

UNITED WISDOM

Images and recordings as evidence in the age of synthetic media

World and Consciousness

Andrey Zarinsh & Maria Evilkut

W-06-N04 · EN

@andrey.zarinsh @ilkutki

Images and recordings as evidence in the age of synthetic media

Epigraph

> *A wise man, therefore, proportions his belief to the evidence.*

David Hume, *An Enquiry concerning Human Understanding*, X, §4, SBN 110; English original as displayed by Hume Texts Online. The point is proportionate confidence, rather than automatic rejection of unfamiliar events.

What this article is about

An image can feel like a direct window onto an event. We recognize a face, hear a voice, and almost unnoticed move from a recording to certainty. Yet the file, the event, and the story attached to it answer different questions. A genuine recording can carry a false caption; a synthetic reconstruction can illustrate something real while remaining unsuitable as a recording of that occurrence. This article asks what makes an image useful evidence, how provenance differs from truth, and why doubt also needs reasons. We bring Hume's account of testimony into a bounded comparison with contemporary approaches to media provenance and digital verification. Hypothetical cases then explore a disputed photograph, a voice message, and several reposts of one clip. The aim is language for a limited, revisable conclusion that respects the people represented. This is neither forensic training nor a promise of detecting fabricated media by sight.

I. Name the claim first

Consider a hypothetical scene. A neighbourhood group receives a photograph of a flooded street, captioned as water approaching the local school. One person demands an immediate warning; another notices unusually smooth reflections and declares the picture AI-generated. They argue about authenticity, although the practical question is narrower: is this particular street flooded now, and is the proposed route usable? A perfectly genuine photograph from another town cannot answer that question. A mistaken judgment about reflections does not establish that the route is safe either. Start by separating what appears in the image from the claim someone asks you to accept because of it.

This separation does not make vision worthless. It identifies where its present support ends. The photograph may show water, a building, and a car, while the place name, date, and cause come from an accompanying message. Even text visible inside the frame belongs to the

@andrey.zarinsh @ilkutki

material under examination; it is not automatically independent confirmation. Someone might have photographed a sign beside a similar building, recorded a demonstration, or attached the wrong date. A possible explanation is not evidence that it happened. Its value is in showing which further observation would distinguish the alternatives, instead of inviting an endless collection of suspicions.

In this scene a limited conclusion might be: ‘We have received a flood photograph; its location and time have not yet been confirmed.’ That wording leaves room to check current messages from the school or relevant local service. It neither requires accusing the sender nor turns missing knowledge into evidence of safety. Depending on the available circumstances, someone may temporarily choose another route. That is a precautionary decision, not acceptance of every detail in the caption. Different actions need different support. Avoiding a questionable stretch of road and asserting that a named person deliberately circulated a forgery cannot be justified by the same minimal information.

II. Evidence has a mediator

In section X of the Enquiry, Hume discusses eyewitness reports and the conditions under which we trust them. This eighteenth-century philosophical argument belongs to a controversy about miracles; it is not an experiment on contemporary platforms. Its useful distinction here is between the supposed fact and our reasons for believing a report of it. A report has no necessary connection with an event merely because someone delivers it confidently. Yet Hume does not tell us to dispense with testimony. He describes this kind of reasoning as common, useful, and necessary to human life. Preserving both sides matters more than turning a short epigraph into a slogan of total distrust.

A camera changes the form of mediation without eliminating it. Someone chooses its direction, the start of recording, the method of saving, and the moment of publication. Recipients often do not know what intermediate operations occurred. This is not an accusation of manipulation against everyone who records something. Framing is unavoidable even in conscientious work. The distinction is between limitation and deception. A shortened clip can adequately answer a narrow question, such as whether particular audible words occur in this recording. The same clip may poorly support a conclusion about the speaker’s intention when an earlier reply has been removed. Evidence remains valuable when its scope is established, rather than when completeness is assumed.

Mediators form a chain that can be described without invention. ‘A friend sent it to me’ establishes the latest route of receipt; it does not make that friend an eyewitness. ‘Ten accounts published it’ establishes circulation, not ten observation points. A reliable source can also be mistaken outside what it actually checked. Instead of the general label ‘I trust them’, ask which step this person witnessed, which file version they received, and how they checked

the caption. The conversation may feel slower than an ordinary argument, but it distributes responsibility more accurately. The sender can honestly say they merely forwarded a recording; the recipient need infer neither expertise nor intent. A new clarification changes the assessment of this particular message.

III. File integrity and event truth

Three levels are worth separating. Integrity concerns whether this particular file changed after a specified point. Provenance concerns the available history of its creation and movement. Support for a claim concerns how the material helps establish a particular circumstance. These questions interact, but none substitutes for the others. An invented scene created yesterday can be preserved perfectly. A reliable recording of an event can arrive with part of its technical history missing. The author of a photograph may be known while the cause of the recorded action remains uncertain. ‘Verified’ therefore needs a complement: what was checked, in which version, and in relation to which conclusion? Otherwise the word promises more than the work performed.

Suppose an investigator receives a clip and records its hash, a computed digital identifier of the particular bytes. A matching hash in a later comparison helps establish that the compared version has not changed; it does not itself certify that the depicted events happened before a camera. If the file was synthetic before receipt, it remains the same synthetic file. If an authentic image acquired a false caption, hashing the image does not correct that caption. Understanding this logical boundary does not require learning the calculation. A means of checking changes does not become a witness to the past. It protects the possibility of investigating the same object and discussing a reproducible check.

The Berkeley Protocol, developed for digital investigations of human rights violations, separates analysis of the source, the file, and its content. We borrow that distinction rather than transfer professional investigative duties into a casual group chat. If two people view different excerpts, their disagreement may concern different objects. Specifying duration, the start of a fragment, and the route of receipt can be more useful than adding another confident opinion. Public discussion does not require releasing the entire original, which may expose faces, addresses, or sensitive information. Verifiability matters together with appropriate access conditions, not through automatic disclosure of everything to anyone who asks.

IV. What provenance can add

Contemporary provenance systems seek to make a file's history inspectable. In the C2PA 2.4 specification, assertions about creation and changes are bound to content and a digital signature. Its scope is verifiable provenance assertions within a defined trust model; the standard does not decide whether an entire depicted story is true. That boundary is substantive, not a defect to conceal behind an attractive badge. For example, confirming use of a generation tool helps identify the kind of material but does not establish why its author chose the scene. Confirming a signer likewise does not establish every human inference attached to the file.

The distinction becomes clear in a hypothetical advertisement. A photographer honestly preserves capture information, after which a company places the image of a happy family beside a promise about its service. The picture's history may be transparent while the promise remains unsupported. A professional photograph really represents people at a moment of capture; it does not prove the advertised service caused their happiness. Provenance checking strengthens one part of an argument, while its causal bridge still needs separate support. Conversely, an openly synthetic illustration does not become fraudulent merely because it is not a photograph. What matters is whether it is presented as observation or clearly offered as an artistic representation.

Missing provenance information should be described literally. 'No verifiable history was found' is different from 'A forgery was found.' The file may have arrived as a screenshot or undergone conversion; what happened in this case still needs investigation. Yet missing history cannot become positive confirmation either: 'Since nothing has been disproved, everything is genuine.' In that interval it makes sense to seek another form of support for the claim that matters. If a caption announces a school closure, a current message from the school itself may be more relevant than further inspection of the picture. We examine a particular circumstance and potential action, rather than pursue a trust symbol for its own sake.

V. A detector reports the result of its method

A synthetic-content detector may be useful as an additional tool, but its output needs interpretation. The NIST AI 100-4 overview describes dependence on materials and conditions, including file transformations, and errors in both directions: synthetic content may be missed, and genuine content may be flagged. This is a technical overview, not our own evaluation of a particular service. We therefore provide neither a universal accuracy figure nor a recommended-product ranking. Learning that a program produced a high score is not the same as learning the probability of fabrication in this situation. The score's meaning, evaluation conditions, and fit to the material's domain matter.

A deliberately invented numerical example clarifies the issue without making a claim about an existing detector. Suppose ten of a thousand files are actually synthetic. An imaginary method flags eight of those ten and mistakenly flags ten genuine files as well. There are eighteen flags, eight of which concern synthetic material. A good ability to notice synthetic content in this hypothetical case does not turn each flag into a confident verdict. Change the collection’s composition, and the meaning of the same flag for a decision can change. The example explains the logic of prevalence and error. It does not assess the market, predict an individual check, or justify inserting these figures into a real dispute.

The cost of error matters too. If a program helps select items for further investigation, a false flag may mean extra work. If it is used for public accusation, the same mistake affects a person’s reputation. The tool’s label and the editorial decision should therefore remain distinct records. One can say, ‘This method returned this result; the context is still being checked.’ Without further support, that cannot become ‘The author knowingly lied.’ The phrase artificial intelligence does not make an inference impartial, just as human participation alone does not remove error. Combined checking serves a task; it is not a ritual that gives every result the appearance of final truth.

VI. Several publications are not several observations

Corroboration, comparing independent support, is valuable when the additional sources actually contribute information. Imagine three publications about one road incident: a news clip, a shortened messenger version, and a frame in a blog. If all derive from one original, they show its circulation. Their number does not establish three independent observations. This does not make the original false. It simply means its support has not tripled through repetition. The same principle applies to reports that ‘experts confirmed it’: find what the named experts checked and whether everyone is repeating a single original assessment. Independence is a relation between grounds, not a count of logos surrounding a message.

Another camera angle can add information, but it must also be connected to the event. Similar cars in two pictures are not enough to establish the same place and time. Matching stable building details, object positions, and movement sequences provides more definite grounds for comparison. This describes a logical task, not a promise to geolocate any photograph. Sometimes a quite different record matters: an event schedule, an organizer’s explanation, or the original complete excerpt. Additional material should distinguish competing explanations. Another emotionally compelling retelling of the same caption can increase engagement without increasing knowledge of what happened. The useful addition is one that changes what can reasonably be concluded.

Support sometimes conflicts. Then identify the area of disagreement instead of choosing a source by sympathy. One witness may describe when something began, another what they

saw on arrival; both could honestly have observed different portions of an event. Elsewhere the contradiction really concerns one circumstance, and ‘everyone has their own truth’ cannot resolve it. A location does not become two locations out of respect for narrators. Respect means allowing clarification without humiliation, preserving uncertainty, and not substituting majority pressure for examination. New material changes a particular conclusion. Keeping the earlier wording in a record makes clear which ground proved mistaken or incomplete, rather than quietly pretending the correction was always known.

VII. A familiar voice and a verified request

An audio recording poses a similar question in a more intimate form. Suppose a relative receives a message in a familiar voice asking for an unusual action immediately. Recognition may matter, but it does not exhaust verification of the request. Who initiated this contact? Does the request correspond to the person’s actual intention? Can it be confirmed through another previously known route? Establishing the technology used to create the recording need not come first. Checking human intention may be more useful than investigating acoustic origin. Calling through an already saved contact is one possible step in this hypothetical scene. Circumstances determine the response; this article does not prescribe every financial or emergency procedure.

Even a genuine voice can belong to a different context. The sentence might have been spoken for a performance, read from a page, extracted from an old conversation, or paired with someone else’s explanation. Synthetic origin does not automatically establish malicious intent either: a person may knowingly use voice synthesis for accessibility or creative work. Examine how the material is presented and what is being requested. A recording offered as a current conversation without disclosure of transformations raises one question. An openly announced artistic reconstruction raises another. The same sound does not have the same evidential value in every use; its value depends on the connection to a person and event that the sender asserts.

In public discussion of a recording, recognizable sound alone should not assign words to a person. Describe what is audible and separately state the support for voice attribution. ‘Resembles’, ‘confirmed’, and ‘not established’ name different states when their grounds are explained. Yet a cautious word should not conceal the circulation of an accusation: a headline saying ‘Perhaps he confessed’ can still launch a damaging rumour. Sometimes the responsible outcome is to withhold publication pending a material clarification. This is not a demand to remain silent about real danger. It is a choice of wording that conveys the supported concern without turning an unverified identity into established conduct by a named individual.

VIII. Doubt should not erase the witness

Synthetic media make it tempting to dismiss an inconvenient recording with a single sentence: ‘Anything can be fabricated now.’ That describes a general possibility; it does not examine this file. Applied equally to every item, it ensures that no evidence can change a prior belief. This is a logical trap regardless of how many fabrications circulate. Doubt remains meaningful when it identifies what is unestablished and what could help establish it. It becomes insulation from scrutiny when every possible confirmation is dismissed as another possible fabrication, leaving no circumstances under which the conclusion could change. Demands for evidence need intelligible grounds as much as assertions do.

Someone reporting harm should not have to become a technical laboratory merely to be heard. Listening, recording what was reported, and recognizing the need for investigation are distinct, compatible actions. Taking an account seriously does not by itself settle every question of responsibility. There is no need to choose between treating every account as established in advance and suspecting every narrator in advance. The relevant point here is that examination of an image should not quietly become a judgment of the witness’s human value. A poor-quality file can concern an event worth investigating. Authentic material can carry an inference needing clarification. These states do not describe a person’s dignity.

Within its professional domain, the Berkeley Protocol considers investigator and participant safety alongside handling of material. For an ordinary reader, a limited ethical question follows: who could be harmed by another public repost? A humiliating image need not be republished to explain doubts about it. A possible witness’s address need not be disclosed to demonstrate thoroughness. Reasonable access limits do not, however, permit an invented forensic result. One can honestly explain the type of check performed and why some material is withheld. Examination sustains the possibility of trust when its boundaries can be assessed and people do not become expendable objects in an argument.

IX. Conclusions need a route to correction

A verification result is useful through its form as well as its confidence. ‘Fake’ may refer to an invented scene, a wrong date, misidentified person, editing, satire, or merely unverified reporting. Such a label often leaves readers with stronger emotion and less precise knowledge. In a hypothetical case, an old clip can be genuine while its present caption is mistaken. The correction should identify that relation: ‘The recording belongs to an earlier event; the current caption is unsupported.’ Readers can then see what remains and what changes. Removing the entire recording or declaring everything about it true are both excessive responses to the narrower error actually found.

A good explanation states the object of the conclusion, its grounds, and the remaining boundary. For example: ‘An earlier publication of the same excerpt was found, so this video does not support the new caption date. The filming location was not separately checked.’ That limitation does not make the finding pointless. It prevents readers from extending a particular check to a question not investigated. Listing every imaginable unknown is unnecessary too; it can obscure the point. Include boundaries that could change understanding or action. When the claim concerns who spoke, attribution quality matters. When only a file’s publication date is being checked, a complete explanation of everyone’s intentions is outside the promised result.

Correction should reach the setting where the earlier conclusion had an effect. Someone who forwarded a wrongly captioned recording can send a clarification to the same group and say what changed. Recognizing an error does not require making the sender a villain; consistency does not require retaining a false message. Quietly replacing the caption as though no uncertainty existed deprives recipients of a chance to reconsider actions already taken. This is an editorial norm for responsible communication, not a measured effect of a particular technique. It follows from a simple relation: if earlier wording shaped someone’s understanding, the new grounds should allow that understanding to change. A record of the change makes correction intelligible rather than obstructing it.

X. Trust as an open practice

Trust in media need not depend on an infallible eye or machine. A more useful object of trust can be an intelligible practice: where material came from, which claim was checked, what was supported, what remained unknown, and how errors will be corrected. This is an editorial proposal about criteria, not an established law of social behaviour. It does not require every person to perform every check independently. On the contrary, division of labour is useful when a specialist or newsroom states the limits of its result and provides enough information for reasonable assessment. Trust then concerns a particular competence and procedure while preserving revision, rather than granting one name the authority to declare final truth.

A synthetic image can have an honest place within this practice. It can represent the impossible, explain a design, support an artistic narrative, or offer a reconstruction when its status is clear. The problem arises when the ability to represent is confused with a claim to have observed. Readers need to know whether they are offered evidence, illustration, a model, reconstruction, or an invented scene. These distinctions create no aesthetic hierarchy. An artistic image may be stronger than a photograph as art and weaker as support for a particular moment of capture. Keeping both evaluations respects the different tasks media can perform, instead of requiring every beautiful frame to record a real occurrence.

Return to the flooded street. Examination may reveal that the photograph is genuine but old, that the new warning is supported by an independent message, or that adequate information

remains unavailable. Each outcome changes the wording, but none makes the original question unnecessary. We wanted to understand a present circumstance and choose an action, not win a contest in guessing fabrications. The more precisely a claim is stated, the clearer it becomes which grounds add knowledge. Evidence does not lose its value in the age of synthetic media. It requires attention to the transition from image to conclusion, where a recognizable face and a digital signature alike cannot perform every necessary step on our behalf.

What this means in practice

For personal use, choose one non-intimate item without dangerous content and write the claim it is supposed to support. Do not begin with ‘real’ or ‘fake’. Separate what can be seen or heard from information supplied by the caption: location, time, voice attribution, or cause. Then name your actual route of receipt. If it was a forward, record that; uncertainty about the original author need not be filled with a guess.

Select one material unknown and consider which independent information could clarify it. This might be an original publication, fuller context, or current information from the person or organization concerned. Do not upload sensitive recordings to a random service out of curiosity. If examination requires professional access or creates risk for people, limit your own action and leave the question open. This exercise neither replaces expertise nor has an established effectiveness here.

Write the outcome in one paragraph: what was supported, on which grounds, and what it still does not establish. Keep a possible action separate from your confidence. Precaution may be warranted before a final conclusion; public accusation needs different support. If you have already circulated a wrong caption, send an intelligible correction to the same recipients. The last question is not ‘Did I guess the origin?’ but ‘Is my conclusion more precise, and can someone else see its grounds?’ An honest ‘not established yet’ can also be an outcome.

Key quotation

> *We frequently hesitate concerning the reports of others.*

David Hume, *An Enquiry concerning Human Understanding*, X, §6, SBN 112; displayed English original. Hume next describes weighing opposing circumstances. Hesitation leads into assessment of grounds; it does not make every account equally well supported.

Connections between traditions

1. **Hume: proportionate trust / C2PA: verifiable provenance.** Both invite a question about the grounds of confidence. Hume offers philosophical assessment of testimony in an eighteenth-century controversy; C2PA binds assertions to a file and signer technically. The standard neither implements Hume's entire epistemology nor establishes narrative truth.

2. **Hume: weighing testimony / Berkeley Protocol: source, file, and content analysis.** Their comparable question concerns the warranted move from a received account to a fact. The Protocol specifies professional work within a particular domain; philosophical reasoning does not replace that work or its safety requirements. This is a conceptual comparison, without a claim of historical borrowing.

3. **C2PA: provenance / NIST: synthetic-content detection.** These approaches can answer different parts of the media question. Inspectable history is distinct from a classifier's result. Agreement or conflict between signals requires contextual interpretation. Neither a signature nor a detector score by itself establishes a person's intention.

4. **Hume: proportionate inference / NIST: conditions of a detector result.** The comparable question concerns how strong a claim its grounds support. Philosophical assessment of testimony and technical error evaluation do different work: a historical argument computes no probability for a particular file, while a method's flag does not replace examining the claim. This is conceptual comparison, with no universal accuracy or shared mechanism asserted.

5. **C2PA: bound provenance / Berkeley Protocol: retaining a verifiable version.** Both approaches need to identify the material to which a check refers. Technical binding of assertions to a file and hash handling in professional work are not interchangeable operations. They help retain an object of examination, while source and content analysis remain separate; the professional procedure is not reducible to a digital signature.

Primary sources

1. David Hume. An Enquiry concerning Human Understanding, X, §§4–8. Primary philosophical account of testimony; displayed English original and labelled own Russian translations.

2. C2PA. Content Credentials, Technical Specification 2.4. Primary technical standard, §§1.2–1.3.1 and definitions: provenance-validation boundaries.

3. Chandra et al. NIST AI100-4, 2024. Technical overview, selected passages from §§3.2.2,4.2–4.3 on conditions and detection errors; not a new experiment by this article.

4. OHCHR / UC Berkeley. Berkeley Protocol, 2022. Professional methodological guidance; VI.C,VI.E and introductory safety provisions. Exact reading limits are retained in research_notes.

Tags

evidence, synthetic media, images, recordings, digital provenance, C2PA, NIST, Hume, trust, verification, captions, context, authenticity, file integrity, eyewitnesses, corroboration, source independence, voice, detectors, false flags, uncertainty, claims, correction, Berkeley Protocol, responsibility, human dignity, reconstruction, illustration

About the authors

Andrey Zarinsh is the author and founder of United Wisdom. He explores inner transformation, the unity of traditions, intercultural dialogue, and contemporary spiritual discipline. Instagram: @andrey.zarinsh.

Maria Evilkut is a co-author of United Wisdom. Alongside her research, she seeks to test these ideas in practice and observe how principles of unity and mutual understanding work in real life between people and cultures. Instagram: @ilkutki.