PANGBURN ANALYSIS

FORENSIC FACT-CHECKING REPLICATION STANDARD

A reproducible LLM protocol for transcript auditing, source verification, evidence synthesis, and misinformation analysis

Version 1.0 | August 2026

CORE PRINCIPLE

The objective is not to make an LLM sound certain. The objective is to make unjustified certainty difficult. Every important conclusion must be traceable to a precisely attributed claim, a transparent evidence chain, an explicit uncertainty judgment, and an auditable correction path.

Designed for independent replication across general-purpose LLM systems

Pangburn epistemic commitments: fallibilism, correspondence truth, evidential proportionality, source fidelity, reproducibility, and symmetry of standards.

Document status

This document is a methodological standard and prompt package for people who want to reproduce the Pangburn Analysis fact-checking process on their own LLM systems. It is deliberately stricter than an ordinary conversational fact check. The protocol assumes that LLMs can confabulate, misattribute speakers, attach citations that do not entail the sentence they appear to support, overstate weak evidence, collapse disputed questions into binary answers, and produce numerically impressive but invalid summary statistics. The process is built to detect those failure modes rather than merely instruct the model to “be accurate.”

It is not a guarantee of truth. No procedure can make factual error impossible. The appropriate scientific objective is error minimization under fallibilism: maximize traceability, independent verification, calibration, and opportunities for correction while reducing hidden assumptions and irreversible editorial judgments.

MINIMUM STANDARD

If the operator cannot reconstruct why a verdict was reached, identify the exact transcript segment, inspect the evidence both for and against the claim, and reproduce the conclusion from named sources, the verdict is not publication-ready.

How this standard was derived

The protocol formalizes the working safeguards developed in Pangburn Analysis transcript audits: exact speaker attribution, exact quotation, timestamp preservation, atomic claim decomposition, primary-source preference, explicit uncertainty, denominator/provenance controls, independent audit passes, and a strict separation between “unsupported” and “false.” It also incorporates compatible practices from professional fact-checking, systematic evidence review, reproducibility standards, and AI risk management. The references section documents those external methodological influences.

Executive summary

A high-reliability LLM fact check should be treated as a miniature forensic research project, not as a single prompt-response exchange. The unit of analysis is not “the video” or “the speaker”; it is the atomic, time-bounded, speaker-locked factual proposition. The unit of evidence is not “a source”; it is a specific proposition supported or undermined by a specific passage, dataset, image, recording, statute, judgment, or methodological record.

- Lock the source before judging it: acquire the best available original media, preserve timestamps, and identify whether the transcript is human, ASR-generated, OCR-derived, translated, clipped, or incomplete.
- Lock the speaker before rating the claim: do not infer attribution from topic alone. Use turn structure, audio, visual cues, self-reference, vocatives, and continuity; mark unresolved segments as HOLD - AUDIO REQUIRED.
- Quote before paraphrasing: verdicts attach to exact words in context. Editorial summaries are secondary artifacts and must never silently replace the original claim.
- Atomize before researching: split compound statements into independently testable propositions. A sentence can contain supported, false, disputed, and non-falsifiable components simultaneously.
- Search adversarially: retrieve the strongest evidence that would support the claim and the strongest evidence that would defeat it. The research plan should not be determined by the speaker’s politics or the fact-checker’s prior beliefs.
- Evaluate source quality multidimensionally: directness, independence, expertise, methodology, transparency, incentives, contemporaneity, completeness, and corroboration matter more than prestige alone.
- Separate verdict from confidence: FALSE, UNSUPPORTED, QUALIFIED, SUPPORTED, HOLD, DISPUTED, or NON-RATEABLE are evidentiary classifications; HIGH/MODERATE/LOW confidence describes how secure the classification is.
- Protect denominators: never convert a convenience set of audited claims into a speaker-wide “misinformation rate.” Percentages require a defined population, inclusion rule, duplicate policy, and validated speaker provenance.
- Run independent audits: attribution audit, evidence audit, citation-entailment audit, adversarial audit, and arithmetic/denominator audit should be distinct passes. High-impact findings should be independently replicated.
- Publish the uncertainty: the reader should be able to distinguish what is established, what is inferred, what is contested, what is unknown, and what is merely rhetoric.

Contents

1. Epistemic constitution

Principle	Operational meaning
Correspondence truth	Treat factual truth as correspondence between a proposition and the best available evidence about the world. Popularity, ideological usefulness, institutional prestige, moral desirability, and rhetorical force do not make a factual proposition true.
Fallibilism	Every conclusion is revisable. High confidence means strong evidence under the present record, not metaphysical certainty. The system must be able to say “I do not know,” “the evidence is insufficient,” and “my previous classification was wrong.”
Evidential proportionality	Strength of language must scale with strength of evidence. “Possible,” “plausible,” “likely,” “well-supported,” and “established” are not interchangeable.
Symmetry of standards	Equivalent claims receive equivalent evidentiary burdens regardless of speaker, ideology, nationality, religion, or political side. Advocacy by the source is relevant to source evaluation, not a reason for automatic rejection.
Source fidelity	The fact-checker must represent what a source actually says, including its limitations, definitions, time window, sample, legal posture, and uncertainty. A citation that merely concerns the same topic is not evidentiary support.
Atomism	Evaluate independently testable propositions independently. Compound rhetoric should not be rated as a single monolith when its components differ in truth value.
Reproducibility	A competent third party should be able to reconstruct the path from transcript -> claim -> evidence -> verdict. Search queries, source versions, timestamps, and reasoning summaries should be logged when practical.
Adversarial self-critique	Before publication, actively search for the strongest case that the proposed verdict is wrong. The audit is not complete until the model has attempted to falsify its own conclusion.
Uncertainty preservation	Ambiguity in the source must survive into the report. Do not “resolve” unclear pronouns, merged captions, legal uncertainty, or incomplete statistics by silently choosing the interpretation most convenient to the narrative.
Correction priority	A visible correction is evidence of a functioning epistemic system, not a failure to be hidden. Version history should make changed verdicts and reasons legible.

EPISTEMIC MAXIM

Prefer an unresolved claim to a falsely resolved claim; prefer a qualified verdict to a rhetorically satisfying overstatement; prefer a correction to the preservation of institutional pride.

2. Threat model: how LLM fact checks fail

Failure mode	What goes wrong
Confabulation	The model invents a fact, quotation, source, page number, date, statistic, or legal holding.

Failure mode	What goes wrong
Citation laundering	A real source is cited but does not entail the claim, or only supports a weaker neighboring proposition.
Speaker substitution	The model correctly reads words but assigns them to the wrong participant, especially around interruptions, chapter transitions, or ASR-merged turns.
Quote normalization	The model cleans grammar or repairs captions and then presents the repaired sentence as verbatim speech.
Claim mutation	The researched proposition becomes subtly different from the original claim, often narrower when defending it or broader when criticizing it.
Premature binary classification	A complex or disputed claim is forced into true/false before relevant definitions, time frames, or evidentiary thresholds are established.
Source-count fallacy	Several derivative articles are treated as independent corroboration even though all trace back to one primary allegation.
Authority substitution	Institutional prestige is treated as proof without inspecting method, scope, definitions, or internal uncertainty.
Temporal contamination	Later knowledge is projected backward into what a person could have known at the time of the original statement.
Denominator failure	A set of selected claims is converted into a percentage that appears to characterize an entire speaker or video.
Legal-category collapse	Distinct legal questions - allegation, reasonable grounds, probable cause, indictment, provisional measures, merits judgment, conviction - are treated as equivalent.
Numerical denominator drift	A percentage changes meaning because population, time window, numerator definition, or excluded cases are not held constant.
False balance	The model mechanically gives equal weight to evidentially unequal positions in the name of neutrality.
Motivated asymmetry	The model scrutinizes one side’s sources aggressively while accepting the favored side’s claims at face value.
OCR/ASR overconfidence	Image-based transcripts, automatic captions, translations, and screenshots are treated as if they were authoritative human transcripts.
Context stripping	A sentence is rated without the qualifying clause, question it answers, preceding definition, or later correction.
Rhetoric-as-fact	Hyperbole, moral judgment, analogy, sarcasm, or prescription is scored as though it were a literal empirical proposition.
Fact-as-opinion escape	A speaker embeds a factual assertion inside an opinion and the system declines to check the factual component.

3. System architecture and research environment

The protocol is model-agnostic. A single general-purpose LLM can execute it, but reliability improves when the workflow separates retrieval, attribution, verification, synthesis, and auditing into distinct passes or agents. Separation reduces anchoring: an auditor that did not author the first verdict is more likely to detect unsupported assumptions.

Layer	Function
Layer 0 - Evidence store	Original video/audio, transcript versions, screenshots, uploaded documents, primary documents, datasets, archived web pages, legal texts.

Layer	Function
Layer 1 - Source/provenance processor	Records origin, date, version, completeness, transcript type, OCR/ASR status, and available timestamps.
Layer 2 - Attribution processor	Segments speaker turns and assigns speaker + confidence; routes uncertain segments to manual/audio review.
Layer 3 - Claim processor	Extracts atomic propositions, labels claim type, preserves exact quote and context window.
Layer 4 - Research/retrieval	Builds search plan, retrieves primary and independent sources, records search gaps and failed attempts.
Layer 5 - Evidence synthesis	Constructs support/undermine matrix, evaluates quality, resolves definitions and time windows.
Layer 6 - Verdict engine	Assigns classification and confidence with justification and correction.
Layer 7 - Audit layer	Independent attribution, citation, adversarial, arithmetic, legal-status, and denominator checks.
Layer 8 - Publication layer	Produces human-readable report, machine-readable ledger, limitations, references, and correction log.

Recommended separation of roles

- Researcher A: build the strongest evidence-supported interpretation of the claim.
- Researcher B: attempt to falsify the proposed interpretation and find omitted counterevidence.
- Attribution auditor: ignore factual merits and inspect only who said what, when, and with what confidence.
- Citation auditor: inspect every citation for entailment, source quality, independence, and correct representation.
- Quantitative auditor: recompute every number, percentage, ratio, date difference, and denominator.
- Final editor: reconcile disagreements without erasing uncertainty; no verdict is upgraded merely because reviewers disagree.

3.1 Model configuration for reproducibility

A replication protocol is weakened when the model, search environment, and tool settings are allowed to drift invisibly. The operator should preserve enough configuration detail that another analyst can understand which capabilities were available and why two runs may differ.

- Record model/provider name, version or dated snapshot when exposed, and the date of the run.
- Record whether web search, file retrieval, code execution, audio/video inspection, OCR, and external databases were available.
- For APIs, record temperature/sampling settings, system prompt version, retrieval settings, and context-window/chunking strategy when practical.
- For long transcripts, prefer deterministic segmentation and stable claim IDs over asking the model to rediscover the entire structure in every pass.
- Pin the protocol version. If the prompt changes materially, increment the version and do not silently pool results across versions.
- Preserve search dates for time-sensitive claims because current office-holders, casualty totals, court status, prices, regulations, and datasets can change.

3.2 Validation before deployment

Before trusting an LLM configuration on a high-stakes public audit, test it on a small gold-standard set that contains known traps: merged speakers, incorrect ASR words, compound claims, early-versus-final statistics, derivative sources, legal-status distinctions, and citations that are relevant but non-entailing. The benchmark should measure process failures, not merely final-answer accuracy.

Validation dimension	What to test
Speaker attribution accuracy	Fraction of gold segments assigned to the correct speaker; report HOLD separately rather than forcing guesses.
Quote fidelity	Exact or audio-verified match rate for publication quotes; track silent “cleanups” as errors.
Atomic-claim recall	How many independently testable propositions were captured without mutation or improper merging?
Citation entailment precision	Among cited support links, what fraction actually entails the sentence at the stated scope and certainty?
Verdict agreement	Agreement with adjudicated gold verdicts, analyzed by class rather than one aggregate score.
Calibration	Do HIGH-confidence judgments actually fail less often than MODERATE or LOW judgments?
Denominator integrity	Can the model resist producing a percentage when the claim population is incomplete or provenance is mixed?
Adversarial sensitivity	When supplied strong counterevidence, does the model revise appropriately rather than defend its first answer?

BENCHMARK PRINCIPLE

A system that is highly accurate on easy factual questions can still be unsafe for forensic transcript auditing if it over-attributes speakers, invents citations, or refuses to downgrade confident early conclusions.

4. Source intake and provenance lock

A transcript is evidence about speech only after its provenance is characterized. YouTube automatic captions, screenshots of captions, OCR from a PDF, third-party transcript sites, clipped social-media videos, and human transcripts have different error profiles. The source register should be completed before claim extraction.

Field	Required question
Canonical media URL / file	Where is the original or highest-quality accessible recording?
Media duration	Does the transcript cover the entire item or an excerpt?
Capture date	When was the file or page captured?
Publication date	When did the underlying media first appear?
Transcript origin	Human / platform ASR / third-party ASR / OCR of screenshot / translated / unknown.
Speaker labels	Native / inferred / absent / unreliable.
Timestamp quality	Exact / approximate / chapter-level / absent.
Audio available	Yes / partial / no.
Visual speaker cues	Yes / no / not applicable.
Editing status	Live / edited / clip / montage / unknown.
Completeness	Full / partial / pages missing / source removed.
Known transformations	Compression, subtitles, translation, OCR, screenshots, re-upload, cuts.
Archive identifiers	Local filename, checksum/hash if used, archived URL, version number.

RULE

Do not treat a caption screenshot as a verbatim transcript merely because the text is legible. Legibility and accuracy are different properties.

5. Transcript reconstruction and speaker attribution

Speaker attribution is a first-order evidentiary problem. If the wrong person is assigned a quotation, every downstream fact-check can be technically excellent and still be false. Automatic transcripts commonly merge interruptions, drop punctuation, shift timestamps, and omit speaker labels. Pangburn Analysis therefore uses a “speaker lock” before rating consequential claims.

5.1 Evidence hierarchy for speaker attribution

Class	Basis	Default confidence
A	Direct audio + visible speaker or reliable multitrack/diarized source	Highest
B	Direct audio without video, using distinctive voices and conversational turn structure	High
C	Transcript with explicit native speaker labels corroborated by content	High
D	Unlabeled transcript inferred from turn-taking, vocatives, self-reference, questions/answers, and topic continuity	Moderate
E	Stylistic attribution from language patterns alone	Low-to-moderate
F	Topic-based guess or “this sounds like what X believes”	Unacceptable

5.2 Speaker-lock checklist

- Identify the last unambiguous speaker before the disputed segment.
- Check whether the segment grammatically answers the preceding question or continues the prior speaker’s sentence.
- Inspect vocatives (“Mr. X,” “Hasan,” “Piers”), second-person address, and first-person self-reference.
- Use topic continuity only as supporting evidence, never as the sole basis.
- If video exists, inspect mouth movement, camera cuts, waveform changes, or platform speaker labels.
- If audio exists, listen through at least several seconds before and after the boundary.
- For ASR captions, assume that punctuation and line breaks are not reliable turn boundaries.
- If two interpretations remain plausible, split the segment or mark attribution uncertain.
- Assign attribution confidence separately from factual verdict confidence.

MANDATORY HOLD CONDITION

If a materially consequential claim could plausibly belong to more than one speaker and the available evidence cannot resolve the turn, classify the segment as HOLD - SPEAKER ATTRIBUTION / AUDIO REQUIRED. Do not guess.

5.3 Recommended confidence labels

Label	Use when
HIGH	Multiple convergent cues or direct audio/video establish the speaker; a reasonable alternative attribution is remote.
MODERATE	Attribution is more likely than not and supported by turn structure, but one important verification channel is missing.
LOW	Attribution is tentative; use only for internal triage, not publication-level accusation.
HOLD	Speaker identity is unresolved enough that rating the claim would risk misattribution.

6. Quotation and context preservation

Layer / rule	Requirement
Verbatim layer	Store the exact transcript string as captured, including awkward grammar. This is the audit anchor.
Audio-corrected layer	If the transcript is wrong and audio verifies the correction, record the corrected wording separately and identify that it is audio-verified.
Publication quote	Use the shortest quote that preserves the proposition and relevant qualifiers. Ellipses must not alter meaning.
Context window	Preserve enough preceding/following material to resolve pronouns, conditions, concessions, sarcasm, quotation of others, and speaker changes.
No silent repair	Do not silently replace “could” with “did,” singular with plural, “about” with an exact number, or garbled proper nouns with guessed names.
Quoted third parties	If a speaker quotes another person, route the factual claim and attribution separately: “X accurately quoted Y” is distinct from “Y’s underlying statement is true.”

7. Claim extraction and atomic decomposition

The most important transformation is from natural language to atomic propositions without changing meaning. The system should first preserve the full utterance, then decompose it into propositions that could in principle receive different verdicts.

ATOMICITY TEST

If one part of the sentence could be true while another part is false, the sentence contains more than one atomic claim.

Compound statement	Atomic decomposition
“Country X invaded Y because it wanted oil and killed 10,000 civilians.”	(1) X invaded Y. (2) The invasion was motivated by oil. (3) 10,000 civilians were killed. (4) The deaths were caused by the invasion. Each requires different evidence.
“There was never a war; it was genocide.”	(1) The situation does not meet a relevant definition of war/armed conflict. (2) The conduct constitutes genocide. (3) War and genocide are mutually exclusive categories, if that implication is asserted.

Compound statement	Atomic decomposition
“Polls show 80% support.”	(1) A specific poll exists. (2) The relevant population is the one implied by the speaker. (3) The exact question measures the concept claimed. (4) 80% is numerically correct. (5) The poll is representative enough for the inference.

8. Claim taxonomy and rateability

Claim type	Treatment
Descriptive factual	Who/what/when/where; directly checkable.
Quantitative	Counts, rates, percentages, rankings, trends, comparisons.
Historical	Claims about past events, motives, chronology, continuity, terminology.
Causal	X caused Y; requires more than correlation or sequence.
Intent/motive	Claims about purpose or mental state; require statements, documents, pattern evidence, or inferential justification.
Legal	Claims about legal status, crimes, court findings, jurisdiction, treaty obligations.
Scientific/medical	Empirical claims requiring domain-appropriate evidence hierarchies.
Polling/public opinion	Claims dependent on question wording, sampling, universe, field dates, and subgroup interpretation.
Quotation/attribution	Whether a named person actually said or wrote something.
Prediction	Future-oriented; generally not rateable until outcome window passes, though premise claims may be.
Normative	What should be, good/bad, just/unjust; not directly fact-checkable, but embedded factual premises may be.
Rhetorical/analogical	Hyperbole, metaphor, analogy; test only the factual premises and literal claims actually asserted.

9. Research design and search protocol

Research begins with a claim-specific evidence plan. Do not browse randomly until a persuasive story emerges. State what evidence would support the claim, what evidence would refute it, and what evidence would leave it unresolved.

9.1 Pre-search specification

- Write the atomic proposition in neutral language.
- Define ambiguous terms operationally before searching (“war,” “ethnic cleansing,” “majority,” “ordered,” “caused,” etc.).
- Fix the relevant time window and geographic/population scope.
- Identify the best possible primary record if one exists.
- List at least one plausible alternative explanation.
- Specify what would count as disconfirming evidence.
- Record whether the claim is about what was known at the time or what is known now.

9.2 Retrieval sequence

Order	Target	Purpose
1	Primary record	Original video/audio; official document; statute; judgment; dataset; contemporaneous statement; underlying poll questionnaire; paper.
2	Methodological source	Codebook, technical appendix, methods section, court docket, sampling notes, dataset documentation.
3	Independent corroboration	High-quality reporting or scholarship not merely copying the same originating claim.
4	Counterevidence	Strongest credible source challenging the claim or offering a competing interpretation.
5	Context sources	Background needed to interpret terminology, chronology, or institutional process.
6	Claimant source	If available, identify what the speaker relied on and whether they later clarified/corrected the statement.

SEARCH DISCIPLINE

Search queries should be generated from both the claim and its negation. A fact-check that searches only for confirming phrasing is vulnerable to search-engine confirmation bias.

10. Source hierarchy and evidence-quality assessment

“Primary source” is not synonymous with “true,” and “secondary source” is not synonymous with “weak.” A perpetrator’s statement can be primary but self-serving; a peer-reviewed archival synthesis can be secondary but methodologically superior for a historical inference. Rank sources by fitness for the specific proposition.

Dimension	Question
Directness	Does the source directly observe/measure the proposition or merely infer/repeat it?
Proximity	How close in time and causal chain is the source to the event?
Independence	Is this genuinely independent corroboration or derivative reporting?
Expertise	Does the source have relevant domain competence?
Method transparency	Can methods, sample, definitions, and data provenance be inspected?
Bias/incentives	What institutional, political, financial, reputational, or litigation incentives might distort reporting?
Completeness	Is the source full-length or excerpted? Does it disclose limitations and contrary evidence?
Replicability	Can another reviewer inspect the same record or recompute the result?
Specificity	Does the source support the exact proposition rather than a nearby weaker claim?
Recency/version	For changing facts, is this the current authoritative version?

10.1 Practical source tiers

Tier	Type	Examples / constraints
Tier A	Direct / authoritative record	Original media; original dataset; statute/treaty; court judgment/docket; official contemporaneous record; peer-reviewed study reporting the underlying data.
Tier B	High-quality independent synthesis	Systematic reviews; major scholarly monographs; respected investigative reporting with transparent sourcing; institutional technical reports.
Tier C	Competent secondary reporting	Reputable news summaries, expert explainers, professional reference works.
Tier D	Advocacy / partisan / interested sources	Useful when transparent and evidentially specific; require corroboration and explicit interest disclosure.
Tier E	Unverified social media / anonymous / tertiary aggregation	Lead-generating only unless uniquely evidentiary; do not use as sole basis for consequential findings.

10.2 Independence firewall

Five news stories repeating the same wire report are one evidentiary lineage, not five independent confirmations. The ledger should include an “origin chain” field when multiple sources may derive from the same witness, institution, press release, or dataset.

11. Evidence synthesis and burden of proof

For each atomic claim, construct an explicit matrix rather than a narrative first. Narrative prose should be the output of the matrix, not a substitute for it.

Matrix field	Required content
Evidence supporting claim	Strongest pro-claim items, including claimant’s best source.
Evidence undermining claim	Strongest contradictory, qualifying, or disconfirming items.
Source quality concerns	Bias, indirectness, missing data, disputed authenticity, methodology, conflicts.
Scope fit	Does evidence concern the same date, population, geography, legal standard, or definition?
Residual uncertainty	What important unknowns remain?
Alternative explanations	Competing interpretations consistent with the same evidence.
Burden threshold	What level of evidence is appropriate for the strength of the accusation or factual claim?

11.1 Certainty domains adapted from evidence synthesis

Pangburn Analysis can borrow the logic - not mechanically the scoring - of evidence-grading frameworks such as GRADE: inspect risk of bias, inconsistency, indirectness, imprecision, and missing/selective evidence before increasing confidence. These domains are adapted to heterogeneous fact-checking records rather than clinical effect estimates.

- Risk of bias: are the key records produced under incentives that could systematically distort them?
- Inconsistency: do high-quality sources materially disagree, and is the disagreement explained?

- Indirectness: does the evidence answer the exact proposition, or only an adjacent one?
- Imprecision: are dates, counts, ranges, speaker identities, or causal inferences too uncertain for categorical wording?
- Missing/selective evidence: could unavailable records, selection effects, or publication practices materially change the conclusion?

12. Verdict ontology and confidence

Verdict	Definition
SUPPORTED	The material proposition is well-supported by the best available evidence; important qualifiers are preserved.
QUALIFIED / MISLEADING	A factual core is supported, but wording, scope, number, causality, universality, attribution, or omitted context materially changes the reader's likely understanding.
UNSUPPORTED	The available evidence does not justify the proposition at the claimed level of certainty. This does not mean the proposition has been disproven.
FALSE	The proposition is contradicted by sufficiently strong evidence or fails a clear definitional/numerical test.
DISPUTED	Credible, methodologically serious sources support materially different conclusions and the record does not justify collapsing the dispute.
HOLD / UNVERIFIABLE	Essential evidence is unavailable or source ambiguity prevents a responsible verdict.
NON-RATEABLE	Normative judgment, rhetorical flourish, subjective preference, or prediction lacking a presently testable factual proposition.

12.1 Confidence

Confidence	Meaning
HIGH	Multiple high-quality, direct, mutually consistent sources or a decisive authoritative record; minimal unresolved scope/attribution issues.
MODERATE	Conclusion is better supported than alternatives but important limitations or indirectness remain.
LOW	Tentative classification for working purposes; publication should usually prefer HOLD or a narrower proposition.

DO NOT COLLAPSE VERDICT AND CONFIDENCE

"Unsupported (High confidence)" means the record clearly fails to support the claim. "False (Low confidence)" is usually a warning that the proposition may need to be narrowed or moved to HOLD.

13. Special protocols

13.1 Quantitative claims

- Write the numerator and denominator explicitly.
- Record unit, geography, population, time period, inclusion/exclusion criteria, revision status, and whether the figure is provisional.
- Recompute percentages from raw counts when possible.

- Distinguish stock from flow, people from cases, reported from verified, gross from net, nominal from real, and point estimate from interval.
- If multiple data systems disagree, explain definitional differences before selecting one.
- Never cite a rounded early estimate as if it were the final revised count without temporal qualification.

13.2 Polling claims

- Obtain the original questionnaire and exact wording.
- Identify target population, sample frame, sample size, field dates, mode, weighting, response options, order effects, and subgroup n.
- Do not convert “desirable but impractical” into “supports forced implementation” unless wording justifies it.
- Distinguish all adults from citizens, voters, party supporters, ethnic/religious subgroups, or likely voters.
- Treat single-poll results as time-bounded measurements, not permanent social facts.
- Report uncertainty and question-wording sensitivity when relevant.

13.3 Legal claims

Stage	Do not confuse with
Allegation / accusation	A party, expert, NGO, prosecutor, or commentator alleges conduct.
Investigative finding	A commission or investigator finds evidence under a stated evidentiary standard.
Provisional measure	A court orders temporary measures without deciding final merits.
Arrest warrant / charge / indictment	A tribunal finds the threshold for process, not guilt.
Merits judgment	A court resolves the legal claim under the applicable standard.
Conviction / final judgment	Adjudicated liability or guilt, subject to appeal/finality rules.

13.4 Historical claims

- Separate contemporaneous primary records from later memoir, oral history, local tradition, and retrospective scholarship.
- Record whether the claim is about occurrence, scale, intent, planning, or interpretation. These often have different evidentiary bases.
- Do not treat absence from one primary source as proof of non-occurrence unless the source would reasonably be expected to record it.
- Trace sensational details to their earliest available source; later repetition is not independent corroboration.
- Use current scholarship to map genuine historiographical disputes rather than declaring consensus from one historian.

13.5 Causal and motive claims

- Establish temporal order, mechanism, and plausible counterfactual.
- Identify confounders and alternative motives.
- Distinguish documented stated purpose from inferred intent.
- Patterns can support intent, but the report must identify the inferential bridge.
- Avoid “because” when evidence supports only “after,” “associated with,” or “consistent with.”

13.6 Medical and scientific claims

Use domain-specific evidence hierarchies. Prefer systematic reviews, consensus guidelines, high-quality controlled studies, and authoritative surveillance data where appropriate. Distinguish mechanistic plausibility from demonstrated clinical outcomes. In high-stakes health contexts, current professional guidance and primary literature should be checked, and uncertainty should be explicit.

14. Denominator and provenance firewall

The denominator firewall prevents the most dangerous statistical error in media misinformation audits: converting a curated list of interesting claims into a rate that appears to describe the entire speaker, video, or corpus.

A valid misinformation percentage requires all of the following:

- A defined population of eligible claims (e.g., every independently testable factual assertion in the full video).
- A documented inclusion/exclusion rule for opinions, repetition, questions, quotations of third parties, corrections, and compound claims.
- Speaker provenance for every included claim.
- A duplicate policy: repeated identical claims counted once or per occurrence, declared in advance.
- A classification rule defining which verdicts count as “misinformation.”
- A complete or scientifically sampled denominator, not a convenience set selected because claims looked suspicious.
- An audit showing that every numerator item belongs to the denominator population.
- Uncertainty around sampling-based estimates where a sample rather than census is used.

PUBLICATION PROHIBITION

If these denominator conditions are not met, report counts within the audited claim set only. Do not publish “X% of the speaker’s claims were misinformation” or imply a speaker-wide rate.

15. Multi-pass audit and independent replication

Audit pass	Question
Pass A - Source integrity	Is the source authentic, complete, correctly dated, and correctly transcribed?
Pass B - Speaker attribution	Who said each material statement? Are merged turns resolved?
Pass C - Atomic claim integrity	Did decomposition preserve meaning? Were qualifiers or negations lost?
Pass D - Evidence completeness	Was the strongest evidence for and against retrieved? Are sources independent?
Pass E - Citation entailment	Does each source actually support the sentence and strength of wording?
Pass F - Verdict calibration	Does the classification match the evidentiary record? Could a narrower correction be more accurate?
Pass G - Quantitative audit	Recompute all arithmetic; inspect denominator, sample, dates, units, and revisions.
Pass H - Adversarial review	Assume the verdict is wrong and build the best case against it.
Pass I - Consistency audit	Are similar claims by opposing speakers treated with the same evidentiary threshold?
Pass J - Publication audit	Check exact quotes, timestamps, page references, legal posture, references, and limitations.

15.1 Independent LLM replication

For high-impact reports, run at least two independent model sessions with no shared draft verdict. Give both the same source packet and protocol. Compare speaker attribution, claim boundaries, retrieved evidence, verdicts, and confidence. Disagreements are not averaged away; they are investigated as error signals.

Agreement metrics

For large corpora, operators may compute raw agreement and an inter-rater statistic such as Cohen's kappa or Krippendorff's alpha. These are process diagnostics, not substitutes for adjudication. High agreement can reflect shared bias; low agreement can identify ambiguous ontology or evidence gaps.

15.2 Human review requirements

The protocol is designed to make LLM work auditable, not to eliminate human judgment. Human or expert review is especially important when the source audio is ambiguous, a claim could cause serious reputational harm, legal terminology is dispositive, a medical conclusion could influence treatment, a statistical result depends on non-obvious methodology, or key sources materially disagree.

- A human reviewer should listen to disputed audio rather than asking the language model to infer indefinitely from text.
- Subject-matter experts should be consulted when correct interpretation depends on specialized methodology rather than general reading comprehension.
- The final editor should verify that report headlines and summaries do not become stronger than the underlying claim entries.
- No model should be treated as an independent corroborating source for a factual proposition; models are analytical tools, not witnesses.

16. Publication format and correction policy

Each published claim entry should be independently auditable. A reader should not have to trust the report's narrative summary.

Field	Purpose
C-ID	Stable claim identifier.
Timestamp	Exact or best available time range.
Source page	Transcript page/image locator.
Speaker	Name + attribution confidence.
Exact quote	Verbatim or explicitly audio-corrected text.
Atomic claim	Neutral proposition being tested.
Claim type	Quantitative, historical, legal, causal, etc.
Verdict	Supported / Qualified / Unsupported / False / Disputed / Hold / Non-rateable.
Verdict confidence	High / Moderate / Low.
Evidence for	Strongest support.
Evidence against	Strongest undermining evidence.
Why it matters	Specific defect: number, scope, attribution, causal leap, omitted context, etc.
Correction	Publication-safe statement closest to the evidence.
Sources	Named primary and independent sources.
Video check	Any required audio/visual verification before publication.
Audit status	Attribution / citation / arithmetic / adversarial passes completed.

16.1 Corrections

- Maintain version numbers and dates.
- Never overwrite a material verdict change without a correction note.

- Identify whether the error was attribution, transcription, sourcing, arithmetic, interpretation, or new evidence.
- Recompute any downstream statistics after a correction.
- If a corrected claim appeared in summaries, graphics, or social posts, correct those derivative artifacts too.

17. The master LLM system prompt

The following prompt is intended to be pasted into the highest-priority instruction layer available in the operator's LLM environment. It should be used together with access to web/search tools or an evidence packet. Tool names are deliberately generic so the prompt is portable.

```
PANGBURN ANALYSIS - FORENSIC FACT-CHECKING SYSTEM PROMPT v1.0

ROLE
You are a forensic fact-checking research system. Your highest objective is not speed, persuasion, ideological
balance, or conversational smoothness. Your objective is to minimize factual error and unjustified certainty while
producing an auditable path from source material to conclusion.

EPISTEMIC CONSTITUTION
1. Treat truth as correspondence between propositions and the best available evidence.
2. Operate under fallibilism. Every conclusion is revisable.
3. Calibrate language to evidence. Never express greater certainty than the record supports.
4. Apply identical evidentiary standards regardless of speaker, ideology, nationality, religion, or political side.
5. Distinguish fact, inference, allegation, interpretation, opinion, rhetoric, prediction, and legal judgment.
6. Never silently fill a gap in the source. Preserve ambiguity.
7. Prefer HOLD / UNRESOLVED over a guessed attribution or verdict.
8. Unsupported is not synonymous with false.
9. A citation is not evidence unless it entails the specific proposition at the strength stated.
10. Never compute a misinformation rate without a defensible denominator and speaker-provenance audit.

THREAT MODEL
Assume you are vulnerable to: confabulation; fabricated citations; citation-topic matching without entailment;
wrong-speaker attribution; transcript/OCR errors; quote normalization; claim mutation; scope drift; temporal
contamination; source-count inflation; authority bias; political motivated reasoning; false balance; denominator
errors; legal-status collapse; and arithmetic mistakes. Design every stage to catch these failures.

SOURCE INTAKE
Before rating claims, characterize the source:
- original media or derivative;
- full item or excerpt;
- publication/capture date;
- transcript type: human, ASR, OCR, translation, screenshot, unknown;
- speaker labels present/absent;
- timestamp quality;
- audio/video availability;
- known edits or missing sections.
If the transcript is automatic or image-derived, do not assume punctuation, line breaks, spelling, or turn
boundaries are reliable.

SPEAKER ATTRIBUTION LOCK
For every consequential claim:
A. identify the last unambiguous speaker;
B. inspect question-answer structure, grammar, vocatives, self-reference, interruptions, and topic continuity;
C. use audio/video when available;
D. do not infer speaker from ideology or topic alone;
E. assign attribution confidence HIGH / MODERATE / LOW / HOLD.
If a claim could plausibly belong to more than one person and the record cannot resolve it, mark HOLD - SPEAKER
ATTRIBUTION / AUDIO REQUIRED. Do not fact-check it as belonging to either person.

QUOTE FIDELITY
Maintain three layers when needed:
1. raw transcript string;
2. audio-verified corrected transcript;
3. publication excerpt.
Never present a cleaned paraphrase as a verbatim quote. Preserve qualifiers, negations, conditionals,
approximations, and relevant context. Mark uncertain words.

ATOMIC CLAIM EXTRACTION
Split compound statements into independently testable propositions. If one component could be true while another is
false, create separate claim IDs. Preserve the complete original utterance as context.
Label each claim: descriptive, quantitative, historical, causal, motive/intent, legal, scientific/medical, polling,
quotation/attribution, prediction, normative, rhetorical/analogical.
Do not rate purely normative propositions; extract and rate embedded factual premises if present.

PRE-SEARCH SPECIFICATION
Before searching, write:
- exact atomic proposition;
- operational definitions of ambiguous terms;
```

```
- relevant time window;
- population/geographic scope;
- what evidence would support it;
- what evidence would refute or materially qualify it;
- plausible alternative explanation(s);
- whether the claim concerns what was known THEN or what is known NOW.

RESEARCH ORDER
1. Find the best primary/direct record.
2. Find methodology or technical documentation for that record.
3. Find independent high-quality corroboration.
4. Deliberately search for the strongest credible counterevidence.
5. Find context needed to resolve definitions, chronology, or legal posture.
6. If possible, identify the claimant's own source or later clarification/correction.
Search both the claim and its negation. Do not stop when you find confirming evidence.

SOURCE EVALUATION
Evaluate each source for:
- directness;
- temporal/causal proximity;
- independence;
- relevant expertise;
- methodological transparency;
- incentives/conflicts;
- completeness;
- replicability;
- specificity to the exact proposition;
- version/recency.
Do not count derivative articles as independent confirmations. Trace source lineage.
Do not assume official = true or activist = false. Evaluate evidence quality proposition by proposition.

EVIDENCE MATRIX
For each atomic claim, create:
A. strongest evidence supporting the claim;
B. strongest evidence undermining or qualifying it;
C. source-quality limitations;
D. scope/definition fit;
E. unresolved uncertainties;
F. plausible alternatives;
G. appropriate burden of proof for the strength of the allegation.
Draft the narrative only after this matrix exists.

VERDICT ONTOLOGY
SUPPORTED: material proposition is well supported with important qualifiers preserved.
QUALIFIED / MISLEADING: factual core exists, but number, scope, causal force, universality, attribution, or omitted
context materially alters understanding.
UNSUPPORTED: evidence does not justify the proposition at the stated certainty. Do not imply disproven.
FALSE: strong evidence contradicts the proposition or it fails a clear definitional/numerical test.
DISPUTED: credible high-quality sources support materially different conclusions and the record does not justify
collapsing the dispute.
HOLD / UNVERIFIABLE: essential evidence, attribution, or source fidelity is insufficient.
NON-RATEABLE: normative, rhetorical, subjective, or presently untestable proposition.

CONFIDENCE
Assign verdict confidence separately: HIGH / MODERATE / LOW. Explain the principal reason confidence is not higher.

QUANTITATIVE FIREWALL
For every number: specify numerator, denominator, unit, population, geography, time window, source revision status,
and exclusions. Recompute arithmetic. Distinguish estimates from verified counts. Never project an early estimate
into a later period without qualification.

POLLING FIREWALL
Obtain original questionnaire when possible. Record target population, n, field dates, mode, weighting, subgroup n,
exact wording, response options, and question context. Do not substitute a broader concept for what the poll
actually asked.

LEGAL FIREWALL
Distinguish allegation, investigative finding, provisional measure, warrant/charge, merits judgment,
conviction/final judgment. Quote the relevant legal standard. Do not say a court 'found' a matter that it merely
treated as plausible, within jurisdiction, or sufficient for provisional action.

HISTORICAL FIREWALL
Separate contemporaneous records from later memoir/oral history/secondary synthesis. Trace sensational details to
their earliest source. Distinguish occurrence, scale, planning, motive, and interpretation. Map genuine
historiographical disagreement.

CAUSAL/INTENT FIREWALL
Do not infer causation from sequence. Evaluate mechanism, temporal order, confounding, alternatives, and
counterfactual. Separate stated purpose from inferred intent. If intent is inferred from pattern, state the
inferential bridge.
```

```
DENOMINATOR / PROVENANCE FIREWALL
Do not publish a speaker-wide or video-wide misinformation percentage unless you have:
- defined the complete eligible claim population or a valid sample;
- documented inclusion/exclusion rules;
- audited speaker provenance;
- declared duplicate/repetition policy;
- declared which verdicts count as misinformation;
- verified every numerator item belongs to the denominator.
Otherwise report only counts within the audited set.

MANDATORY AUDIT PASSES
Run separate passes for:
1. source integrity;
2. speaker attribution;
3. atomic claim fidelity;
4. evidence completeness;
5. citation entailment;
6. verdict calibration;
7. arithmetic/denominator verification;
8. adversarial review trying to prove the verdict wrong;
9. symmetry/consistency across speakers;
10. publication details (quote, timestamp, page, source, legal posture).

CITATION AUDIT
For every externally verifiable sentence in the final report, ask:
- Which exact source passage supports it?
- Does the source support the same subject, predicate, date, scope, quantity, and certainty?
- Is the citation primary when a primary record is reasonably available?
- Is this source independent of other cited sources?
If not, revise the sentence or citation.

OUTPUT FOR EACH CLAIM
C-ID | timestamp | transcript page/source locator | speaker | attribution confidence | exact quote | context |
atomic claim | claim type | verdict | verdict confidence | evidence for | evidence against | source-quality notes |
why problematic (if applicable) | correction | sources | video/audio check | audit status.

WRITING RULES
- Use restrained academic language.
- Do not accuse a person of lying unless evidence establishes knowing deception; factual error is not equivalent to
dishonesty.
- Preserve competing credible interpretations.
- State limitations prominently.
- Do not hide uncertainty in footnotes while writing categorical headlines.
- Use “according to [source]” when describing a contested source-specific finding.
- Use “the evidence supports” rather than “obviously,” “clearly,” or “everyone knows,” unless the proposition is
genuinely elementary and undisputed.

STOPPING RULE
Do not stop merely because enough evidence exists to defend a preferred verdict. Stop when:
- the best available primary evidence has been located or its absence documented;
- strong counterevidence has been actively sought;
- major source conflicts are understood;
- further searching is unlikely to change the classification materially;
- the remaining uncertainty is explicitly stated.

FINAL ADVERSARIAL QUESTION
Before publication ask: “What is the strongest fair-minded argument that this verdict is wrong, and have I
represented it accurately?” If you cannot answer, the audit is incomplete.
```

18. Phase-specific prompts

A. Transcript intake / attribution prompt

```
You are the attribution auditor. Do NOT fact-check content yet. Reconstruct speaker turns from the supplied
transcript/media. For every segment with a factual claim, output timestamp, raw text, proposed speaker,
attribution confidence, and the concrete cues supporting attribution. Treat ASR punctuation and line breaks
as unreliable. If two speakers are plausible, mark HOLD - AUDIO REQUIRED. Do not infer speaker from
ideology.
```

B. Atomic claim extraction prompt

```
Using only the speaker-locked transcript, extract every independently testable factual proposition. Preserve the full original quote and timestamp. Split compound statements whenever components could receive different verdicts. Label normative/rhetorical material NON-RATEABLE but extract embedded empirical premises. Do not research yet.
```

C. Research-plan prompt

```
For each atomic claim, write a neutral proposition, definitions, scope, time window, best possible primary source, strongest plausible disconfirming evidence, and search queries for both the claim and its negation. Identify special protocols required: quantitative, polling, legal, historical, causal, medical, quotation.
```

D. Evidence-for / evidence-against prompt

```
Research this claim adversarially. First build the strongest evidence that supports it. Then independently build the strongest evidence that undermines or qualifies it. Trace source lineage so derivative reports are not treated as independent. Prefer original documents/datasets/audio. Return an evidence matrix; do not issue a verdict until both sides are populated.
```

E. Citation-entailment audit prompt

```
Ignore whether you agree with the conclusion. For every sentence and citation, determine whether the cited source entails the same subject, predicate, date, scope, number, and degree of certainty. Flag topic-only citations, source laundering, unsupported causal language, overgeneralized populations, and missing qualifiers. Recommend narrower wording where needed.
```

F. Adversarial verdict audit prompt

```
Assume the current verdict is wrong. Build the strongest evidence-based case against it. Identify missing sources, definitional alternatives, contradictory primary records, later corrections, denominator errors, speaker-attribution uncertainty, or asymmetric standards. Then state whether the verdict survives unchanged, should be narrowed, downgraded, upgraded, or moved to HOLD.
```

G. Final compilation prompt

```
Compile the audited claims into a publication-ready Pangburn Analysis report. Preserve exact quotes, timestamps, speaker attribution confidence, verdict confidence, strongest evidence for and against, correction, and sources. Do not compute speaker-wide misinformation percentages unless the denominator/provenance firewall has passed. Include limitations and unresolved holds.
```

19. Structured schemas and templates

19.1 Minimal machine-readable claim record

```
{
  "claim_id": "C-001",
  "source_id": "video_or_document_id",
  "timestamp_start": "00:12:34",
  "timestamp_end": "00:12:52",
  "source_page": 7,
  "speaker": "Name",
  "speaker_confidence": "HIGH|MODERATE|LOW|HOLD",
  "raw_quote": "...",
  "audio_verified_quote": null,
  "context_before": "...",
```

```
  "context_after": "...",
  "atomic_claim": "...",
  "claim_type": "historical",
  "definitions": ["..."],
  "time_scope": "...",
  "geographic_scope": "...",
  "verdict": "SUPPORTED|QUALIFIED|UNSUPPORTED|FALSE|DISPUTED|HOLD|NON-RATEABLE",
  "verdict_confidence": "HIGH|MODERATE|LOW",
  "evidence_for": [{"source":"...","passage":"...","quality_notes":"..."}],
  "evidence_against": [{"source":"...","passage":"...","quality_notes":"..."}],
  "correction": "...",
  "citation_audit": "PASS|FAIL|PENDING",
  "attribution_audit": "PASS|FAIL|PENDING",
  "arithmetic_audit": "PASS|NA|FAIL|PENDING",
  "adversarial_audit": "PASS|FAIL|PENDING",
  "video_check": "...",
  "notes": "..."
}
```

19.2 Claim-ledger spreadsheet columns

Claim ID | Source ID | Timestamp start | Timestamp end | Transcript page | Speaker | Speaker confidence | Raw quote | Audio-verified quote | Context | Atomic claim | Claim type | Definitions | Scope | Verdict | Verdict confidence | Evidence for | Evidence against | Source lineage | Correction | Sources | Video check | Attribution audit | Citation audit | Arithmetic audit | Adversarial audit | Version | Notes

20. Quality gates and stopping rules

Gate	Pass condition
Gate 1 - Source	Original or best available source identified; source type and completeness documented.
Gate 2 - Attribution	Every consequential quote has speaker + confidence; low-confidence items not published as accusations.
Gate 3 - Quote	Exact wording and context verified; transcript corrections labeled.
Gate 4 - Atomicity	Compound claims decomposed; no verdict hides mixed truth values.
Gate 5 - Retrieval	Primary evidence sought; counterevidence deliberately searched.
Gate 6 - Evidence quality	Directness, independence, methods, incentives, scope, and uncertainty assessed.
Gate 7 - Citation	Every load-bearing citation entails its sentence.
Gate 8 - Quantitative	Arithmetic and denominators independently recomputed.
Gate 9 - Adversarial	Strongest fair-minded countercase documented.
Gate 10 - Symmetry	Equivalent claims receive equivalent standards.
Gate 11 - Statistics	No invalid misinformation rate or speaker-wide inference.
Gate 12 - Publication	Limitations, holds, legal posture, and correction path visible.

STOPPING RULE

A claim is ready when additional reasonable searching is unlikely to change its classification materially AND remaining uncertainty is explicit. “I found enough to support my verdict” is not a valid stopping rule.

21. Worked methodological examples

These examples illustrate process errors rather than serving as a comprehensive fact-check of the underlying political subjects.

21.1 Merged-turn transcript example

In image-based YouTube transcript captures, a page may contain an interviewer's question, a guest's answer, an interruption, and another answer without reliable speaker labels. Treating each page or caption line as one speaker is unsafe. The correct procedure is to anchor on unambiguous questions, identify direct answers, and use audio for consequential boundaries. If the transcript says "We never commit crimes" immediately after the interviewer asks about the guest's organization, the grammatical response structure is evidence for the guest attribution; the next interruption beginning "Sorry, I have to interrupt..." is evidence for the interviewer. This is still a turn-level inference and should be confirmed against audio when publication stakes are high.

21.2 Compound legal-rhetorical claim example

Statement: "There was never a war; it was genocide, the opposite of war." A rigorous audit should not assign one global verdict. It should separately test whether the hostilities meet a relevant academic/legal armed-conflict definition; whether the conduct meets a genocide standard under the specified authority or scholarly framework; and whether the two categories are logically or legally mutually exclusive. The correction should preserve any supported concern about genocide while rejecting an incorrect category exclusion if applicable.

21.3 Historical sensational-detail example

Statement: "Rebels beheaded babies." The protocol should trace "rebels killed children" and "multiple babies were beheaded" separately. A contemporaneous source may establish an infant's killing without specifying method, while a later local history may describe decapitation of older children. The correct verdict can therefore be "supported core, qualified detail" rather than treating the entire sentence as simply true or false. The source chain itself becomes part of the finding.

21.4 Percentage-denominator example

Suppose an audit identifies 8 false, 12 unsupported, and 21 qualified claims from 41 selected entries. Those numbers describe the audited set. They do not prove that 19.5% of all factual statements in the video are false, because the denominator may exclude supported claims, repetitions, unselected statements, or claims attributed to other speakers. The firewall requires a census or valid sampling frame before generalizing.

22. References and recommended reading

R1 International Fact-Checking Network (IFCN). Code of Principles: commitments on non-partisanship, source transparency, multiple sources, evidence for and against, methodology, and corrections. Source

R2 Full Fact. "How we fact check." Methodology covering claim selection, understanding the claim, gathering evidence, claimant contact, expert consultation, context, independent review, and correction. Source

R3 National Academies of Sciences, Engineering, and Medicine. Reproducibility and Replicability in Science. Washington, DC: National Academies Press, 2019. Source

R4 National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1, 2023. Source

R5 Autio C, Schwartz R, Dunietz J, et al. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1, 2024. Source

R6 Higgins JPT, Thomas J, Chandler J, et al., eds. Cochrane Handbook for Systematic Reviews of Interventions, version 6.5, 2024. Source

R7 Schünemann HJ, Higgins JPT, Vist GE, et al. Cochrane Handbook, Chapter 14: grading certainty of evidence using risk of bias, inconsistency, indirectness, imprecision, and publication bias. Source

R8 Aggazzotti C, Andrews N, Smith EA. "Can Authorship Attribution Models Distinguish Speakers in Speech Transcripts?" Transactions of the Association for Computational Linguistics 12 (2024): 875-891. Source

R9 Aggazzotti C, Wiesner M, Smith EA, Andrews N. "The Impact of Automatic Speech Transcription on Speaker Attribution." Transactions of the Association for Computational Linguistics 13 (2025): 1578-1596. Source

R10 Fadeeva E, et al. "Fact-Checking the Output of Large Language Models via Token-Level Uncertainty Quantification." Findings of ACL 2024. Source

R11 Cohen J. “A Coefficient of Agreement for Nominal Scales.” Educational and Psychological Measurement 20(1), 1960: 37-46. Source

R12 Krippendorff K. Content Analysis: An Introduction to Its Methodology. Sage. (For reliability concepts including Krippendorff’s alpha.) Source

Methodological note

This standard adapts ideas from professional fact-checking, evidence synthesis, reproducible research, and AI risk management. It does not claim endorsement by any referenced organization and should not be represented as an IFCN, Cochrane, GRADE, NIST, or National Academies standard.

Appendix A. One-page operator checklist

☐ Have I identified the original or best available source?
☐ Do I know whether the transcript is ASR, OCR, human, translated, partial, or edited?
☐ Is the speaker attribution locked for every consequential claim?
☐ Is the exact quote preserved before paraphrase?
☐ Have I split compound claims into atomic propositions?
☐ Have I defined ambiguous terms, scope, and time window?
☐ Did I search for the best primary record?
☐ Did I deliberately search for evidence against my emerging conclusion?
☐ Are cited sources genuinely independent?
☐ Does each citation entail the exact sentence and certainty level?
☐ Have I separated unsupported from false?
☐ Have I separated legal allegation/process from final adjudication?
☐ Have I recomputed every number and denominator?
☐ For polls, did I inspect the original wording and sample?
☐ For historical claims, did I trace sensational details to their earliest source?
☐ For causal/intent claims, did I examine alternatives and the inferential bridge?
☐ Have I run an attribution audit independent of the factual audit?
☐ Have I run an adversarial audit trying to defeat my verdict?
☐ Have I checked whether the same standard would be applied to an opposing speaker?
☐ Am I publishing only statistics justified by a valid denominator?
☐ Are uncertainty, holds, and limitations visible in the main text?
☐ Can a third party reconstruct the verdict from the claim ledger?

Appendix B. Full claim-ledger template

Use one record per atomic proposition. For long projects, maintain this as a spreadsheet or database and render only the relevant fields into the final report.

Field	Entry
C-ID	C-001
Source ID	Stable video/document identifier
Timestamp	00:00:00-00:00:00
Transcript page	p.__
Speaker	Name
Speaker confidence	HIGH / MODERATE / LOW / HOLD
Raw quote	Exact transcript string

Field	Entry
Audio-verified quote	If applicable
Context before/after	Enough to preserve meaning
Atomic claim	Neutral proposition
Claim type	Historical / quantitative / legal / etc.
Definitions	Operational meanings
Time scope	Relevant dates
Geographic/population scope	Relevant universe
Evidence for	Primary + independent support
Evidence against	Best counterevidence
Source lineage	Which sources are derivative?
Verdict	Classification
Verdict confidence	HIGH / MODERATE / LOW
Why problematic	Specific defect, if any
Correction	Publication-safe wording
Sources	Named sources
Video/audio check	Required verification
Attribution audit	PASS/PENDING/FAIL
Citation audit	PASS/PENDING/FAIL
Arithmetic audit	PASS/NA/PENDING/FAIL
Adversarial audit	PASS/PENDING/FAIL
Version notes	Changes and reasons

Appendix C. Audit log / reproducibility record

A reproducibility record need not expose private chain-of-thought. It should preserve externally inspectable research actions and decision inputs: source identifiers, searches, retrieved records, version dates, calculation steps, and reviewer decisions.

```
PROJECT: _____
SOURCE / VIDEO: _____
CAPTURE DATE: _____
MODEL(S) / VERSION(S): _____
TOOL ACCESS: web / local files / audio / video / code / other

SOURCE REGISTER
- canonical source:
- transcript origin:
- completeness:
- timestamp quality:
- speaker labels:
- known transformations:

RESEARCH LOG
[date/time] claim C-___ | query / source opened | result / relevance | retained? why?

DECISION LOG
C-___ | initial verdict | reviewer challenge | final verdict | reason for change/no change

CALCULATION LOG
C-___ | formula | numerator | denominator | source values | independent recomputation result
```

```
CORRECTION LOG
version | date | affected claims | type of error | old wording | new wording | downstream outputs updated
```

Closing standard

THE PANGBURN TEST

A fact-check is not rigorous because it contains many citations, a confident verdict, or a long explanation. It is rigorous when the exact speaker and claim are stable, the evidence chain is direct and independently testable, the strongest counterevidence has been confronted, the uncertainty is calibrated, the mathematics is reproducible, and a skeptical reader can audit the path from source to conclusion.