

Spotting the Synthetic

A Handbook for Pastoral Media Discernment

To cultivate pastoral prudence and critical media literacy in a digital environment where authentic, edited and synthetic media can all be used to tell the truth or to mislead.

The internet has entered a synthetic age. Artificial intelligence can now generate convincing text, images, voices and video with little technical skill. Yet the central pastoral challenge is wider than artificial intelligence. A genuine photograph can carry a false caption. A real speech can be cut so that it says the opposite of what the speaker intended. A cloned voice can be used to ask for money. A completely synthetic video can also be clearly labelled and used responsibly for education or art.

For this reason, the safest question is no longer simply, "Does this look fake?" The stronger question is, "What exactly is being claimed, and what evidence establishes that claim?" This handbook develops that habit. It combines the Church's teaching on truth and human dignity with practical methods for source checking, lateral searching, tracing media to its origin, examining visual and audio evidence, and responding pastorally when a claim is false or uncertain.

Scriptural refrain *Test everything; hold fast what is good* 1 Thessalonians 5:21

1) The Pastoral Responsibility for Truth

Media literacy is a pastoral responsibility because false information affects real people. It can injure a person's reputation, inflame prejudice, provoke communal tension, steal money, silence victims, distort faith and weaken trust in the Church. A future priest will not encounter these problems only as a private user of social media. He may receive messages in the name of a bishop, be asked to comment on an inflammatory clip, counsel a family harmed by impersonation, or decide whether a parish channel should circulate an urgent claim.

Christian discernment does not begin with universal suspicion. It begins with love of truth, respect for the person and willingness to test what we receive. The command to "test everything" does not end with doubt; it continues, "hold fast what is good." The purpose of verification is therefore not cynicism. It is responsible trust.

Christus Vivit and the Digital Environment

In *Christus Vivit*, Pope Francis recognises the digital world as a place of participation and pastoral possibility, while also warning about manipulation, exploitation, violence and isolation. He notes that closed information circuits can favour contact among people who already think alike and can facilitate fake news, prejudice and hatred. He also observes that the reputations of individuals can be endangered by summary trials conducted online, and that the Church and her pastors are not exempt from this harm (*Christus Vivit*, 86-90).

This teaching asks seminarians to resist two temptations. The first is passive consumption: allowing an algorithm to decide what deserves belief and attention. The second is impulsive judgement: condemning a person from a short clip before finding

the full recording and context. Pastoral prudence requires patience, proportion and a concern for the person whose dignity is at stake.

Artificial Intelligence and the Wisdom of the Heart

In his 2024 message for the World Day of Social Communications, Pope Francis describes the danger of "cognitive pollution": partially or completely false narratives are believed and broadcast as if they were true. He explicitly mentions deepfake images and audio messages that use a person's voice to make that person appear to say what was never said. The problem is not simulation in itself. Simulation becomes perverse when it distorts our relationship with other people and with reality.

The message also insists that algorithms are not neutral. They embody human choices and can be directed toward knowledge and encounter, or toward profit, polarisation and groupthink. Technical skill must therefore be joined to the wisdom of the heart: the capacity to see a person rather than merely a data point, target or viral opportunity.

Antiqua et Nova

Antiqua et Nova is a joint note of the Dicastery for the Doctrine of the Faith and the Dicastery for Culture and Education. It warns against treating artificial systems as if they possessed human intelligence, moral conscience or personal responsibility. It also addresses AI hallucinations, manipulated content, fraudulent impersonation and deepfakes. The document makes a crucial pastoral point: even when an image or recording is artificial, the injury to a person's dignity can be real (Antiqua et Nova, 84-89).

The note calls everyone involved in producing and sharing AI content to truthfulness, accuracy and careful verification. This responsibility includes the ordinary user who forwards a message. A person may not have created the deception, but an uncritical forward can increase its harm.

Magnifica Humanitas

Magnifica Humanitas places the subject within the Social Doctrine of the Church. Pope Leo XIV cautions that technical descriptions of AI can become outdated quickly and therefore calls for continuing scientific study together with moral and spiritual discernment (Magnifica Humanitas, 97-100). He later presents truth as a common good. Public truth requires verification, cross-checking of sources and responsible argumentation; it also grows through relationships of trust and honest exchange (132-140).

The encyclical does not ask Christian communities to idolise technology or to demonise it. It asks them to communicate transparently, pursue facts honestly, strengthen critical thought and educate people to use digital tools responsibly. This course answers that educational call in a practical way.

A Pastoral Synthesis

- Truth is a common good. It is not a weapon to be owned by the loudest voice.
- Human dignity comes before virality. Do not enlarge the humiliation of a victim merely to prove that a post is false.
- Prudence slows down judgement until adequate evidence is available.

- Charity corrects without ridiculing the person who was deceived.
- Justice requires timely action when synthetic media is used for fraud, harassment, sexual exploitation or incitement.

2) Understanding Information Disorder

The popular expression "fake news" is too vague for careful investigation. It can refer to an honest mistake, a deliberate lie, a real image with a false caption, a joke, a scam or an edited recording. The following distinctions help us diagnose what has happened and choose a proportionate response.

Type	Meaning	Pastoral example
Misinformation	False or misleading information shared without an intention to deceive.	A parishioner forwards an old flood photograph believing that it shows today's disaster.
Disinformation	False information deliberately created or distributed to deceive or cause harm.	A fabricated notice claims that the diocese has cancelled all Sunday Masses.
Malinformation	Information based on reality but used out of context or weaponised to harm.	A private pastoral message is leaked and framed so as to humiliate the sender.
Miscontextualised media	Authentic media is attached to a false date, place, person or explanation.	A genuine video from another country is shared as a local communal incident.
Cheapfake	Real media is deceptively cropped, slowed, reordered, captioned or selectively edited without advanced AI.	A homily is cut immediately before the sentence that explains its meaning.
Deepfake	A person's face, body or voice is created or altered with machine-learning techniques.	A cloned voice imitates the bishop and asks for an urgent transfer.
Fully synthetic media	The people, place, event, image, sound or video may have been generated rather than recorded.	A realistic video depicts an event that never occurred.
AI slop or synthetic persona	Large amounts of low-accountability synthetic content or a fabricated identity are produced for reach, money or influence.	A false spiritual adviser publishes daily AI-generated testimonies to sell products.
Satire or parody	Manipulated or synthetic media is used for humour or criticism without presenting itself as factual reporting.	A clearly labelled parody places invented words in a public figure's mouth.

The categories can overlap. A genuine image with a false caption is miscontextualised. If someone knowingly attaches the false caption to inflame hostility, it is also disinformation. If a synthetic voice asks for money, the same incident is both impersonation and fraud. The label helps us understand the method; evidence is still needed to establish what occurred.

The Three Layers of Every Claim

When you receive a suspicious post, separate three layers:

1. The media file. Is the image, audio or video recorded, edited, generated, or a mixture?
2. The context. Are the date, place, identity, sequence and surrounding circumstances accurate?
3. The claim. Does the evidence support what the caption, narrator or sender says happened?

These layers prevent the false choice between "real" and "fake." A file can be authentic while its caption is false. A file can be synthetic while its message is clearly labelled and educational. A file can be manipulated without changing the central meaning. The investigator's task is to state exactly what is known about each layer.

3) How Generative AI Produces Convincing Media

A basic understanding of generation removes some of the mystery. It also prevents exaggerated claims that an AI system simply "copies and pastes" a whole article or photograph from its training material. Modern systems learn statistical patterns from very large collections of data and use those patterns to generate new outputs in response to instructions. The result may resemble familiar human work even when no identical original exists.

Text Generation

Large language models generate text by estimating which sequence of words or tokens is likely to follow the prompt and the text already produced. They can imitate styles, organise explanations and produce fluent answers. Fluency is not the same as truth. A model can generate a confident statement, quotation, reference or name that is inaccurate because its task is to produce plausible language, not to bear witness as a human person does.

Image and Video Generation

Diffusion systems often begin with visual noise and repeatedly refine it toward an image that matches the prompt. Video systems extend this process across time while trying to maintain identity, motion, lighting and continuity from frame to frame. Other systems can animate a still image, replace a face, alter expressions, synchronise lips to new speech, remove objects or construct a scene from several sources. A single clip may therefore contain recorded, edited and generated elements together.

Voice Cloning and Synthetic Audio

Voice-cloning systems learn characteristics such as pitch, rhythm, pronunciation and timbre from samples of a person's speech. They can then generate new sentences in a voice that resembles that person. Background noise, compression and a short urgent message may hide weaknesses. This is why a familiar-sounding voice can no longer serve as sufficient proof of identity in a financial or reputational emergency.

Synthetic Personas and Production Pipelines

A deceptive operation may combine several systems: a language model writes a script, an image generator creates a face, a voice model produces narration, a video tool animates the image, and scheduled accounts publish the result repeatedly. The important clue may not appear inside one video. It may appear in the account's history, repetitive posting pattern, missing identity, copied descriptions, strange commercial links or the absence of any independent trace of the alleged person.

Two Essential Distinctions

AI-generated does not automatically mean false. A labelled reconstruction, illustration, translation or accessibility aid may communicate responsibly. Ethical evaluation asks whether

the audience is deceived, whether consent and dignity are respected, and whether the claim is true.

Authentic does not automatically mean trustworthy. A real photograph, document or recording may be old, selectively edited, staged, or described falsely. Many harmful posts therefore require no generative AI at all.

4) Why Intelligent People Are Deceived

People do not usually share a false claim because they have carefully studied it and chosen error. They share because the post reaches them through a trusted relationship, confirms something they already fear, offers a simple explanation, creates urgency, or appears to come from authority. Effective media literacy therefore needs both technical habits and self-awareness.

Emotion and Urgency

Anger, fear, disgust, wonder and pity can narrow attention. A message that says "share before this is deleted" tries to convert urgency into obedience. Emotion is not proof that a claim is false. It is a signal to slow down because the cost of a mistaken share may be high. The spiritual discipline of the pause creates room for freedom.

Familiarity and Repetition

Repeated statements can begin to feel true simply because they are familiar. Repetition across many accounts may look like independent confirmation even when all accounts copied the same original source. Count evidence, not reposts. Ten posts that point back to one unverified clip still provide only one unverified source.

Authority and Relationship

A priest, bishop, religious superior, family member or well-known institution carries relational authority. Impersonators borrow that trust. The more sensitive the request, the more important it is to confirm identity through a separate channel. Pastoral courtesy does not require us to abandon verification. A genuine leader will understand a prudent call-back before money, data or a public statement is released.

Confirmation and Group Identity

A claim that praises our group or condemns a rival can bypass scrutiny. We often demand evidence from opponents while accepting flattering claims about our own side. A disciplined investigator applies the same standard to content that agrees with him and content that disturbs him.

The Liars Dividend

Deepfakes create a second danger beyond believing a false recording. Once people know that convincing media can be generated, a guilty person can dismiss authentic evidence as synthetic. This is called the liar's dividend. A healthy response avoids both extremes: believing every vivid recording and rejecting every inconvenient one. Preserve originals, verify provenance, consult independent evidence and state the level of certainty honestly.

5) From Detection to Discernment

Earlier generations of AI media often produced obvious defects: extra fingers, garbled words, immobile faces, strange reflections or very short clips. Those examples remain useful for showing how generation can fail, but they are unsafe as fixed rules. Generators improve, editing can hide defects, and ordinary compression can make genuine media look unnatural. A person trained only to hunt glitches may become overconfident.

The central task is claim verification. Begin outside the pixels: identify the claim, source, date, place and requested action. Search laterally for independent evidence. Trace the item to its earliest available form. Only then inspect the media itself and use specialist tools where they add value.

The PASTOR Verification Method

Letter	Action	Guiding question
P	Pause	Why do I feel an immediate urge to believe, reject or share this?
A	Ask about the source	Who first published or sent this, and what can I verify about that source?
S	Search laterally	What do independent reliable sources say about the claim or source?
T	Trace to the original	Where is the earliest, longest and least edited version?
O	Observe the evidence	What do the media, metadata, provenance and inconsistencies actually show?
R	Respond responsibly	Should I verify, withhold, correct, report, preserve evidence or accompany a victim?

P Pause

Stop the automatic action. Do not forward, download repeatedly, accuse a person or send money. Read the exact wording. Notice urgency, emotional pressure and the consequences of being wrong. Ask whether the sender wants you to act before you think.

- Write the central claim in one neutral sentence.
- Separate what the post shows from what the caption says it shows.
- Choose "unsure" when the available evidence is insufficient.

A Ask About the Source

Do not confuse the account that forwarded the item with the source that created it. A trusted friend may unknowingly pass along a deceptive post. Open the profile or site. Look for the full name, institutional link, publication history, corrections, contact information and evidence of identity. A blue tick or polished design can be copied and is not enough by itself.

- Who recorded, wrote or uploaded the item first?
- Does the account have a credible history before this viral post?
- Is the official website linked from another confirmed institutional source?
- Is the message asking for secrecy, money, credentials or an unusual favour?

S Search Laterally

Lateral reading means leaving the suspicious page and opening new searches to learn what other reliable sources say. Professional fact-checkers do this because a deceptive page controls its own appearance, biography and claims. Search the person's name, organisation, exact quotation, key event and distinctive phrases. Use several sources that are independent of one another.

- Search the exact sentence in quotation marks.
- Search the name plus words such as statement, transcript, clarification, fraud or fact check.
- For Church claims, look for the original diocesan, bishops' conference or Vatican source.
- For a local event, consult credible local reporting and people who can directly confirm it.

T Trace to the Original

A repost often strips away date, caption, credit, resolution and context. Find the earliest available upload, the complete video, the full homily, the original photograph or the official document. Compare what was removed or added. When many accounts repeat the same item, trace their links backward until you reach the first known source or a dead end.

- Prefer the longest available version over a short extract.
- Check whether the original caption differs from the viral caption.
- Check the upload date against the date claimed in the post.
- Preserve the URL and a screenshot if the matter may require reporting.

O Observe the Evidence

Only now inspect the media closely. Look for internal inconsistencies, but treat them as clues. Use reverse-image search, keyframe extraction, metadata or Content Credentials when appropriate. Compare the person's known speech, setting and timeline. A strong conclusion normally rests on several forms of evidence rather than one visual oddity.

R Respond Responsibly

The correct response may be to share a verification, ask the sender for the original, contact the person concerned, warn a parish group, report an account, preserve evidence for police, or simply withhold judgement. Do not turn every uncertainty into a public accusation. Do not repost harmful material more widely while trying to condemn it.

6) Inspecting Media Without False Confidence

Inspection can reveal reasons for further investigation. It rarely establishes the whole truth on its own. Social platforms re-encode files, messaging apps reduce quality, stabilisation changes motion, portrait mode alters edges, and ordinary editing changes colour and sound. The same symptom may have innocent and deceptive explanations.

Possible Visual Clues

- Temporal inconsistency: a face, object, garment, logo or background changes shape across frames.
- Interaction errors: hands pass through objects, contact points slip, or an object appears without a plausible cause.
- Lighting and reflection conflicts: shadows, highlights, mirrors or eyeglasses do not agree with the scene.
- Identity drift: facial features, age, hairline, teeth or accessories subtly change during movement.
- Text and symbol instability: writing changes between frames or remains semantically meaningless.
- Boundary artefacts: face edges, hair, ears or jewellery blend unnaturally, especially during turns.
- Biological or physical impossibility: motion, anatomy, gravity or cause-and-effect is inconsistent.

None of these proves that a file is synthetic. A low frame rate, motion blur, beauty filter, poor connection or edited background can produce similar effects. Conversely, a high-quality deepfake may show none of them.

Possible Audio Clues

- Cadence or emphasis does not fit the speaker's normal manner or the meaning of the sentence.
- Breathing, mouth sounds, room echo or background noise changes abruptly between phrases.
- Pronunciation of names, local words or emotional expressions is unusual.
- Lip movement and sound drift apart, especially after a cut or head turn.
- The voice remains unnaturally clean while the surrounding scene is noisy, or the acoustic space does not match the image.

Muffled sound is not a reliable AI test. WhatsApp forwarding, old microphones, noise reduction and repeated compression often produce muddy audio. When identity matters, confirm through a known number or another independent channel.

Text Documents and Quote Cards

- Search a distinctive sentence from the alleged quotation in quotation marks.
- Find the complete document, speech, homily or interview rather than relying on the graphic.
- Check the issuing office, date, reference number, signature block and official distribution channel.
- Compare the language and formatting with verified documents from the same institution.
- Watch for invented citations, impossible dates, wrong office titles or a link that does not belong to the institution's domain.

Weak Rules That Should Not Be Taught as Tests

Weak rule	Why it fails	Better question
The video is under fifteen seconds	Real clips are cropped for Reels and Shorts; synthetic video can be extended and edited.	Can I trace the full clip and confirm the event independently?

Weak rule	Why it fails	Better question
The hands look correct	Current systems often generate convincing hands; blur can make real hands look wrong.	Are there several consistent clues, and what does provenance show?
The throat does not move	Anatomy, angle, clothing, frame rate and compression make this unreliable.	Does the speaker confirm the recording, and is there a full original?
The audio is muffled	Compression and poor microphones commonly cause this.	Does identity hold across a separate trusted channel?
There is only one camera angle	Many real events have only one surviving recording.	Is there independent evidence beyond the video itself?
A detector says ninety-nine percent	Detectors can produce false positives and fail on unfamiliar generation methods.	What other evidence supports or contradicts the detector?

7 The Investigators Toolkit

Tools are most useful when they answer a specific question. Begin with the fastest and strongest evidence. A direct call to a known number may establish identity more reliably than ten minutes of pixel inspection. A reverse-image search may reveal that a supposedly new photograph is five years old. A tool result should be recorded as evidence, not treated as an oracle.

Level One Direct Confirmation and Basic Search

Direct call-back. For requests involving money, credentials, secrecy or urgent action, end the incoming call and contact the person through a number already known to you. Do not use the number supplied in the suspicious message.

Shared verification phrase. A diocese, seminary or parish office can establish a private phrase or two-person approval process for exceptional financial requests. The phrase should not be transmitted in public posts.

Ordinary web search. Search exact quotations, names, dates, places and document titles. Open the source itself rather than trusting the search-result snippet.

Official sources. For a papal or Vatican quotation, use the Holy See website. For a diocesan announcement, use the diocese's official channels. For a Government of India claim, the PIB Fact Check Unit verifies only claims relating to Union Government bodies and policies.

Level Two Reverse Image and Video Search

Google Lens. Upload the image or a clear screenshot. Crop to the distinctive subject when necessary. Review visually similar results, earlier dates and pages that use the same picture with another caption. Reverse search does not tell you automatically which use is correct; it helps you trace the image's history.

InVID WeVerify. Use the browser extension to extract keyframes from a video, inspect available information and reverse-search individual frames. Choose frames with faces, signs, buildings or unique objects rather than blurred transitions.

A practical InVID sequence:

1. Copy the public URL of the suspicious video or save a lawful working copy.
2. Open the Keyframes function and generate representative frames.

3. Choose two or three distinctive frames from different points in the video.
4. Reverse-search each frame through more than one available engine when possible.
5. Compare the earliest result, original caption, date, place and length with the viral post.
6. Record the URLs and what each result actually proves.

Level Three Provenance and Metadata

Content Credentials. The C2PA standard allows signed provenance information to describe how an asset was created and edited. A verification interface can show the signer, capture or creation information and declared edit history when such credentials are present and valid.

Two limitations are essential. First, the absence of Content Credentials does not prove that an image is fake; many cameras, apps and platforms do not preserve them. Second, provenance can help establish origin and editing history, but it does not automatically prove that the caption or interpretation is true. A genuine photograph of a real flood can still be falsely described as today's local flood.

Ordinary metadata. EXIF and file metadata may contain date, device, location or software information. Social platforms often strip these fields, and fields can be edited. Metadata is supporting evidence, not a complete verdict.

Level Four Account and Domain Investigation

- Check when the account began posting and whether older posts support the claimed identity.
- Look for abrupt changes in topic, language, profile name or audience, which may indicate a hijacked account.
- Compare the account link with the institution's confirmed website and public directory.
- Use ICANN Lookup for domain-registration information, while remembering that privacy services and new legitimate sites limit what registration data can prove.
- Treat follower count, polished branding and frequent posting as weak evidence. All can be purchased or automated.

Level Five AI Detectors

Commercial detectors estimate whether patterns resemble known synthetic media. They may help triage a large set of files or point an analyst toward closer examination. They can also misclassify authentic content, especially after cropping, recompression, filters or unfamiliar generation methods. Detector scores should therefore be recorded as one item of evidence and never announced as a final judgement by themselves.

If you demonstrate a detector in class, test it with several kinds of material: an original photograph, a compressed screenshot of that photograph, a clearly generated image, and a generated image that has been edited or reposted. The differences teach more than a single impressive percentage.

An Evidence Log

For a serious case, write down the investigation rather than relying on memory:

- exact claim being tested;
- URL, account and time received;
- original file or best available copy;
- search terms used and sources consulted;

- earliest source found and contextual differences;
- tool results and their limitations;
- people contacted for confirmation;
- conclusion, confidence level and unresolved questions.

8 Pastoral Cases

The Bishop Requests an Urgent Transfer

A seminarian receives a WhatsApp voice message from an account bearing the bishop's photograph. The voice sounds familiar and asks for an immediate transfer to help a person in distress. The sender says the matter is confidential and that there is no time to call.

Apply PASTOR. Pause and make no payment. Ask whether the number is already known. Search is secondary here; direct confirmation is strongest. Call the bishop's office or another authorised person through a saved number. Preserve the message and account details. If it is fraudulent, alert those who may be targeted and report the attempt. Do not publicly circulate the cloned voice unless needed by investigators.

A Priest Appears to Make an Inflammatory Statement

A short video appears to show a priest insulting another community. Anger spreads before the priest can respond. The clip may be synthetic, selectively edited or authentic.

Apply PASTOR. Write down the exact alleged statement. Identify the first uploader. Find the full recording and event. Search for other recordings, a transcript or independent witnesses. Contact the priest or institution before defending or condemning him. Examine cuts, audio continuity and lip synchronisation only as supporting evidence. Publish a correction only after the evidence is adequate, and avoid repeating the harmful fragment more than necessary.

A Miraculous Statue Moves on CCTV

A low-resolution clip appears to show a statue moving. Some viewers see a miracle; others immediately accuse the parish of fraud.

Apply PASTOR. Protect faith from both gullibility and contempt. Obtain the original CCTV file rather than a phone recording of a screen. Confirm date, camera position, frame rate, maintenance history and the people with access to the footage. Look for compression, dropped frames, insects near the lens, reflections, camera movement or editing. Consult competent technical help. Do not make a theological claim that exceeds the evidence.

The Pope or Bishop Says

A beautifully designed quotation card attributes a striking sentence to the Pope or bishop. The sentence fits what the audience hopes to hear, but no speech or document is cited.

Apply PASTOR. Search a distinctive part of the sentence in quotation marks. Search the official website and complete documents. Check whether the card paraphrases, mistranslates or joins separate sentences. If the quotation cannot be found, describe it as unverified rather than inventing a source. When correcting it, provide the actual link so that people can read the full context.

A Communal Rumour Arrives During Tension

A dramatic video claims that members of one community attacked another nearby. The location cannot be confirmed, but people are already planning a response.

Apply PASTOR. Do not forward the clip to ask whether it is true; that can spread it further. Contact verified local authorities and trusted leaders on the ground. Reverse-search frames and compare landmarks, weather, language, vehicle plates and earlier uploads. Preserve evidence and alert police if there is a credible risk of violence. Public communication should state only verified facts, discourage retaliation and provide a clear source for updates.

A Sexualised Deepfake Targets a Student or Parishioner

Synthetic or manipulated intimate imagery is a form of abuse. The immediate pastoral task is the safety and dignity of the person targeted, not satisfying public curiosity. Do not request additional copies or circulate the image for informal analysis. Preserve necessary evidence carefully, restrict access, help the victim report the content and seek professional or legal support. Avoid language that blames the victim. In India, cybercrime complaints can be filed through the National Cyber Crime Reporting Portal; urgent financial cyber fraud can be reported through helpline 1930. Local police should be contacted in an emergency.

9 Correcting Falsehood Without Enlarging Harm

A correction can fail if it repeats the sensational falsehood, humiliates the sharer or leaves people without a believable explanation. The aim is to replace error with a clear account of what the evidence shows.

A Clear Correction Pattern

1. Lead with the verified fact in plain language.
2. Identify the false or misleading claim briefly and only as much as necessary.
3. Explain how the verification was done and link to the original evidence.
4. Offer the accurate context or alternative explanation.
5. Tell the audience what to do: delete, stop forwarding, report, contact the institution, or consult the verified update channel.

Example. *"The diocese has not issued an order cancelling Sunday Masses. The image circulating on WhatsApp uses an old letterhead and does not appear on the diocesan website or official channel. The Chancellor's office has confirmed that the notice is false. Please delete it and share this verified clarification instead."*

Correct the Claim Without Shaming the Person

People often share false information because they are anxious, protective or trusting. Public ridicule can push them deeper into defensive belief. Begin with a shared value: concern for the community, safety, faith or truth. Then show the evidence. A correction should be firm about the claim and gentle toward the person unless that person is deliberately causing harm.

Protect the Victim

Do not embed the harmful video in every correction. A description, blurred frame or link available only to investigators may be sufficient. Remove personal contact details. Obtain consent before using a victim's story in training. If a false clip concerns alleged abuse or

misconduct, verification must not become an excuse to dismiss a complainant or suppress legitimate investigation. Truth-seeking includes uncomfortable facts about Church institutions as well as false accusations against them.

A Parish or Diocesan Response Protocol

- Designate who receives suspicious-content reports and who can speak publicly.
- Maintain an updated directory of official phone numbers and communication channels.
- Require a second confirmation for exceptional financial requests.
- Preserve the original message, URL, time and account details before content disappears.
- Assess immediate risks to safety, finance, reputation and communal peace.
- Contact the impersonated person and relevant authorities through trusted channels.
- Issue a concise, evidence-based correction from the recognised official account.
- Review what allowed the incident to spread and update training or procedures.

10 Practical Investigation Exercise

The aim of the exercise is not to guess correctly from appearance. It is to produce a transparent conclusion that another person can follow. Work in groups of three or four. Give each group one post containing an image, audio clip, video or quotation card. At least one example should be authentic but miscaptioned, and at least one should look strange but be authentic.

Round One First Impression

Allow two minutes. Each group records one of three responses: authentic, synthetic or unsure. They must also record the emotion and visual detail that influenced the first impression.

Round Two PASTOR Investigation

Allow twelve to fifteen minutes. The group must:

- write the exact claim;
- identify the first available source;
- conduct at least two lateral searches;
- trace the media toward an original version;
- use one appropriate tool, such as Lens, InVID or Content Credentials;
- state what the evidence proves and what remains unknown.

Round Three Pastoral Response

Allow five minutes. The group prepares a response of no more than one hundred words for a parish WhatsApp group. It should state the verified facts, avoid unnecessary repetition of the falsehood, protect the dignity of persons involved and give a clear action.

Evaluation Questions

- Did the group investigate the claim rather than merely inspect the pixels?
- Did it use sources independent of the original post?
- Did it distinguish evidence from inference?
- Did it allow an unverified conclusion when evidence was insufficient?
- Would its public response reduce harm and help people act responsibly?

The PASTOR Media Verification Worksheet

Title or description of media _____

Platform and URL _____

Date and time received _____

Person or group who sent it _____

P Pause

☐ I have not forwarded, paid, accused or publicly reacted.

☐ I can state the central claim in one neutral sentence.

☐ I have noticed the emotion or urgency the post creates.

Exact claim _____

A Ask About the Source

☐ I identified the original creator or earliest available uploader.

☐ I checked the account or website outside the post itself.

☐ I checked whether the request involves money, secrecy, credentials or unusual action.

Source findings _____

S Search Laterally

☐ I searched the exact quotation or key phrase.

☐ I consulted at least two independent and relevant sources.

☐ I checked the appropriate official source where one exists.

Independent sources _____

T Trace to the Original

☐ I found the earliest available image, full video, transcript or document.

☐ I compared the original date, caption, place and length with the viral post.

☐ I preserved the relevant URL or evidence if reporting may be necessary.

Original source and context _____

O Observe the Evidence

☐ I inspected visual, audio or textual consistency without treating one clue as proof.

☐ I used reverse search, keyframes, provenance or metadata only where relevant.

☐ I treated any detector score as supporting evidence rather than a verdict.

Media and tool findings _____

R Respond Responsibly

Choose the most accurate present finding:

☐ Verified Evidence sufficiently supports the claim and context.

☐ Unverified Available evidence is insufficient.

☐ Miscontextualised The media may be genuine, but the caption or context is false.

☐ Manipulated Genuine material has been altered or selectively edited.

☐ Synthetic Evidence supports AI generation or substantial AI alteration.

☐ Satire or parody The media is not presented as factual reporting.

☐ Scam or impersonation The content seeks trust, payment, access or disclosure.

Confidence level ☐ Low ☐ Moderate ☐ High

What remains uncertain _____

Responsible next action _____

Quick Reference Before You Believe or Share

P Pause. Name the claim and notice the emotional pressure.

A Ask about the source. Identify who first created or published it.

S Search laterally. Leave the post and consult independent reliable sources.

T Trace to the original. Find the earliest, longest and least edited version.

O Observe the evidence. Inspect media, provenance and tool results as clues.

R Respond responsibly. Verify, withhold, correct, report or accompany.

Do not ask only whether the media looks fake. Ask how you know the claim is true. "I do not know yet" is a responsible conclusion when evidence is incomplete.

Final pastoral maxim

Verify before you amplify. When the evidence is incomplete, say so. When a person is harmed, protect dignity while pursuing truth.

References and Further Reading

- C2PA. Content Credentials specifications. <https://spec.c2pa.org/specifications/specifications/2.4/>
- Chandra, B., Dunietz, J., Roberts, K., Lee, Y., Fontana, P., and Awad, G. 2024. Reducing Risks Posed by Synthetic Content. NIST AI 100-4. <https://www.nist.gov/publications/reducing-risks-posed-synthetic-content-overview-technical-approaches-digital-content>
- Dicastery for the Doctrine of the Faith and Dicastery for Culture and Education. 2025. Antiqua et Nova. https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_dcf_doc_20250128_antiqua-et-nova_en.html
- Francis. 2019. Christus Vivit. https://www.vatican.va/content/francesco/en/apost_exhortations/documents/papa-francesco_esortazione-ap_20190325_christus-vivit.html
- Francis. 2024. Artificial Intelligence and the Wisdom of the Heart. Message for the 58th World Day of Social Communications. <https://www.vatican.va/content/francesco/en/messages/communications/documents/20240124-messaggio-comunicazioni-sociali.html>
- InVID Project. InVID WeVerify Verification Plugin. <https://www.invid-project.eu/tools-and-services/invid-verification-plugin/>
- Leo XIV. 2026. Magnifica Humanitas. <https://www.vatican.va/content/leo-xiv/en/encyclicals/documents/20260515-magnifica-humanitas.html>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., and others. 2020. The Debunking Handbook 2020. <https://digitalcommons.unl.edu/scholcom/245/>
- McGrew, S., and Breakstone, J. 2023. Civic Online Reasoning Across the Curriculum. AERA Open. <https://doi.org/10.1177/23328584231176451>
- NSA, FBI, and CISA. 2023. Contextualizing Deepfake Threats to Organizations. <https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-DEEPFAKE-THREATS.PDF>
- Wardle, C., and Derakhshan, H. 2017. Information Disorder. Council of Europe. <https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>

Practical Links for India

- National Cyber Crime Reporting Portal. <https://cybercrime.gov.in/>
- Indian Cybercrime Coordination Centre and helpline information. <https://i4c.mha.gov.in/ncrp.aspx>
- PIB Fact Check Unit for Union Government related claims. <https://www.pib.gov.in/aboutfactcheck.aspx>
- ICANN Registration Data Lookup. <https://lookup.icann.org/>
- Content Credentials Verify. <https://contentcredentials.org/verify>