AI Hallucination & Primary-Source Audit Checklist (Aligned with NCSC & UK Standards)

Atomic claim extraction, primary-source crosswalks, and independent Judge AI verification gates.

2026 RESEARCH BENCHMARK: 2026 Verification Standard: An AI model writing 'approved' in its own JSON output is NOT independent verification. Author and Judge models must be separated by a hard evidence boundary.

1. 2026 Core Comparison & Operational Decision Matrix

Source Tier	Authority Level	Examples	Publication Policy
Tier 1: Statutory / Official	Highest (Primary)	GOV.UK, ONS, Companies House, NCSC, Legislation.gov.uk	Cite directly with retrieval date
Tier 2: Peer-Reviewed	High (Empirical)	NIH/PubMed RCTs, IEEE, ACM, University Labs	State sample context & metric
Tier 3: Industry Benchmarks	Medium-High	BCC/Tussell, McKinsey, Adobe Analytics datasets	Attribute named report & year
Tier 4: First-Party Telemetry	Internal Truth	Verified company SQLite logs & live server tests	Label clearly as internal benchmark
Tier 5: Unattributed LLM Text	Zero Trust	Raw ChatGPT/Claude prose without URL proof	BLOCK until verified or removed

2. Core System Architecture & Operational Pillars

UltimateTruth.co.uk is an independent claim-verification and hallucination-auditing platform designed for UK legal, financial, medical, and B2B publishing teams. Standard large language models generate fluent prose by predicting token probabilities rather than verifying factual truth, frequently fabricating statutory references, numerical benchmarks, or URLs. Operating on an independent 'Author-vs-Judge' separation architecture aligned with UK NCSC secure AI principles and strict primary-source verification standards, Ultimate Truth breaks any document or webpage into atomic factual assertions. Each assertion is checked against authoritative primary registries (such as GOV.UK, Companies House, ONS, and peer-reviewed journals), producing a line-by-line provenance ledger before publication.

Module Code	System Component	Target Metric / Output	Stage
CLAIM #01 · VERIFIED (98%)	'UK SME direct public procurement reached £45.2B (21%).'	Primary Source: BCC / Tussell 2025-2026 Report	Verified Match
CLAIM #02 · FLAGGED	'All UK SMEs are automatically fined £50,000 for late VAT filing.'	Correction: HMRC uses a graduated points-based penalty regime	Hallucination Caught
CLAIM #03 · VERIFIED (100%)	'Bulgaria adopted the Euro (€) on 1 January 2026.'	Primary Source: ECB / EU Council Official Journal	Verified Match
CLAIM #04 · PARTIAL	'AI code assistants always reduce security bugs.'	Context: Veracode/NCSC note 40%+ vulnerability rate without review	Needs Caveat

3. 5-Step Immediate Implementation Checklist

[ ] STEP 01	Extract every number, percentage, date, price, and statutory claim from the draft into a table.
[ ] STEP 02	Verify that every cited external URL returns HTTP 200 and actually contains the claimed figure.
[ ] STEP 03	Never relabel an older benchmark date as '2026' unless freshly re-verified from the primary source.
[ ] STEP 04	Remove self-serving fake review stars or unsubstantiated '#1 in the UK' superlatives.
[ ] STEP 05	Require an independent Judge pass (separate from the writer prompt) before live deployment.

4. Deterministic ROI & Performance Calculation Formula

Document Truth Index (%) = (Claims Backed by Tier 1–4 Primary Sources) ÷ (Total Atomic Factual Claims) × 100

5. Key Implementation Questions & Field Notes

• Why can't the same AI model that wrote an article verify its own facts?
When a model reviews its own output in the same context window, it suffers from confirmation bias and repeats its own hallucination. Independent verification requires a separate Judge architecture that only scores atomic claims against retrieved external documents.

• What is an 'Atomic Claim Decomposition'?
Instead of asking 'Is this paragraph true?', our system splits the text into individual verifiable statements (dates, percentages, named laws, prices) and requires a live HTTP 200 primary source for each one.

• What is inside the free 2-page Audit Checklist PDF?
Page 1 provides the 5-Tier Source Credibility Hierarchy (Primary Statutory -> Peer-Reviewed -> Named Industry Report -> Practitioner -> Unverified). Page 2 contains the 12-Point Pre-Publication Hallucination Gate.

• Does this help with Google Search and AI Overviews compliance?
Yes. Google's helpful content and spam policies heavily penalise scaled, unverified claims. Publishing a verified provenance trail builds genuine E-E-A-T.

• How is the PDF delivered?
It is emailed to you immediately upon joining the waitlist and available for instant download on the thank-you page.

© 2026 Ultimate Truth (ultimatetruth.co.uk) · All Rights Reserved · Direct Desk: hello@ultimatetruth.co.uk · Page 2 of 2