

Evidence Pack

Sealed, reproducible analysis of review authenticity — prepared to evidentiary standards

Case ID: LS-4EFA12A7
Subject: Ristorante Fittizio (DEMO)
Jurisdiction: DE
Reviews analyzed: 9
Platforms: google
Generated (UTC): 2026-07-10T12:52:47.555Z

Cryptographic seal

Report integrity (SHA-256): 550ba3a56018f6b75df215b5a1ec44d0e50462f49de7d407af49381309367387
Dataset digest (SHA-256): cd37de1e0f30b98546a22239c289fa878c3868c9c56c4d39b8486c326c70fa83
Timestamp: **not requested**

This report is prepared to evidentiary standards and designed to be defensible. It presents probabilistic signals and a reproducible methodology; it does not assert that any individual review is fake with certainty, nor does it guarantee admissibility, which is a determination for the tribunal.

FIKTIVES DEMONSTRATION (BEISPIEL)

1. Independent Authenticity Opinion

Of 9 reviews analyzed, 4 exhibited one or more authenticity-risk signals at or above the reporting threshold (0 high-confidence, 4 likely, 0 suspicious). 5 were assessed as clean or low-concern.

VERDICT DISTRIBUTION

Strong multi-family signal convergence (85–100): 0

Converging independent signals (65–84): 4

Multiple signals marked for review (40–64): 0

Isolated signals marked (20–39): 0

No signals marked (0–19): 5

IMPACT ON DISPLAYED RATING

As displayed (all reviews): 2.8 / 5

On reviews without marked signals: 4.2 / 5

2. SCOPE & SOURCE ATTESTATION

The corpus comprises 9 review(s) from google, dated 2026-01-14T12:05:00Z to 2026-06-21T10:03:00Z. The reviews were supplied for analysis by the subject or its representative; the analysis provider did not select which reviews to include. Each exhibit below is bound to its content by a SHA-256 hash so any later alteration is detectable.

Source attested by: _____

Date: _____

FIKTIVES DEMONSTRATIONSBEISPIEL

3. Methodology & Tool Disclosure

The analysis runs a deterministic ensemble of detection engines. Each engine contributes signals; a fusion stage combines them into a calibrated 0–100 risk score and verdict band.

- **AI-generated-text analysis**

Probability that a review's text was produced by a language model, via independent detectors. Corroborative only — never used in isolation to reach a conclusion.

- **Stylometric fingerprinting**

Whether multiple reviews under different names share an authorship fingerprint (vocabulary, syntax, rhythm), indicating a single hand behind many identities.

- **Reviewer-network analysis**

Cross-review author relationships and persona patterns that distinguish organic reviewers from coordinated accounts.

- **Temporal & coordinated-burst analysis**

Unnatural timing clusters and synchronized posting consistent with an orchestrated campaign rather than organic arrival.

- **Sentiment-rating incongruity**

Mismatch between the emotional content of the text and the star rating, a common artifact of fabricated reviews.

- **Photo forensics**

EXIF metadata, perceptual hashing and reverse-image checks to flag recycled, stock or manipulated images attached to reviews.

- **Signal fusion**

An ensemble that combines the above into a single 0–100 authenticity-risk score and a calibrated verdict band, with hard-fail overrides for unambiguous artifacts (e.g. embedded prompt-injection text).

LIMITATIONS & HONEST CAVEATS

- These signals indicate likelihood, not proof. No method — automated or human — identifies an inauthentic review with certainty.
- Peer-reviewed research shows AI-generated review text is now statistically indistinguishable from human writing: in a 2025 study human raters averaged 50.8% accuracy at telling real from AI-generated product reviews (chance is 50%), and the best-performing language model (GPT-4o) reached only 50.0% (Meng et al., "Large Language Models as Hidden Persuaders", arXiv:2506.13313, 2025). AI-text signals are therefore corroborative only and are never used alone to reach a verdict.
- AI-text detectors have a documented bias against non-native English writing: across seven commercial GPT detectors an average of 61.3% of non-native (TOEFL) essays were misclassified as AI-generated, while native-English writing was classified near-perfectly (Liang et al., "GPT detectors are biased against non-native English writers", Patterns 4(7):100779, 2023). Such signals are down-weighted within multi-signal convergence and this caveat is disclosed rather than hidden.
- A one-shot audit without longitudinal author history runs the network and persona signals in a degraded mode; conclusions rely more heavily on content-intrinsic signals.
- Reviews that were assessed and cleared are listed as exceptions. Their inclusion is part of the integrity of the method, not an omission.
- This report is prepared to evidentiary standards and is designed to be defensible. Whether it is admitted, and the weight it is given, is determined by the tribunal — not by the analysis provider.

TOOL VERSIONS (FOR REPRODUCIBILITY)

engine 0.1.0
product 0.1.0
runtime v20.19.5

4. Per-Review Evidence

Every analyzed review is listed, including those assessed and cleared (disclosed as exceptions).

EX-001 NO SIGNALS MARKED (score 4)

google · 2026-01-14T12:05:00Z · 5/5 · reviewer: Rezensent 1

Signals: none above reporting threshold

Gate disposition: auto_clear — low score (4) and no medium+ signal

"Fantastische Pasta, aufmerksamer Service. Kommen wieder!"

SHA-256 c8966a940efd426eb9c12fb9eac634081cf5e3547ff2919e059a35bf7d1cd82b

EX-002 NO SIGNALS MARKED (score 4)

google · 2026-02-20T19:40:00Z · 4/5 · reviewer: Rezensent 2

Signals: none above reporting threshold

Gate disposition: auto_clear — low score (4) and no medium+ signal

"Gutes Essen, etwas laut am Wochenende, aber empfehlenswert."

SHA-256 78cbbf30c5a20df5b2e55168000a3cde4b148438911b277dc5cefaa5f26a9b74

EX-003 NO SIGNALS MARKED (score 4)

google · 2026-03-28T13:22:00Z · 5/5 · reviewer: Rezensent 3

Signals: none above reporting threshold

Gate disposition: auto_clear — low score (4) and no medium+ signal

"Bestes Tiramisu in Berlin, ehrlich."

SHA-256 22292eff0bd0967d582a67e48956f93b41a2940fdf8e599d135eab9c89b9d23d

EX-004 NO SIGNALS MARKED (score 4)

google · 2026-05-06T20:15:00Z · 5/5 · reviewer: Rezensent 4

Signals: none above reporting threshold

Gate disposition: auto_clear — low score (4) and no medium+ signal

"Nette Atmosphäre, faire Preise."

SHA-256 b17b330709b9e00183982bc8b804f1e6f9fc62a2aefa8d9a590a3f12ee17332c

EX-005 NO SIGNALS MARKED (score 12)

google · 2026-06-02T18:30:00Z · 2/5 · reviewer: Rezensent 5

Signals: none above reporting threshold

Gate disposition: auto_clear — low score (12) and no medium+ signal

"Pizza war leider kalt als sie ankam. Schade, sonst nett."

SHA-256 530230237d8bd8a327b6808b916cc51b751926620c26254c0f1d0649124ebf17

EX-006 CONVERGING INDEPENDENT SIGNALS (score 84)

google · 2026-06-20T18:02:00Z · 1/5 · reviewer: Rezensent 20

Signals: coordinated_negative_burst (sev 78, conf 0.86) · stylometric_match (sev 65, conf 0.8) · burner_account (sev 55, conf 0.75)

Gate disposition: auto_contest — score 84 corroborated by 3 independent evidence families (coordination, stylometry, account)

"Absolut schrecklich, niemals wieder, Finger weg!"

SHA-256 8602ca831dc45b52e3da67b9c0da8acd1fb3ce96062a07bb9f7a8c4dfa2b6b72

EX-007 CONVERGING INDEPENDENT SIGNALS (score 84)

google · 2026-06-20T18:44:00Z · 1/5 · reviewer: Rezensent 21

Signals: coordinated_negative_burst (sev 78, conf 0.86) · stylometric_match (sev 65, conf 0.8) · burner_account (sev 55, conf 0.75)

Gate disposition: auto_contest — score 84 corroborated by 3 independent evidence families (coordination, stylometry, account)

"Schrecklicher Laden, Finger weg, niemals wieder!"

SHA-256 00295b715dd67b207cd0422b1962b8c93113bad2e5f9f648a0269321b2b57bb7

EX-008 CONVERGING INDEPENDENT SIGNALS (score 84)

google · 2026-06-21T09:15:00Z · 1/5 · reviewer: Rezensent 22

Signals: coordinated_negative_burst (sev 78, conf 0.86) · stylometric_match (sev 65, conf 0.8) · burner_account (sev 55, conf 0.75)

Gate disposition: auto_contest — score 84 corroborated by 3 independent evidence families (coordination, stylometry, account)

"Niemand wieder, absolut schrecklich, Finger weg!"

SHA-256 46f3b90f14a8923082f6e9a8009425dc68836453dfd3a1379902ebb63ac27a38

EX-009 CONVERGING INDEPENDENT SIGNALS (score 84)

google · 2026-06-21T10:03:00Z · 1/5 · reviewer: Rezensent 23

Signals: coordinated_negative_burst (sev 78, conf 0.86) · stylometric_match (sev 65, conf 0.8) · burner_account (sev 55, conf 0.75)

Gate disposition: auto_contest — score 84 corroborated by 3 independent evidence families (coordination, stylometry, account)

"Finger weg von diesem schrecklichen Laden, nie wieder!"

SHA-256 024420fa47ee5d5735df5b567a122007ff8d07505f47892bf8dad0324b396e7e

FIKTIVES DEMONSTRATIONSBEISPIEL

5. Coordinated-Inauthentic-Behaviour Findings

Beyond per-review scoring, the analysis tests for CASE-LEVEL coordination across reviewers, accounts, and surfaces (statistically validated network: exact hypergeometric test with Bonferroni family-wise correction, plus cross-platform and cross-domain linkage engines).

Finding: no significant global coordination pattern was statistically validated for this corpus at the thresholds applied. Where individual reviews carry coordination-family signals, those are per-review indicators listed with each exhibit in Section 4 — they do not, at this corpus size, aggregate into a statistically validated global pattern, and this report does not claim one.

FIKTIVES DEMONSTRATIONSBEISPIEL

6. Chain of Custody

The following record binds this report to the exact data and tooling used, so the analysis can be independently reproduced and any tampering detected.

Report generated (UTC): 2026-07-10T12:52:47.555Z

Case ID: LS-4EFA12A7

Reviews in corpus: 9

Report integrity (SHA-256): 550ba3a56018f6b75df215b5a1ec44d0e50462f49de7d407af49381309367387

Dataset digest (SHA-256): cd37de1e0f30b98546a22239c289fa878c3868c9c56c4d39b8486c326c70fa83

Reproduction: re-running the same engine version over the attested corpus regenerates an identical report-integrity hash. The dataset digest is the RFC 6962 Merkle root over the per-exhibit content hashes: the hashes are sorted (making the digest order-independent), each leaf is SHA-256(0x00 || content-hash bytes), and each interior node is SHA-256(0x01 || left || right).

QUALIFIED ELECTRONIC TIMESTAMP

No qualified timestamp was attached at generation time. The report integrity hash above can be submitted to a Qualified Trust Service Provider at any time to obtain the eIDAS Art. 41–42 presumption of integrity.

INDEPENDENT VERIFICATION

This exhibit is self-verifying — no trust in SherlockCore is required. Attached to this PDF: manifest.json, the canonical evidence record in RFC 8785 (JCS) form, whose SHA-256 equals the report integrity hash printed above — recompute it with any conformant JCS + SHA-256 tool to reproduce the seal. No RFC 3161 token was requested for this run — the seal is content-integrity only (JCS hash + Merkle root).

FIKTIVES DEMONSTRATIONS BLEND