier snapshot to Webflow’s 2111 unpublished-placeholder value in a later snapshot. No article changed from one real date to another real date; the only observed claimed-date changes are real-to-unpublished.

Significance

14 PhotoRobot blog articles claim publication dates between 2021-03-19 and 2021-08-09 that cannot be reconciled with any independent technical evidence of 2021 existence in the preserved dataset. The claimed publication dates are contradicted by:

- Absence from the contemporary March 2021 blog-index snapshot, which enumerated 53 other blog articles existing on that date
- Absence from any Wayback snapshot during the 11-17 months between claimed publication and next observed evidence
- Absence from the standalone per-article capture tree for the entire observation period

Two additional articles are direct evidence that the pubSort field is not authoritative for when an article first existed — their claimed publication dates are 4+ months after their demonstrable first appearance on the site.

The benign explanation for the 14 multi-year discrepancies — “articles were published in 2021 but remained unindexed by every snapshot and crawler until August 2022” — is possible in principle but implausible at scale for 14 distinct articles across 5 distinct claimed publication dates on a commercial marketing site whose primary function is search-engine visibility.

This examination does not claim the 14 articles constitute manufactured content. It documents that their claimed publication dates cannot be reconciled with the independent technical evidence of their first appearance in this dataset.

Connecting Context

Unlike Finding 2, this finding is fully adjudicable from the static corpus alone — it rests on internal contradiction between the site’s own claimed-publication-date field and the dataset’s own independent evidence of first appearance. No live CDX query is required to surface the contradiction, though one per article URL would further corroborate or reverse it (see Recommendations).

Finding 3 does not overlap at the page-slug level with Finding 2’s headline cluster pages (Virtual Catwalk, Turning Platform, Robots Catwalk) — the 14 flagged articles are all from the article-tier corpus that was absent from the Wayback per-article tree. Finding 3 and Finding 2 illuminate two different mechanisms operating on the same site during the same observation window.

Reference evidence: forensic_analysis_v3/evidence/pubdate_discrepancy.csv,
forensic_analysis_v3/evidence/pubdate_discrepancy_flagged.csv,
forensic_analysis_v3/evidence/claimed_pub_dates.csv,
forensic_analysis_v3/visualizations/pubdate_vs_earliest_observed.png.

Finding 4 — HIGH: Patent-Novelty Vocabulary Inserted on Product Pages in the Observation Window

Background

The collector produced 38 Google Docs compare files in December 2022 by copy-and-pasting the text of PhotoRobot pages from adjacent Wayback captures into Google Docs and using Docs’ built-in compare feature to surface the inter-capture differences. The resulting .docx files preserve every insertion and deletion as native tracked-change XML, attributed to a single author (“PhotoRobot”) with contemporaneous timestamps.

For this examination, the 38 compare files were machine-parsed to extract the full insertion and deletion corpus (824 tracked-change runs total). A watchlist of 68 phrases drawn from US Patent 11,577,177 — Guertin’s patent — was applied to the insertion text. The watchlist is extracted from the patent’s specification and claims and is organized by semantic category (apparatus, synchronization, virtual production, scope expansion, filming, digital character, orbital motion).

Findings

Tracked-change insertions of patent-novelty phrases with zero offsetting deletions:

Patent-novelty phrase	Insertions	Deletions	Net	Source documents
infinite runway	3	0	+3	_CATWALK.docx (1), _TURNING PLATFORM.docx (2)
filming product videos	1	0	+1	_CATWALK.docx (1)
producing online fashion shows	1	0	+1	_CATWALK.docx (1)

Three specific phrases drawn directly from US Patent 11,577,177's novelty-axis vocabulary appear as tracked insertions with no offsetting deletions anywhere in the 38-document corpus. All three concentrate in the two product-page documents most directly adjacent to the subject matter of the patent: the Virtual Catwalk product page (the rotating-platform-photography product that was cited as candidate prior art during the observation window) and the Turning Platform product page (also rotating-platform hardware).

Corroborating temporal evidence on the Virtual Catwalk page:

- **Mid-series phrase introduction.** The phrase infinite runway is absent from all 7 captures of blog_catwalk-for-apparel-photography-and-online-fashion-shows spanning 2020-11-28 through 2022-05-22 and is present in all 5 captures from 2022-09-25 through 2023-01-05. Across the full 3,195-capture corpus, this is the only clean mid-series introduction of a watchlist phrase on a single page with a pre/post split of this clarity.
- **Raw HTML size growth.** The Virtual Catwalk page HTML grew from 100,087 bytes (2022-12-06) to 121,951 bytes (2022-12-28) — a **+22.2% (+22,223 bytes) content addition** in the same 22-day window that concentrates Finding 2's IA-footer-timestamp cluster.

Corroborating temporal evidence on the Turning Platform page:

- infinite runway appears on robots_turning-platform for the first time in the 2022-12-06 FFC section capture — **the same date as Finding 2's #2 cluster date** — and is inserted twice via _TURNING PLATFORM.docx in that window.

Significance

The concentration of three specific patent-novelty phrases as insertions (with zero offsetting deletions) on the two product pages most directly adjacent to the patent's subject matter, concurrent with a +22% HTML size growth on the flagship page and a cross-corpus-unique mid-series phrase introduction on the same flagship page, represents a convergence of three independent text-level signals on a single product-page area during the same observation window.

The phrases themselves are notable:

- infinite runway — the term semantically encodes the patent's "unlimited distance in unlimited directions" novelty axis, applied to apparel filming
- filming product videos — represents a scope expansion from still photography (the product's stated original function) into video production (the patent's novelty)
- producing online fashion shows — represents a further scope expansion into event-format content (adjacent to the patent's "creating content" framing)

This examination does not claim the insertions constitute manufactured prior art. It documents that three specific phrases drawn from US Patent 11,577,177's novelty vocabulary were inserted on the Virtual Catwalk product page and Turning Platform product page during December 2022, with zero offsetting deletions, concurrent with a +22% HTML size growth on the Virtual Catwalk page and a cross-corpus-unique mid-series phrase introduction on the same page.

Connecting Context

Finding 4 concentrates temporally on the same 22-day window as Finding 2's IA-footer-timestamp cluster. The Virtual Catwalk page specifically is the single page on which all three of Findings 2, 4, and (in Finding 5's vocabulary axis) the newer-cluster vocabulary shift jointly concentrate. The cross-corroboration matrix in the Synthesis section consolidates this.

Reference evidence: `forensic_analysis/evidence/ground_truth_edits.csv`,
`forensic_analysis_v2/evidence/watchlist_insertions.csv`,
`forensic_analysis_v2/evidence/watchlist_midseries_introductions.csv`,
`forensic_analysis_v2/evidence/watchlist_section_timeline.csv`.

Finding 5 — HIGH: Observer's MAXQDA Code Analysis Reproduces Quantitatively Under Independent Examination

Background

During the observation window in December 2022, the collector performed a qualitative coding pass on the full PhotoRobot blog corpus using MAXQDA and produced a visual heatmap documenting the distribution of 140 hand-selected semantic codes across articles ordered by their purported publication date. The collector's visual observation: the top rows of the heatmap (most recently published articles) share a dense, narrow band of terminology; the bottom rows (older articles) share a different, sparser band; the transition appears discrete rather than gradual.

The original MAXQDA project file was not preserved. The visual heatmap (Matrix.png, 4,688 × 2,899 px) was preserved.

This examination reconstructs the collector's code list via OCR of the heatmap image, applies the codes across the full preserved corpus, and tests the collector's visual observation quantitatively via clustering and separating-code information-gain analysis.

Findings

Code extraction. 140 codes were OCR-extracted from the heatmap's row labels (Tesseract on a 4× Lanczos-upscaled left strip of the image). The extracted code list is preserved at `forensic_analysis_v3/evidence/maxqda_codes.txt`.

Corpus matrix computation. The 140 codes were applied across the 62 blog article slugs in the EVERY_WAYBACK_SAVE_PER_PAGE tree, using each slug's most recent Wayback capture as the current-state measurement and each slug's earliest capture as the baseline measurement. Per-cell output: binary presence and raw frequency.

Separating-code analysis. Slugs were partitioned into the older third and newer third by earliest-capture datetime (used as a proxy for publication date; see Finding 3 for the limitations of the site's own claimed dates). For each code, old-third presence rate, new-third presence rate, Δ , and Shannon-entropy information gain were computed.

Top separating codes:

Code	Old-third presence	New-third presence	Δ	Information gain (bits)
scanning	0.0%	35.0%	+35.0	0.2020
elevation	0.0%	30.0%	+30.0	0.1692
real-time	0.0%	25.0%	+25.0	0.1379

Code	Old-third presence	New-third presence	Δ	Information gain (bits)
axis	5.0%	35.0%	+30.0	0.1117
robotic camera	5.0%	35.0%	+30.0	0.1117
robotic camera arm	5.0%	35.0%	+30.0	0.1117
rotation	30.0%	65.0%	+35.0	0.0905
rotary	5.0%	30.0%	+25.0	0.0852
frame	15.0%	45.0%	+30.0	0.0800

The top eight separating codes are all technical-apparatus and motion/orbital terms. The only two codes that occur more frequently in older articles than in newer articles (integration, AI) are general-purpose marketing terms; their decline is consistent with a vocabulary shift from generic marketing language toward apparatus-specific language.

Article clustering. KMeans clustering on the binary-presence vectors with $k=2$ partitions the 62 slugs into a 52-slug “older” cluster and a 10-slug “newer” cluster. The clusters are separable on the temporal axis: the newer cluster’s slugs have earliest captures concentrated in late 2021 through 2022; the older cluster’s slugs have earliest captures spanning 2020-08 through 2021-10.

Overlap with US Patent 11,577,177 novelty vocabulary. 22 of the 140 MAXQDA codes substring-match at least one phrase in the patent-novelty watchlist used in Finding 4. The overlap includes (non-exhaustive): 3d design, 3d modeling, automation, avatar, axis, camera, cue, illusion, motion, motorized, photogrammetry, precise, render, rotation, turntable, virtual, virtually.

Significance

The collector’s visual observation — that newer articles share a different, narrower vocabulary band than older articles — reproduces quantitatively under independent examination. Eight codes separate the old and new corpora by $\Delta \geq 25$ percentage points; KMeans clustering produces a clean two-cluster partition along the temporal axis. 22 of the 140 codes the collector chose as semantically significant during the qualitative pass overlap with Guertin-patent novelty vocabulary.

This establishes that the vocabulary shift observed by the collector is not an artifact of visual bias or selective attention; it is a real, quantified, temporally-localized pattern in the corpus text. The pattern is consistent with either (a) organic editorial evolution over time as the site’s content focus shifted, or (b) a discrete authorship or content-regeneration event introducing a different

vocabulary band to newer articles. The $k=2$ clean separation favors (b); a continuous (a)-type evolution would produce a gradient rather than a two-cluster partition. This examination does not adjudicate between (a) and (b).

Connecting Context

Finding 5 complements Finding 4 by demonstrating that the patent-novelty vocabulary convergence observed in the 824 tracked-change insertions is part of a broader vocabulary shift visible across the full blog corpus, not only on the two product pages Finding 4 analyzed. 22 of the collector's 140 independently-chosen MAXQDA codes overlap with the Guertin-patent watchlist — two independently-sourced vocabulary lists (one by a subject-matter-informed human coder, one extracted from the patent itself) converging on overlapping terminology.

Finding 5 does not overlap at the article-slug level with Finding 3 (the 14 flagged articles are not in the EVERY_WAYBACK_SAVE_PER_PAGE tree that Finding 5 analyzed).

Reference evidence: forensic_analysis_v3/evidence/maxqda_codes.txt,
 forensic_analysis_v3/evidence/matrix_current_presence.csv,
 forensic_analysis_v3/evidence/separating_codes.csv,
 forensic_analysis_v3/evidence/article_clusters.csv,
 forensic_analysis_v3/evidence/maxqda_patent_overlap.csv,
 forensic_analysis_v3/visualizations/matrix_reproduction_current.png,
 forensic_analysis_v3/visualizations/cluster_vs_pubdate.png.

Cross-Corroboration Synthesis

The five findings partition the examined evidence into complementary slices rather than converging on a single narrow subset:

Finding	Corpus slice examined	Temporal concentration
1 — CSV reproduction	All 1,110 collector-observation rows	2022-12-09 → 2023-01-15 (observation window)
2 — IA timestamp anomaly	1,117 HTML captures with JS references	2022-12-06 → 2022-12-28 (22-day cluster)
3 — Publication date discrepancy	110 FFC-cataloged blog articles	Earliest evidence 2022-08-11 for 14 articles
4 — Patent-novelty insertions	38 .docx tracked-change files, 824 runs	Insertions dated in the observation window
5 — MAXQDA corpus separation	62 blog article slugs in TREE_A	Cluster boundary in late 2021 - 2022

No single article is flagged by all five findings. At the **window level**, however, all five findings concentrate signal in the **August 2022 - December 2022**

period, with a tight intra-window concentration on the 2022-12-06 → 2022-12-28 cluster. The observation coheres across findings on the time axis without requiring artificial article-level coincidence.

On the single product page `blog_catwalk-for-apparel-photography-and-online-fashion-shows`, three of five findings jointly concentrate:

- Finding 2 cluster dates (Dec 6, Dec 9, Dec 11, Dec 28) — present in this page's Wayback capture timeline for the relevant Webflow builds
- Finding 4 — 3 tracked-change insertions (infinite runway, filming product videos, producing online fashion shows) via `_CATWALK.docx`, mid-series introduction of infinite runway on 2022-09-25, +22.2% HTML size growth 2022-12-06 → 2022-12-28
- Finding 5 — cluster-membership and separating-code signals consistent with the newer-vocabulary corpus

The Virtual Catwalk product page is the single highest-density cross-finding observation in the examination.

Limitations & Alternative Explanations

1. **This examination did not re-query the live Internet Archive.** Several findings could be strengthened or reversed by a live CDX query per affected URL. Specifically: Finding 2's A-vs-B adjudication (benign crawler backfill versus contemporaneous archive modification) can be closed by the live CDX queries enumerated in the Recommendations section; Finding 3's 14 flagged articles can each be individually corroborated or reversed by a single CDX query per article URL.
2. **The `trafilatura` HTML-to-text extractor is imperfect on Webflow marketing pages.** Diff-reproduction agreement between the tracked-change `.docx` ground-truth set and this examination's independent HTML-pair diff pipeline is bounded below by extraction normalization differences. Cross-document agreement on specific watchlist phrases (Finding 4) is unaffected by this limitation.
3. **Image forensics on the product brochure PDF is not diagnostic of synthesis.** The `catwalk_EN.pdf` brochure's 11 embedded images all passed through Adobe InDesign CC 13.0's PDF export pipeline on 2017-11-24. Classical image-forensics metrics (Error Level Analysis, frequency analysis, noise-residual analysis, JPEG quantization-table fingerprinting) applied to InDesign-recompressed JPEGs are not reliably diagnostic of whether the source images were photographic or synthetic. The examination recorded a null result on this axis (neither supporting nor refuting synthetic-image claims) rather than a positive or negative finding.
4. **The MAXQDA code list was reconstructed via OCR from a preserved image.** The collector's original MAXQDA project file was not preserved. 140 codes were OCR-extracted with human review; a small

number of potentially misread codes are documented in the v3 methodology notes. The substantive findings are not sensitive to single-code OCR uncertainty.

5. **The Finding 3 pubSort authoritativeness assumption can admit a benign reading.** If any of the 14 flagged articles has a legitimate explanation — e.g., a scheduled-publication mechanism where the pubSort field reflects an intended-release date rather than an actual first-visible date — the discrepancy could be partially explained. That explanation would not account for why 11–17 months elapsed between each claimed publication date and the earliest observation of the article on the live site.
 6. **Potential benign explanations** for each finding are documented in the v1/v2/v3 source reports and briefly recapped where applicable. Where benign and non-benign readings coexist as plausible, this report declines to adjudicate and documents both.
-

Possible Explanations (Unresolved; Not Proven by This Report)

A minimally-speculative list, documented here solely to make the space of explanations legible:

1. **Coordinated December 2022 operator activity on the PhotoRobot site and its Wayback archive record.** Consistent with the temporal clustering in Finding 2, the inflection at the collector’s documented 04:46 CST observation, the tracked-change insertions in Finding 4, and the +22% HTML growth on the flagship page. Not proven; requires the live CDX queries in Recommendations to advance.
2. **Ordinary Webflow CMS authoring activity by a small content team in December 2022.** Consistent with Finding 4’s tracked-change insertions and Finding 5’s newer-cluster vocabulary shift, but does not explain Finding 2’s archival timing pattern (which is exterior to PhotoRobot’s site — it is a property of the Wayback Machine’s own records of PhotoRobot’s site).
3. **Articles in Finding 3 existed in 2021 but remained unindexed by every snapshot and crawler until August 2022.** Possible in principle; implausible at scale for 14 distinct articles on a commercial marketing site whose primary function is search-engine visibility.

No explanation in the above list is proven by this examination. Each is a hypothesis that the data permits; none is a hypothesis that the data establishes.

Recommendations for Independent Verification

For legal proceedings

- This report's findings are reproducible from the preserved `forensic_investigation/` dataset. The SHA-256 baseline manifest in `forensic_analysis/baseline/sha256_manifest.csv` anchors the integrity chain. Any reviewing expert with a copy of the dataset and a Python environment can regenerate every quantitative claim in this report.
- Finding 3 (14 articles claim 2021 pub dates with no surviving 2021 evidence) is the most technically unambiguous finding in the report and requires the least interpretive context to present to a fact-finder. It rests on internal contradiction between the site's own claimed-publication-date field and the dataset's own evidence of first appearance; no live CDX query is required for the contradiction itself (though queries would further corroborate or reverse the findings).
- Finding 1 (99.0% independent row-level reproduction of the collector's thirteen observation CSVs) is the credibility anchor. It establishes that the collector's contemporaneous documentation is accurate and that downstream findings referencing those observations are not relying on unverified claims.
- It is strongly recommended that a credentialed digital forensic examiner (CFCE, EnCE, or equivalent) independently verify the key findings before use in formal legal proceedings.

For independent forensic verification

- **Request live Internet Archive CDX enumeration** for each of the 14 Webflow build-tag filenames in Finding 2. Expected query shape:

```
https://web.archive.org/cdx/search/cdx?
  url=assets-global.website-files.com/.../photorobot.<hash>.js
  &output=json
  &fl=timestamp,statuscode,digest
```

If CDX returns only a single snapshot per filename, dated in the 2022-12-06 → 2022-12-28 cluster window, the December 2022 archival cluster is the only archival event for those bundle URLs, and the 100% one-sided asymmetry is fully explained by that property. If CDX returns multiple snapshots of the same bundle URL with differing content digests, that is direct forensic evidence of content modification on the archived asset over time and would be dispositive for mechanism (B) in Finding 2.

- **Request live Internet Archive CDX enumeration** for each of the 14 flagged article URLs in Finding 3. Expected query shape:

```
https://web.archive.org/cdx/search/cdx?
  url=photorobot.com/blog/<slug>
  &output=json
  &fl=timestamp,statuscode
```

If CDX returns any snapshots dated before 2022-08-11 for any of the 14 articles, the claimed-publication-date for those articles is corroborated by independent IA evidence. If CDX returns no snapshots before 2022-08-11 for any of the 14, the claimed-publication-date for those articles is contradicted by independent IA evidence.

- **Examine Webflow CMS export history** for photorobot.com if available to law enforcement. Webflow maintains CMS revision history; a forensic review of the pubSort field's revision history on the 14 flagged articles would show whether the field was set at or near the claimed publication date or at a later date.
- **Obtain DNS and WHOIS history** for photorobot.com and any related operator entities during the observation window.

For continued documentation

- Preserve the forensic_investigation/ dataset with its SHA-256 integrity manifest intact. The dataset is a contemporaneous record of a specific state of a third-party website's archival record on a specific set of dates; it cannot be reproduced from public sources after the fact.
 - The v1/v2/v3 source analysis packages under forensic_analysis/, forensic_analysis_v2/, and forensic_analysis_v3/ are this examination's working papers. They remain the authoritative technical documentation for reviewers who wish to verify specific scripts or intermediate evidence files.
-

Context: Related Documentation

This examination was performed to provide an independent technical evaluation of observations documented in real time by the collector (Matthew David Guertin) during December 2022 – January 2023. The observation record — contemporaneous emails to patent counsel, preserved digital-forensic notes, screen captures, and related correspondence — is separately documented in the following materials:

- **calls-for-help.html** — the narrative record of the December 2022 – January 2023 observation window presented on MnCourtFraud.com, including the chronological sequence of emails to patent counsel, screen-captured live observations (including the December 9, 2022 04:46 CST event referenced in Finding 2), and the progression of observations that preceded Guertin's January 21, 2023 arrest.
- **EML-D1 and EML-D2 exhibits** (Doc. 64 and Doc. 65 in federal case 0:25-cv-02670-PAM-DLM) — the digitally-authenticated email corpus documenting the real-time observation record. Each email is SHA-256 hashed and anchored via OpenTimestamps to the Bitcoin blockchain.
- **US Patent 11,577,177** — “Motorized Rotatable Treadmill and System for Creating the Illusion of Movement” — filed March 18, 2022

(provisional March 19, 2021), granted February 14, 2023. The patent whose novelty vocabulary is referenced in Findings 4 and 5.

- **The V1/V2/V3 source examiner reports** preserved under `forensic_analysis/`, `forensic_analysis_v2/`, and `forensic_analysis_v3/` — the working-paper-level documentation for each of the three sequential analysis passes. This consolidated report synthesizes from those source reports; the source reports remain the authoritative technical record for individual scripts and intermediate evidence.

This report does not re-narrate the observation-window record. Its purpose is to provide a neutral technical evaluation of the preserved dataset against the observations documented elsewhere. Readers interested in the contemporaneous context are directed to `calls-for-help.html` and the EML exhibits cited above.

Appendix A — Reproducible Scripts and Evidence Files

All quantitative claims in this report are reproducible from unmodified `forensic_investigation/` via the following Python scripts:

Integrity baseline

Script	Output
<code>forensic_analysis/scripts/w0_sha256_manifest.py</code>	<code>forensic_analysis/baseline/sha256_manifest.csv</code>
<code>forensic_analysis/scripts/w0_inventory.py</code>	<code>forensic_analysis/evidence/inventory.csv</code>

Finding 1 (CSV reproduction)

Script	Output
<code>forensic_analysis/scripts/w1_collector_csv_validation.py</code>	<code>forensic_analysis/evidence/collector_csv_validation.csv</code>

Finding 2 (temporal anomaly)

Script	Output
<code>forensic_analysis/scripts/w1_asset_timeline.py</code>	<code>forensic_analysis/evidence/asset_timeline.csv</code>
<code>forensic_analysis/scripts/w1_js_archive_ts.py</code>	<code>forensic_analysis/evidence/js_archive_correlation.csv</code>
<code>forensic_analysis/scripts/w1_visualize.py</code>	<code>forensic_analysis/visualizations/delta_js_vs_html.png</code> , <code>hash_vs_capture_date.png</code> , <code>bundle_timeline_html_vs_js.png</code>

Script	Output
forensic_analysis_v2/scripts/ v2_task1_burst.py	forensic_analysis_v2/evidence/ footer_timestamp_daily_histogram.cs v, burst_statistics.txt, collector_saved_vs_js_archive.csv, before_after_0446_dec9.csv
forensic_analysis_v2/scripts/ v2_visualize.py	forensic_analysis_v2/visualizations/ footer_timestamps_nov2022_jan2023. png, footer_timestamps_dec9_minutes.png , before_after_0446_cdf.png

Finding 3 (publication date forensics)

Script	Output
forensic_analysis_v3/scripts/ v3_task2_pubdate.py	forensic_analysis_v3/evidence/ claimed_pub_dates.csv, observed_existence.csv, pubdate_discrepancy.csv, pubdate_discrepancy_flagged.csv, claimed_pub_dates_changes.csv, pubdate_cross_corroboration.csv
forensic_analysis_v3/scripts/ v3_visualize_pubdate.py	forensic_analysis_v3/visualizations/ pubdate_vs_earliest_observed.png, pubdate_discrepancy_histogram.png

Finding 4 (patent-novelty insertions)

Script	Output
forensic_analysis/scripts/ w2_extract_docx_edits.py	forensic_analysis/evidence/ ground_truth_edits.csv, ground_truth_edits_summary.csv
forensic_analysis/scripts/ w2_html_to_text.py	forensic_analysis/evidence/ page_text_corpus/
forensic_analysis/scripts/ w2_diff_reproduction.py	forensic_analysis/evidence/ diff_reproduction_report.csv
forensic_analysis_v2/scripts/ v2_task2_watchlist.py	forensic_analysis_v2/evidence/ watchlist_terms.txt, watchlist_insertions.csv, watchlist_midseries_introductions.csv , watchlist_section_timeline.csv

Finding 5 (MAXQDA corpus separation)

Script	Output
forensic_analysis_v3/scripts/ v3_task1_matrix.py	forensic_analysis_v3/evidence/ maxqda_codes.txt,

Script	Output
forensic_analysis_v3/scripts/ v3_visualize_matrix.py	matrix_current_presence.csv, matrix_earliest_presence.csv, matrix_delta_presence.csv, separating_codes.csv, article_clusters.csv, maxqda_patent_overlap.csv forensic_analysis_v3/visualizations/ matrix_reproduction_current.png, matrix_reproduction_earliest.png, matrix_delta_current_minus_earliest.p ng, cluster_vs_pubdate.png

One-command reproduction

Each analysis package has its own REPRODUCE.md with a one-command pipeline:

- forensic_analysis/REPRODUCE.md
- forensic_analysis_v2/REPRODUCE.md
- forensic_analysis_v3/REPRODUCE.md

Running the three in sequence against the unmodified forensic_investigation/ tree regenerates every CSV, JSON, text file, and visualization referenced in this report.

Appendix B — Integrity Chain

- **SHA-256 baseline manifest:** forensic_analysis/baseline/sha256_manifest.csv — 77,603 rows, captured at the start of the V1 pass.
- **Per-file integrity verification:** rerun sha256sum on any file under forensic_investigation/ and compare to the manifest. Any mismatch indicates post-examination modification.
- **Analysis package integrity:** Each of forensic_analysis/, forensic_analysis_v2/, and forensic_analysis_v3/ can be independently SHA-256 hashed at publication time and anchored via OpenTimestamps if required for external attestation.

End of report.

Source: Matthew Guertin v. State of Minnesota | 27-CR-23-1886 (Related) | Internet Archive / photorobot.com Preserved Dataset | MattGuertin.com